Lecture Notes in Operations Research

Menggang Li · Guowei Hua · Xiaowen Fu · Anqiang Huang · Dan Chang *Editors*

# IEIS 2022

## Proceedings of 9th International Conference on Industrial Economics System and Industrial Security Engineering

Springer

# Lecture Notes in Operations Research

Lecture Notes in Operations Research is an interdisciplinary book series which provides a platform for the cutting-edge research and developments in both operations research and operations management field. The purview of this series is global, encompassing all nations and areas of the world.

It comprises for instance, mathematical optimization, mathematical modeling, statistical analysis, queueing theory and other stochastic-process models, Markov decision processes, econometric methods, data envelopment analysis, decision analysis, supply chain management, transportation logistics, process design, operations strategy, facilities planning, production planning and inventory control.

LNOR publishes edited conference proceedings, contributed volumes that present firsthand information on the latest research results and pioneering innovations as well as new perspectives on classical fields. The target audience of LNOR consists of students, researchers as well as industry professionals.

Menggang Li · Guowei Hua · Xiaowen Fu ·
Anqiang Huang · Dan Chang
Editors

# IEIS 2022

## Proceedings of 9th International Conference on Industrial Economics System and Industrial Security Engineering

Springer

*Editors*
Menggang Li
National Academy of Economic Security
Beijing, China

Xiaowen Fu
Department of Industrial and Systems Engineering
Hong Kong Polytechnic University
Kowloon, Hong Kong

Dan Chang
School of Economics and Management
Beijing Jiaotong University
Beijing, China

Guowei Hua
School of Economics and Management
Beijing Jiaotong University
Beijing, China

Anqiang Huang
School of Economics and Management
Beijing Jiaotong University
Beijing, China

ISSN 2731-040X ISSN 2731-0418 (electronic)
Lecture Notes in Operations Research
ISBN 978-981-99-3617-5 ISBN 978-981-99-3618-2 (eBook)
https://doi.org/10.1007/978-981-99-3618-2

# Contents

# The Impact of Digitalization on Teleworking Communication and Trust in the Post Epidemic Era

Yuejiao Zhao[1(✉)], Menggang Li[1], Xiaojun Jia[1], Weizhen Gao[1], and Li Zheng[2]

[1] Beijing Jiaotong University, National Academy of Economic Security, Beijing 100044, China
{9756,mgli1}@bjtu.edu.cn
[2] Qingdao University, Business School, Qingdao, China

**Abstract.** The research topic of this paper mainly focuses on the problem of communication and trust in Chinese enterprises' telecommuting under the situation of crisis. Through semi-structured interviews with employees of 32 different enterprises during the epidemic isolation period, 290000 Chinese words of text interview data were obtained. With the help of manual analysis and software NVIVO, the grounded theory method was used for multi-stage coding and relationship analysis. We found that unlike virtual teams and temporary organizations, telecommuting required in special environments relies more on past experience, common experience and commitment to tasks in terms of trust.

**Keywords:** Telecommuting · Digitalization · the Post Epidemic Era · Virtual Team

## 1 Introduction

Telecommuting (telecommuting, long distance working, work at home) was originally defined as using telecommunications technology to partially or completely replace on-site office (Nilles 1995; Mokhtarian 1991), also known as working at home (but should not include overtime at home other than on-site office). Along with telecommuting is the management of virtual team. When a department or a team of the company needs to work remotely for a long time, they form a virtual team.

Before 2020, in China, although this officially defined telecommuting has been involved in some enterprises (especially multinational enterprises and Internet enterprises: Tencent, Alibaba and Fuyao Glass), it is still limited to the stage of "heard of" but not really practiced for the whole market and enterprises. However, a survey on telecommuting in the United States found that 85% of Companies in the United States had telecommuting in 2016 (Hoch and Dulebohn 2017). The advantages of telecommuting lie in saving time and cost, obtaining diversified knowledge resources, having more opportunities to cooperate, solving global problems, helping employees balance their families and lives and reducing fixed costs. It has become an indispensable way of work complementary to on-site office.


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 1–9, 2023.
https://doi.org/10.1007/978-981-99-3618-2_1

On January 29, 2020, when the COVID-19 disease swept the world, the Chinese government called on residents to be isolated at home to cope with the pandemic and keep residents' lives safe. At the same time, almost all enterprises need to turn on-site office into remote office (work from home). Due to the lack of experience in telecommuting in the past and the sudden epidemic requirements, the management of the enterprise was unprepared, but it is also necessary to carry out telecommuting throughout the company. However, teleworking in this situation is different from telecommuting in the existing literature.

First, the situation is different. Chinese enterprises are forced to fully implement telecommuting in an extremely short period of time (unplanned and unexpected), which is different from the previous literature that the construction and development of telecommuting have an advance plan or a small number of groups realize the pilot telecommuting first. This will lead to more operational and psychological problems that are not involved in the literature.

Second, the change of working methods in emergency situations is different from the newly established telecommuting team. China adopts telecommuting. Although employees are isolated from each other at home and cannot see each other, they have accumulated a lot of cooperation experience in the previous long-term on-site office, and their colleagues and superiors and subordinates know each other very well. This is different from the problems faced by the team construction required by the newly established remote office team.

Third, the unique cultural and behavioral preferences of Chinese people will also lead to different problems from those encountered in Western telecommuting. Therefore, based on the above three points, it is necessary for us to specially study the problems and solutions that Chinese enterprises are forced to suddenly and comprehensively implement telecommuting under the situation of crisis.

The research topic of this paper mainly focuses on the problem of communication and trust in Chinese enterprises' telecommuting under the situation of crisis. Through semi-structured interviews with employees of 32 different enterprises during the epidemic isolation period, 290000 words of text interview data were obtained. With the help of manual analysis and software NVIVO, the grounded theory method was used for multi-stage coding and relationship analysis, and the following problems were studied. First, what are the new problems encountered by Chinese enterprises' telecommuting employees in crisis situations (focusing on the new problems in new situations that have not been found in the existing literature). Second, what are the manifestations of trust and distrust among colleagues and between superiors and subordinates in telecommuting? The process and results of trust development and transformation among employees in China's crisis situation? Third, what changes have taken place in the communication in the team and how do they affect the trust between employees. Fourth, what communication strategies can improve team trust and team performance.

The innovations of this paper are as following. Firstly, this paper focuses on forced telecommuting under infectious diseases, which provides the diversity of situations for the research of telecommuting trust and enriches the focus of communication and

improving trust in crisis environment. The sample of this paper is based on the background of China's epidemic events. The sudden major infectious disease makes employees in crisis. The urging of government forces also requires enterprises to implement the company wide remote office without preparation. During this period, the establishment of employee trust, the management of remote team by managers and the impact of communication on trust are unique. But at the same time, this situation can provide reference for changing working methods in the future or in other countries in case of similar emergencies.

Second, we found the dynamic transformation of trust among team members in telecommuting. Most of the existing literature studies trust as a static concept. In a few studies of virtual team trust, it is found that trust is gradually strengthened with the process of team construction, team members' familiarity and cooperation. Based on the actual data, this paper found that under the outbreak of the epidemic, with the process of employees' familiarity with telecommuting from unfamiliar to familiar, the trust between teams decreased from high trust in on-site office to low trust in the early stage of telecommuting to high trust in the end.

Thirdly, it further reveals the impact of communication strategies on trust and enriches the research on the antecedents of trust. The existing research on the antecedent factors of trust in telecommuting team is discussed and verified from four aspects: member characteristics, team characteristics, team management and interpersonal relationship, involving factors at the individual level, team level and relationship level. For the research on the impact of communication on trust, communication is only explored as a situational variable, which obviously ignores the important role of communication in trust building and work efficiency improvement in telecommuting.

Therefore, this paper focuses on the way, content and feedback of remote communication between familiar employees in telecommuting, and explores its impact on trust. This will make up for the literature that affects the elements of trust in telework theory. At the same time, it will provide operational methods of communication to improve trust, so as to guide practice.

## 2 Theory and Research Question

Enterprises will face more problems that are not found in ordinary telecommuting, such as greater challenges to the office system, complex changes in employees' emotions and psychology, fuzzy task assignment and assessment, etc. If we are still confined to the existing literature of telecommuting to guide practice, it will produce possible deviation and invalid results. Therefore, it is necessary for us to take the team and members of telecommuting in emergency environment as the research object to explore the unique problems in communication, the unique factors affecting trust, and how to establish Chinese trust to improve the efficiency of telecommuting at home.

O'Hara Devereaux and Johansen (1994) summarized the possible problems of telecommuting, including low individual commitment, role overload, role ambiguity, absenteeism, and social loafing may be exaggerated, etc. Although these problems occur from time to time in on-site office, these problems are more serious in telecommuting. At the same time, telecommuting has a significant disadvantage in gaining the trust of

customers. Customers will think that they have no permanent, reliable and sustainable sellers (in the interview with Li PJ and Liang CZ). Handy (1995) believes that only trust can solve the above problems. Trust can reduce the obstacles caused by geographical distance and virtual technology, help people take risks and take actions without meeting, and promote telecommuting to play its due role.

Pearson et al. (2019) mentioned three difficulties to overcome in telecommuting: First, communication problems, including the appointment of meeting time, inability to communicate face-to-face, loss of physical communication and expression communication; Second, technical problems, including the skilled use of a series of technologies; Third, the diversity of the team, such as how to build trust, work guidelines, share information, measure performance and so on. Although trust is only a part of it, the establishment of trust can significantly improve cooperation and knowledge sharing, so as to promote the improvement of other difficulties. Jarvenpaa et al. (1998) also believe that trust is the heart of remote work. Our interview data also significantly prove this.

There are many definitions of trust. This paper adopts the most widely accepted definition of trust by McAllister: the definition of trust is that the expectations of another's conduct or an acceptance of vulnerability. The forms of trust include trust behavior, disposition and expectations of subjective environment.

Trust develops during the virtual work. The leaders of the virtual team have to make their effort to create, reinforce, maintain trust between members. However, trust is totally different in the team working in long distance under the emergency circumstance. Because in the past, when we worked in the office, we had established basic trust. For example, I know how my leader works and whether he has a bad temper; One of my colleagues is a little slow in doing things, but she is very serious, and the final result is often satisfactory; Another colleague of mine is a hot and hot character. If you cooperate with him, he will constantly supervise your work progress and so on. However, after becoming telecommuting, I don't know whether my colleagues are working at home or in other places and how they are working, which will have an impact on the communication strategies and trust established in the past.

Trust is the source of promoting subordinates to spontaneously accept leadership authority, the main inducement of endogenous work motivation, and also an effective management means when it is difficult to realize real-time on-site supervision in the special situation of telecommuting. Unlike virtual teams and temporary organizations, telecommuting required in special environments relies more on past experience, common experience and commitment to tasks (companies) in terms of trust.

Overall, we propose the following research questions:

Context: temporary long distance work (work at home) to tackle the problems under emergency.

Research Question 1:what are the new problems employee encountered when they turn to long distance working?

Research Question 2: whether distrust exist between peers and upper levels during working at home? What are the forms of distrust during working at home? What are the forms of trust during working at home under the certain pandemic context?

Research Question 3: what are the kinds of behaviors or communication ways which would trigger distrust during long distance working ?

Research Question 4: what are the strategies to control/reduce distrust so that to enhance working performance?

## 3 Research Method

From February 3 to March 20, 2021, we spent seven weeks interviewing employees with telecommuting experience. The interview form is telephone interview, and the interview outline includes 12 questions. The time of each interviewer is more than 40 min.

Some of interview questions are as follows: How did you work and how did you communicate with leaders, colleagues and subordinates before the epidemic? How did the company communicate when the epidemic first broke out? What were the specific measures? How to update the epidemic information in the enterprise? After the outbreak of the epidemic, what do you think is the most obvious change in telecommuting compared with the previous on-site working? What factors, for examples, your current job position, job content, commitment to the organization, are related to these changes?

We conducted semi-structured interviews with employees of 32 different enterprises and obtained 290000 Chinese words of written interview data. Then, the grounded theory is used to import the data into NVivo software for selective coding and spindle coding. The characteristics of interviewees is shown in Table 1. The age and working years of the interviewees are widely distributed. The sample also includes many types of companies. These shows that shows that our samples are diverse for further analysis.

**Table 1.** Characteristices of sample

| Characteristics | NO | Characteristics | NO |
|---|---|---|---|
| Female | 20 | Private firms | 22 |
| Male | 12 | State-owned | 10 |
| Age (20 ~ 30) | 9 | Listed firms | 25 |
| Age (30 ~ 40) | 23 | Non listed firms | 7 |
| Working year (0 ~ 5) | 12 | Employees (0 ~ 1000) | 11 |
| Working year (6 ~ 10) | 14 | Employees (1000 +) | 21 |
| Working year (11 ~ 15) | 6 | Average working hours | 8.125 |

## 4 Data Analysis

Coding in grounded theory is the important process of analyzing the data. Suppose that the author as an actor in the process so as to intervene, manipulate, act on conceptualize and use specific techniques to generate and discover the concept, relationship and the theory. In this paper, Interview materials are the raw data. The author read every sentences and do the open and selective coding work.

Excerpt from grounded theory analysis is shown in Table 2. In the first example, we get four concepts. The first three belong to the communication methods of the company's leaders to their employees when working at home. The fourth is the result of trust felt by employees. Then we summarize the first three as rough communication and we found that rough communication from leaders may lead low trust. on the contrary, the gentle and positive communication results in high level of trust, in turn bring high performance.

**Table 2.** Grounded Analysis

| Raw Data | Open coding | Selective coding |
|---|---|---|
| After 18 o***'clock**, like after work every day, the leader also asked us to **hold a two-hour summary meeting** and a meeting of more than 100 people……**If you were called by your leader** and it happened you were not online, even if you just happen to be disconnected to internet, it supposed that you were absent from work……so you **have to take the computer or cellphone with you** even you go to toilet* | *forced overtime* | *Rough communication* |
| | *Simple and rough post inspection* | |
| | *Tense working at home* | |
| | *Low job satisfaction* | *low-level trust* |
| *The leader told me in a private letter to **praise me** for doing a good job.* I felt so happy *about it…… The leader approved this thing, and I felt that I had a certain sense of **belonging and my value**……I would **do anything like this again*** | *Praise from leader* | *Positive communication* |
| | *Higher sense of belonging* | *high-level trust* |
| | *willing to work* | *High personal performance* |

What data told us:

1. After telecommuting, the way of communication has changed. After telecommuting, it is mainly to make phone calls and send text messages.

In the past, face-to-face communication was the main work, supplemented by telephone and text software communication. Because face-to-face communication can obtain other implicit information besides the speech content, such as body language, expression, posture, environment and so on. Through these implicit information, we can know whether the other party recognizes or understands the content of the speech, whether it has been well communicated, whether it has been affirmed, etc.

In telecommuting, the way of telephone is used for urgent things and things that are difficult to be explained clearly in words. Because you can get feedback while conveying information, which is conducive to the continuation and development of things. Most of the things that are not in a hurry are text messages. Most of the information with specific

tasks is often accompanied by time nodes to ensure that the other party understands the process of the task.

2. The small module system of information communication, that is, the specific working mode under the epidemic situation, shall be formulated and assessed by specific departments.

Under the epidemic situation, the company gives a broad notice of the universality of telecommuting, mostly through e-mail. The working hours, clock outs, regular meetings and assessment of employees in specific departments are mostly formulated and assessed by the person in charge of specific departments.

3. Under telecommuting, obvious changes include that: network stability, inadaptability of remote office software (system), unable to receive timely and accurate feedback, do not understand the work progress of subordinates or collaborators, and need to communicate at any time, sometimes even supervise, when there is an emergency you can't contact each other, and although the colleague receives the task, he doesn't know whether his understanding of the task is accurate (in Song Y, Cui FF).

4. The timeliness of information reply will affect trust. For example, when you shop on Taobao and contact the merchant through Wangwang, most of the time the merchant can give a timely reply within 10 min. You will think that the merchant is a trustworthy merchant. In a remote office at home, if you can always reply to the information given to you by the leader in time, the leader will tend to believe that you are working during working hours rather than doing housework or other entertainment.

5. In the process of telecommuting under the epidemic in China, the trust between leaders and employees experienced a dynamic process of first decline and then improvement.

Benbasat (2006) believes that familiarity has an impact on trust; Previous experience will have an impact on trust. If you have more experience, you will reduce mistrust to a great extent.

What shocked us is that this study found some different conclusions compared to existing studies from the data. Previous studies have shown the low level of trust in virtual teams during the teleworking. In China, human feelings, sympathy and understanding run through the whole interpersonal relationship. When the epidemic suddenly comes and forces employees to work remotely, the trust between leaders and employees may decline due to the sudden changes in working methods and working environment at the beginning. But with the passage of time, people become familiar with digital working so this trust will gradually increase with the understanding and cooperation accumulated in previous field work. That is to say, the understanding, sympathy and trust from leaders can significantly improve employees' organization commitment and work performance.

For example, one of female participant said "*I take care of baby and work at the same time. My leader understood and showed sympathy to me. So he assigned me the works that can be handled in flexible time. In order to be worthy of the trust from leaders, I usually do it very seriously and work hard.*"

## 5 Conclusion

This study explored how the communication practice influence trust between leaders and employees, organizational commitment and employees' performance during the teleworking in the Post Epidemic Era. This finding provided new insights into communication in new situation in China and showed different conclusions.

Under the epidemic, digitalization promotes communication and trust. Timely communication, positive encouragement from leaders, the way of result control, and previous working experience can promote trust. Result control may not significantly improve employees' performance efficiency in the short term, but in the long run, it improves the quality of results and enhances employees' long-term commitment and performance to the organization.

Strict process control reduces the trust between leaders and employees. Trust is reciprocal, mistrust leads to mistrust, and trust gains trust. Process control may improve employees' work efficiency in a short time, but in the long run, it reduces employees' organizational commitment and performance.

Leaders' communication is aimed at personal work status rather than work content, which will reduce employees' sense of trust, thus reducing organizational commitment and satisfaction. Leaders' vague information reply is considered not to support employees or give employees a negative evaluation, which reduces employees' sense of trust and gives them negative motivation. Leaders not only pay attention to the quality of work results, but also are willing to listen to employees' difficulties in work and explain the reasons for work results to employees, which will improve employees' sense of trust, improve job satisfaction and motivate employees.

This study puts forward some communication practices suitable for telecommuting in Chinese situation, such as providing information to each other in time, frank communication, communicating with a moderate frequency, positive incentives such as more praise from leaders, treating subordinates fairly, and adopting the communication method of "commenting on work content and guiding work attitude".

This research enriches the literatures on communication and trust in telecommuting in the context of epidemic in China. The differences between the conclusions of virtual teams in foreign countries reflect the characteristics of Chinese culture. It is expected that this study can provide reference for business management methods and communication in telecommuting.

## References

Nilles, J.: Telecommunications and organizational decentralization. IEEE Trans. Commun. **23**, 1142–1147 (1975)

Bailey, D.E., Kurland, N.B.: A review of telework research: Findings, new directions, and lessons for the study of modern work. J. Organ. Behav. **23**, 383–400 (2002)

Hacker, J.V., Johnson, M., Saunders, C., et al.: Trust in virtual teams: a multidisciplinary review and integration. Australas. J. Inf. Syst. **23**, 1–16 (2019)

Jarvenpaa, S.L., Knoll, K., Leidner, D.E.: Is anybody out there? antecedents of trust in global virtual teams. J. Manag. Inf. Syst. **14**, 29–64 (1998)

McKnight, D.H., Chervany, N.L.: Trust and distrust definitions: One bite at a time. Trust Cybersocieties **14**, 27–54 (2001)
Handy, C.: Trust and the virtual organization. Long Range Plan. **28**, 126–129 (1995)
Pearlson, K.E., Saunders, C.S., Galletta, D.F.: Managing and using information systems: a strategic approach. John Wiley & Sons (2019)
Marlow, S.L., Lacerenza, C.N., Salas, E.: Communication in virtual teams: a conceptual framework and research agenda. Hum. Resour. Manag. Rev. **27**, 575–589 (2017)
Ferrell, J., Kline, K.: Facilitating trust and communication in virtual teams. People Strat. **41**, 30–36 (2018)
Ravalet, E., Rérat, P.: Teleworking: decreasing mobility or increasing tolerance of commuting distances? Built Environ. **45**, 582–602 (2019)
Abulibdeh, A.: Can COVID-19 mitigation measures promote telework practices? J. Labor Soc. **23**, 551–576 (2020)
Baruch, Y.: The status of research on teleworking and an agenda for future research. Int. J. Manag. Rev. **3**, 113–129 (2001)
Rodríguez-Modroño, P., López-Igual, P.: Job quality and work—life balance of teleworkers. Int. J. Environ. Res. Public Health **18**, 3239–3240 (2021)
Okubo, T.: Telework in the Spread of COVID-19. Inf. Econ. Policy, p. 100987 (2022)
Kowalski, K.B., Swanson, J.A.: Critical success factors in developing teleworking programs. Benchmarking Int. J. **12**, 236–249 (2005)
Abulibdeh, A.: Can COVID-19 mitigation measures promote telework practices? J. Labor Soc. **23**, 551–576 (2020)
Ionescu, C.A., Fülöp, M.T., Topor, D.I., et al.: Sustainability analysis, implications, and effects of the teleworking system in romania. Sustainability **14**, 5273 (2021)
Karia, N., Asaari, M.: Innovation capability: the impact of teleworking on sustainable competitive advantage. Int. J. Technol. Policy Manage. **16**, 181–194 (2016)
Bentley, T.A., Teo, S.T.T., McLeod, L., et al.: The role of organisational support in teleworker wellbeing: a socio-technical systems approach. Appl. Ergon. **52**, 207–215 (2016)
Smith, S.A., Patmos, A., Pitts, M.J.: Communication and teleworking: a study of communication channel satisfaction, personality, and job satisfaction for teleworking employees. Int. J. Bus. Commun. **55**, 44–68 (2018)
Patrutiu-Baltes, L.: The impact of digitalization on business communication. SEA–Practical Appl. Sci. **4**, 319–325 (2016)

# Study on Food Safety Supervision Countermeasure of Agricultural Products Wholesale Market

Sanyang Gu, Zhen Ding, and Huichun Che(✉)

School of Economics, Beijing Wuzi University, Beijing, China
chehuichun@126.com

**Abstract.** Despite the continuous expansion of various emerging distribution channels, the agricultural products wholesale market is still the main channel for the distribution of agricultural products in China. But the agricultural products wholesale market is also the prone and hardest hit area for food safety problems in the circulation of agricultural products. The article first reviews the relevant research of other scholars, then analyzes and summarizes the current problems of food safety supervision in the agricultural products wholesale market in China, and constructs a two-level game model of the government department to the wholesale market and the wholesale market to the wholesaler. We find that strengthening the supervision and punishment of the government department and the wholesale market, and reducing supervision costs will improve the level of food safety in agricultural products wholesale markets. Finally, based on the actual situation, the article puts forward policy recommendations on food safety supervision in agricultural products wholesale markets.

**Keywords:** agricultural products wholesale market · food safety supervision · game analysis

## 1 Introduction

With the rapid development of the economy and the continuous improvement of people's living standards, people's requirements for food safety are also increasing. As the most frequent daily consumer goods for residents, agricultural products are related to the health and well-being of people. Food safety is an important topic that we have to pay attention to. The upstream of the agricultural products wholesale market is the vast number of farmers, agricultural cooperatives, and agriculture-related enterprises, and the downstream is secondary wholesalers, fresh supermarkets, catering enterprises, and final consumers, forming a wholesale-based agricultural product circulation supply chain. The agricultural products wholesale market undertakes the distribution of more than 70% of fresh agricultural products in urban and rural areas and is still the main channel and central link for the distribution of agricultural products in China.

The article was funded by the Beijing Social Science Fund Planning Project "Research on the Transformation and Upgrading of Beijing Agricultural Products Wholesale Market for Food Safety Demands in the Post-epidemic Era". The grant number of the funding is 20JJC030.


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 10–19, 2023.
https://doi.org/10.1007/978-981-99-3618-2_2

For the agricultural products wholesale market, which is an important gripper of food safety supervision in the circulation of agricultural products, the government has introduced many food safety supervision systems and measures, which have also received good supervision effects. But in reality, there are still many problems in the supervision of food safety in the wholesale market of agricultural products, such as the inadequate guarantee of the quality and safety of agricultural products in the market, the insufficient supervision by relevant government departments, and the neglect of supervision to wholesalers by the wholesale market of agricultural products [1, 2]. In addition, the wholesale market of agricultural products has always been a prone place and disaster area for food safety problems in the circulation of agricultural products. The outbreak of COVID-19 in the Huanan seafood market in early 2020 and the recurrence of the epidemic in the Beijing Xinfadi agricultural products market wholesale have exposed many problems in food safety supervision in the wholesale agricultural products market. As a distribution center of agricultural products, the wholesale market of agricultural products has also become a stark fault line in the prevention and control of the epidemic.

In the post-epidemic era, the transformation and upgrading of China's agricultural products wholesale market have become more and more urgent. Many scholars have made a lot of discussions on this, the construction of the public welfare agricultural products wholesale market system should be highlighted [3], and the development of e-commercialization, digitalization, and intellectualization should become an important path for the transformation and upgrading of China's agricultural products wholesale market [4–6], the five general principles of the world agricultural products wholesale market should become the main direction of the transformation and upgrading of China's agricultural products wholesale market [7]. In addition, some scholars have also paid attention to the food safety issues in the agricultural products wholesale market [8, 9], they all have summarized the causes through phenomena and given policy recommendations, but they lacked further discussions.

Moreover, food safety involves the interests of the government department, the agricultural products wholesale market, and the on-site wholesaler. The strategic choice of any party will affect the utility of the other two parties, and game theory is one of the tools to study this behavioral interaction. In view of this, on the basis of previous scholars' research, the article discusses two-level supervision of the government to the wholesale market and the wholesale market to the wholesaler from the perspective of food safety, analyzes the relationship between different stakeholders, then proposes regulatory countermeasures to this. So as to provide theoretical and practical support for the transformation and upgrading of the agricultural products wholesale market.

## 2 Current Problems of Food Safety Supervision

### 2.1 Poor Food Safety Awareness

As the major trader in the agricultural products wholesale market, every move of agricultural product wholesalers is related to product quality and safety. However, in reality, most wholesalers do not have enough sense of responsibility, they do not have a systematic understanding of the laws, regulations, and market operation guidelines on the

quality and safety of agricultural products. When wholesale market managers sign relevant agreements with them, wholesalers will only sign truthfully and have a one-sided understanding of the contents of the documents. Besides, some consumers are concerned about food safety issues and have a strong sense of self-protection, but they do not know enough about the relevant systems of agricultural product quality and safety, and they lack the ability to identify when purchasing.

### 2.2 Imperfect Quality and Safety Management System

The proprietor of the wholesale market will obtain income by charging intermediary fees, and more agricultural products entering the wholesale market can charge more intermediary fees. Therefore, they will think of various ways to approve the entry of agricultural products into the market, and through simple random inspections, the agricultural products will be eligible to enter the market, so that it is easy to allow agricultural products with substandard quality to enter the market. As a result, there is a phenomenon that products can enter the wholesale market as long as they pay a certain fee in the wholesale market, which makes it difficult to implement the market access criterion. In addition, the standards for the origin of agricultural products are relatively lagging, many agricultural products do not have origin labels, and the certification standards are also not complete enough. At present, China's agricultural product quality and safety management system is still a relatively lagging system and needs to be further improved.

### 2.3 Irregular Market Order

Market managers have the responsibility to maintain the order of market operations, supervise wholesalers, guarantee the quality of agricultural products, and provide high-quality services. The transaction method of agricultural wholesale markets is still the backward model of "focusing on fees and ignoring services". Market managers are not well aware of management and service, and fail to perform their corresponding responsibilities. Due to manpower, material resources, and funding issues, market managers are unable to inspect all agricultural products entering the market one by one, and do not pay enough attention to routine inspections such as commodity incoming inspection and unqualified commodity removal. In addition, many employees lack training, and some middle-level managers do not have professional management skills. The internal management of the market is not scientific, and the market management order is not standardized.

### 2.4 Low Efficiency of Government Regulation

At present, food supervision in China is a typical mode of segmented supervision. The supervision model is decentralized and there are multiple managements. The government departments of agriculture, industry, commerce, and quality inspection all have the power to manage the agricultural products wholesale market. The overlapping of functions of various departments results in unclear responsibilities, difficulty in coordination, and difficulty in conveying and implementing decrees. The scope of responsibilities of

each department is not clear and intersects with each other, so it is difficult to form a management mechanism for unified decision-making and efficient operation. Inspectors of government departments often fail to do anything in market management and go through the motions in sampling inspection of agricultural products. The test results are not publicized, so the information is easily falsified and the inspection is often a mere formality. The government does not pay enough attention to the wholesale market of agricultural products, and the policy is not unified. The responsibilities of the management department are unclear, resulting in redundant supervision.

## 3 Game Analysis of Interest Subject

The government department, the wholesale market, and the wholesaler are the relevant stakeholders in the food safety supervision system of agricultural wholesale markets. The government department is the primary regulator and plays a major role in monitoring the quality and safety of agricultural products in wholesale markets. The wholesale market plays an intermediary role. On the one hand, it accepts the supervision of the government department, on the other hand, it undertakes to supervise various operating activities and the behavior of wholesalers in the wholesale market. Wholesalers are suppliers of agricultural products in the wholesale market and are subject to joint supervision from the government department and the wholesale market.

### 3.1 Game Analysis Between Government Department and Wholesale Market

In economic activities, the government requires wholesale markets to supervise wholesalers in strict accordance with legal requirements, strictly test and inspect agricultural products entering the market, check business licenses, health licenses, etc. And punish merchants who provide substandard agricultural products. However, in reality, to maximize their interests, the wholesale market ignores the supervision of wholesalers and lingers in the behavior of following the requirements of government departments to perform supervision functions and not supervising wholesalers for the sake of interests. Therefore, the game between government departments and wholesale markets arises.

*1) Model setup*

Assumption 1: There are two participants in the model, the government department and the wholesale market, who are rational economic persons.

Assumption 2: There are two options for the government department to supervise or not supervise the wholesale market, and the wholesale market also has two options of supervising or not supervising the wholesaler. We set that $P_1$ is the probability that the government department supervises the wholesale market, and $P_2$ is the probability that the wholesale market supervises the wholesalers.

Assumption 3: We set that $C_1$ is the supervision cost saved when the wholesale market does not regulate the wholesaler, $I_1$ is the additional income obtained when the wholesale market does not regulate the wholesaler, and $C_2$ is the supervision cost saved by the government department when it doesn't supervise the wholesale market, $I_2$ is the loss of government reputation caused by agricultural products with substandard quality circulating in the market, $I_3$ is the loss of future income of the wholesale market caused

by agricultural products with substandard quality, and $I_4$ is the fine for the wholesale market not to regulate the wholesaler when the government department regulates the wholesale market. $E_1$ and $E_2$ denote the expected utility functions of the government department and the wholesale market, respectively.

We can easily get the game payment matrix between the government department and the wholesale market (Table 1). For simplicity and without loss of generality, the payoffs of the game subjects in the game payoff matrix are all net gains.

**Table 1.** Game payment matrix

| | | Wholesale market | |
|---|---|---|---|
| | | Supervise | Not to supervise |
| Government department | ***Supervise*** | $(0, 0)$ | $(I_4\text{-}I_2, I_1 + C_1\text{-}I_3\text{-}I_4)$ |
| | ***Not to supervise*** | $(C_2, 0)$ | $(C_2\text{-}I_2, I_1 + C_1\text{-}I_3)$ |

*2) Model solving*

$$E_1 = P_1P_20 + P_1(1 - P_2)(I_4 - I_2) + (1 - P_1)P_2C_2 + (1 - P_1)(1 - P_2)(C_2 - I_2) \tag{1}$$

$$\begin{aligned} E_2 = {} & P_2P_10 + P_2(1 - P_1)0 + (1 - P_2)P_1(C_1 + I_1 - I_3 - I_4) \\ & + (1 - P_2)(1 - P_1)(C_1 + I_1 - I_3) \end{aligned} \tag{2}$$

Find the partial derivative of the above utility function:

$$\frac{\partial E_1}{\partial P_1} = I_4 - P_2I_4 - C_2 = 0 \tag{3}$$

$$\frac{\partial E_2}{\partial P_2} = I_3 + P_1I_4 - I_1 - C_1 = 0 \tag{4}$$

$$\overline{P}_1 = \frac{C_1 + I_1 - I_3}{I_4} \tag{5}$$

$$\overline{P}_2 = 1 - \frac{C_2}{I_4} \tag{6}$$

*3) Conclusion*

According to $0 < \overline{P_1} < 1, 0 < \overline{P_2} < 1$, we get $I_3 < I_1 + C_1 < I_3 + I_4$, that is, the sum of the additional benefits obtained by the inadequate supervision of the wholesale market to the wholesales and the saved supervision cost is more than the future income loss of the wholesale market caused by the substandard quality of agricultural products. And it is less than the sum of the fine imposed by the government on the wholesale market and the future income loss of the wholesale market caused by the agricultural products that do not reach the quality standards; $I_4 > C_2 > C_2 - I_2$, that is to say, the fine imposed by the government on the wholesale market is greater than the difference

between the regulatory cost saved when the government department does not supervise and the reputation loss caused by the substandard products to the government.

*a) Influencing factors of government supervising the wholesale market:* According to $\frac{\partial \overline{P_1}}{\partial C_1} = \frac{1}{I_4} > 0$, $\frac{\partial \overline{P_1}}{\partial I_1} = \frac{1}{I_4} > 0$, we know the probability of the government department's supervision will increase with the increase of the regulatory cost saved by the no supervision of the wholesale market or the additional benefits obtained from the no supervision of the wholesale market. According to $\frac{\partial \overline{P_1}}{\partial I_3} = -\frac{1}{I_4} < 0$, $\frac{\partial \overline{P_1}}{\partial I_4} = -\frac{C_1+I_1-I_3}{{I_4}^2} < 0$, we can see the probability of government supervision will increase with the decrease of fines imposed by the government on wholesale markets or the decrease of future revenue losses caused by inadequate wholesale market supervision.

*b) Influencing factors of wholesale market supervising the wholesalers:* According to $\frac{\partial \overline{P_2}}{\partial C_2} = -\frac{1}{I_4} < 0$, the probability of the wholesale market supervising the wholesaler will increase with the decrease of regulatory costs saved by inadequate government supervision. According to, $\frac{\partial \overline{P_2}}{\partial I_4} = \frac{C_2}{(I_4)^2} > 0$, the probability of wholesale market supervision will increase with the increase of the government fines imposed on the wholesale market.

### 3.2 Game Analysis Between Wholesale Market and Wholesaler

Wholesalers are intermediate traders of agricultural products and the middle of ensuring the quality and safety of agricultural products in wholesale markets. However, in the absence of supervision, wholesalers often provide agricultural products with substandard quality to maximize their profits. While the wholesale market supervises wholesalers, it must also ensure the maximization of its interests. Therefore, the game between the agricultural products wholesale market and the wholesaler arises.

*1) Model setup*

Assumption 1: The model has two participants, the wholesale market and the wholesaler, who are rational economic persons.

Assumption 2: The wholesale market has behaviors that have adequate and inadequate supervision to the wholesaler, and the wholesaler has behaviors of providing agricultural products with up to the standard quality or providing agricultural products with substandard quality. We set that $P_3$ is the probability that the wholesaler provides agricultural products of up to standard quality.

Assumption 3: We set that $I_5$ is the additional income obtained by the wholesaler from providing agricultural products with substandard quality, $I_6$ is the loss of future income caused by providing agricultural products with substandard quality for the wholesaler, and $I_7$ is the penalty for the wholesaler providing substandard agricultural products when the wholesale market supervises the wholesaler. $E_3$ represents the expected utility function of the wholesaler.

We can easily get the game payment matrix between the wholesale market and the wholesaler (Table 2). For simplicity and without loss of generality, the payoffs of the game subjects in the game payoff matrix are also all net gains.

*2) Model solving*

$$E_2 = P_2P_30 + P_2(1-P_3)I_7 + (1-P_2)P_3(C_1+I_1) + (1-P_2)(1-P_3)(C_1+I_1-I_3-I_4) \quad (7)$$

**Table 2.** Game payment matrix

| | | Wholesaler | |
|---|---|---|---|
| | | Guarantee quality | Not to guarantee quality |
| Wholesale market | ***Supervise*** | $(0, 0)$ | $(I_7, I_5\text{-}I_6\text{-}I_7)$ |
| | ***Not to supervise*** | $(C_1 + I_1, 0)$ | $(C_1 + I_1\text{-}I_3\text{-}I_4, I_5\text{-}I_6)$ |

$$E_3 = P_3P_20 + P_3(1 - P_2)0 + (1 - P_3)P_2(I_5 - I_6 - I_7) + (1 - P_3)(1 - P_2)(I_5 - I_6) \tag{8}$$

Find the partial derivative of the above utility function:

$$\frac{\partial E_2}{\partial P_2} = I_3 + I_4 + I_7 - C_1 - I_1 - P_3(I_3 + I_4 + I_7) = 0 \tag{9}$$

$$\frac{\partial E_3}{\partial P_3} = P_2I_7 - I_6 - I_5 = 0 \tag{10}$$

$$\overline{P}_2 = \frac{I_5 - I_6}{I_7} \tag{11}$$

$$\overline{P}_3 = 1 - \frac{C_1 + I_1}{I_3 + I_4 + I_7} \tag{12}$$

*3) Conclusion*

According to $0 < \overline{P_2} < 1$, $0 < \overline{P_3} < 1$, we get $I_6 < I_5 < I_6 + I_7$, that is, the additional income obtained by the wholesaler from providing agricultural products with substandard quality is more than the future loss of income caused by it, and is less than the sum of the future loss caused by the wholesaler providing agricultural products with substandard quality and the penalty imposed by the wholesale market on the wholesaler; $I_7 > C_1 + I_1 - I_3 - I_4$, that is, the penalty imposed on wholesalers by the wholesale market is greater than the sum of the saved supervision cost and the additional income obtained by the inadequate supervision of the wholesale market minus the sum of the loss of future income of the wholesale market and the penalty imposed by the government on the wholesale market.

*a) Factors influencing wholesale market supervise the wholesalers:* According to $\frac{\partial \overline{P_2}}{\partial I_5} = \frac{1}{I_7} > 0$, we know the probability of wholesale market supervision will increase with the increase in the additional income obtained by wholesalers from providing agricultural products of substandard quality. According to $\frac{\partial \overline{P_2}}{\partial I_7} = -\frac{I_5 - I_6}{I_7{}^2} < 0$, $\frac{\partial \overline{P_2}}{\partial I_6} = -\frac{1}{I_7} < 0$, the probability of wholesale market supervision will increase with the decrease of future revenue losses caused by the wholesaler providing substandard quality agricultural products or the reduction of wholesale market penalties imposed on wholesalers.

*b) Factors influencing wholesalers to provide quality agricultural products:* According to $\frac{\partial \overline{P_3}}{\partial C_1} = -\frac{1}{I_3 + I_4 + I_7} < 0$, $\frac{\partial \overline{P_3}}{\partial I_1} = -\frac{1}{I_3 + I_4 + I_7} < 0$, the probability of the wholesaler

providing quality agricultural products will increase with the decrease of saving regulatory costs and additional benefits by inadequate supervision of the wholesale market. According to $\frac{\partial \overline{P_3}}{\partial I_3} = \frac{C_1+I_1}{(I_3+I_4+I_7)^2} > 0$, $\frac{\partial \overline{P_3}}{\partial I_4} = \frac{C_1+I_1}{(I_3+I_4+I_7)^2} > 0$, $\frac{\partial \overline{P_3}}{\partial I_7} = \frac{C_1+I_1}{(I_3+I_4+I_7)^2} > 0$, we can find the probability of the wholesaler providing quality agricultural products will increase with the increase of future revenue loss when the wholesale market supervises the wholesaler or increase with the increase of fines imposed by the government on the wholesale market or the wholesale market imposed on the wholesaler.

## 4 Policy Suggestions

### 4.1 Countermeasures at Government Level

*1) Strengthen the efficiency of government supervision*

The government has an obligatory task of supervising the quality and safety of agricultural products in wholesale markets. It should actively play a leading role, integrate the management functions of various departments, clarify the responsibilities and tasks of each department, perform their duties, and form joint participation in the quality and safety of agricultural products. In the case of respecting the laws of the market, strengthening the supervision of the market, promoting the effect of quality and safety linkage, reduce the supervision cost of government departments. At the same time, it cannot use administrative means to destroy the law of the survival of the fittest in the market itself.

2) *Perfecting relevant laws and statutes*

The wholesale market is an important regulator in the circulation of agricultural products. The affirmation of the main body status of the wholesale market by national laws is conducive to the sound development of the related work in the wholesale market. The healthy development of the wholesale market of agricultural products is closely related to the intervention of the government, and the government's policy and financial support. Although China has formed a regulation system for the wholesale market of agricultural products, there is a lack of targeted law on the wholesale market of agricultural products, so the government should keep up with the legislative rhythm as soon as possible. In this way, the managers and wholesalers of the wholesale market are forced to do their part and work together for the orderly development of the wholesale market.

*3) Establish a feasible incentive and subsidy mechanism*

Due to the high cost of supervision, the wholesale market is unwilling to bear the huge expenditure alone and will relax supervision, so as to cause the behavior of traders on the floor that disrupt the market order. The government should increase the penalties for deregulation and inaction in the wholesale market. The cost subsidies for testing equipment and testing personnel in wholesale markets can be increased. Some agricultural products with excessive testing costs, use costs, quality certification costs, and label management costs should be included in the scope of subsidies, so the subsidy system for agricultural product quality and safety can be gradually improved. The government can also reward and reduce taxes and fees for producers of pollution-free and green agricultural products, and severely crack down and punish those who produce low-quality agricultural products.

### 4.2 Countermeasures at Wholesale Market Level

*1) Enhance the level of market management.*

*a) Improve the market access system and standardize product identification management*: The agricultural products entering the wholesale market should be registered, and the detailed information of each agricultural product must be marked, such as name, origin, wholesaler, business license, health certificate, etc. Guide wholesalers to use electronic data to upload, register, and enter specific information on agricultural products online, and archive them in relevant departments, so as to obtain market entry qualifications for agricultural products. Agricultural product labeling management is not sophisticated enough to cover all agricultural products entering the market in China. The coverage of agricultural product labels should be expanded to achieve unified label management for all agricultural products. Improve the certification standards, so that all agricultural products that lack certification and registration information are not allowed to enter the market for sale.

*b) Improve the information disclosure mechanism and share product information:* When wholesalers supply agricultural products with quality standards, they need to spend more manpower, material resources, and financial resources, so the cost increases, but they do not get enough income in the sales process to offset the cost. Wholesalers will reduce the quality and quality of agricultural products for their own interests. The wholesale market can establish an information disclosure mechanism, so this can release all information on agricultural products in the market promptly, push high-quality agricultural products that meet the quality standards to consumers, promote the long-term connection between wholesalers and consumers, and match consumers with high-quality and affordable high-quality agricultural products. It can also help wholesalers to find the buyer of the goods and increase the wholesaler's revenue, to continue to maintain the supply of high-quality agricultural products. Reducing information asymmetry through the information disclosure mechanism and coordinating the allocation of market resources will increase the market share of high-quality agricultural products.

*2) Build a testing and inspection system and improve the agricultural product traceability system*

*a) Strengthen team building and improve testing and inspection level*: The wholesale market should actively cover agricultural product testing institutions with CMA and CATL certification, equipped with corresponding testing rooms and advanced equipment, increase the testing samples of agricultural products, expand the testing scope, and improve the testing coverage rate, to ensure the accuracy of agricultural product quality testing. At the same time, talent is the top priority, strengthen the construction of the talent team of inspection personnel, mobilize all positive factors, carry out professional knowledge training, and improve the professional knowledge and skills of each quality inspection personnel, to promote the technical progress and efficiency improvement of agricultural product inspection. Build an information-sharing platform for product testing in the market. Through the platform, testing personnel can inquire about the testing process of agricultural products, to improve testing efficiency and save testing costs.

*b) Build an information management platform and perfect the agricultural product traceability system:* According to the requirements of agricultural regulations, the

wholesale market should actively follow up to build the information management platform for agricultural product traceability and perfect the agricultural product traceability system. Establish a product information registration system, monitor and record each workflow from production to sales in detail, and form integrated tracking management of agricultural product information. It will connect the information traceability platform with the market access link, realize data information sharing and resource sharing, and integrate all elements of testing and inspection, such as origin certification, early warning assessment, law enforcement supervision, traceability, and product standards. Relevant departments should conduct strict monitoring and management to achieve real-time connection and tracking of the traceability system. Consumers can also view all the information in the entire production process, including production origin, transportation inspection, and finally the entire process information reaching the customer through the product identification.

## References

1. Zheng, C., Li, X., Zhang, X.: Problems and countermeasures in the construction of China's agricultural modern circulation system. Econ. Rev. J. **04**, 125–128 (2018)
2. Liu, Y.: Analysis on the problems and countermeasures of the quality and safety supervision of agricultural products in China. Agric. Econ. **04**, 136–138 (2020)
3. Cao, B., Yu, R.: Promoting the construction of public welfare agricultural products wholesale market system: Japan's experience and enlightenment. Acad. Res. **12**, 97–104 (2021)
4. Zhao, D.: Research on the intelligent transformation of agricultural products wholesale market. Agric. Econ. **09**, 141–142 (2021)
5. Zeng, Y.: Intelligent upgrade of China's agricultural products wholesale market: Rethinking under the warning of the new crown epidemic. J. Commercial Econ. **23**, 5–8 (2020)
6. Ren, W.: Discussion on the transformation and modernization of China's agricultural products wholesale market. J. Commercial Econ. **24**, 127–130 (2019)
7. Chen, X., Zhang, Z., Zhang, X.: Research on the mode and path of China's agricultural products wholesale market upgrading: based on the experience of five general principles of world agricultural products wholesale market. China Business Market **33**(02), 30–37 (2019)
8. Yan, L., Dong, X., Cui, Y., et al.: Current situation and development suggestions of agricultural product quality safety supervision in Liaoning province. Agric. Econ. **08**, 59–60 (2021)
9. Wang, X.: The role of agricultural products wholesale markets in the quality and safety management of edible agricultural products. China Development Observation **24**, 51–53+43 (2020)

# Blockchain Deposition and Copyright Protection Method for Self-Published Content Infringement

Keke Qu[1] and Liang Wang[2(✉)]

[1] School of Journalism and Publishing, Beijing Institute of Graphic Communication, Beijing, China
1352249088@qq.com

[2] School of Economics and Management, Beijing Institute of Graphic Communication, Beijing, China
wangliang@bigc.edu.cn

**Abstract.** Since the web 3.0 era, everyone can be the source of information dissemination, but with the gradual saturation of the self-publishing market, the video style, copy content, and scene layout of individual creators are directly "copied and pasted" by infringers, and the infringement phenomenon is increasing year by year. Based on the analysis of the characteristics of blockchain technology, this paper aims to analyze the difficulties encountered by self-publishing copyright, identify blockchain as the key to copyright storage and protection, and finally reflect on blockchain technology, which will bring challenges to self-publishing dissemination and regulation.

**Keyword:** self-media · blockchain · infringement · copyright protection

## 1 Introduction

According to a report by Punch News on January 13, 2022, the report points out that the volume of cases of self-media and new media being sued for copyright infringement has been rising year by year. The report argues that Internet service models continue to innovate, the risk of online infringement is growing, and copyright protection faces new situations and problems in the Internet era. According to the report, from January 1, 2020 to December 31, 2021, the Beijing Four Central Courts received a total of 1,566 Internet civil and commercial appeal cases, of which 602 Internet civil and commercial appeal cases were received in 2020, up 86.96% compared with 322 cases in 2019; 964 Internet civil and commercial appeal cases were received in 2021, up 60.13%.

From the case intake network infringement liability disputes accounted for 859 cases, accounting for 59.9%, compared with 2019, the number of cases received in network infringement liability disputes rose significantly, and from the perspective of the subject of the lawsuit, the amount of cases in which self-media and new media are sued for infringement has been climbing year by year. This shows that the issue of self-media copyright protection needs to be paid attention to urgently.

This work is partially supported by the Key Program of Beijing Social Science Fund under contract with No. 19JDXCA001.


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 20–29, 2023.
https://doi.org/10.1007/978-981-99-3618-2_3

## 2 Concept Clarity: Blockchain and Technical Features

Blockchain is a globally distributed, collaborative database storage system that is largely different from traditional database operations. With blockchain, anyone who has the ability to set up server nodes can participate in it. In this way, people from all over the world deploy their own nodes locally and connect to the blockchain network to become a node in this distributed database storage system; once they join the blockchain network, the node enjoys exactly the same rights and obligations as all other nodes. At the same time, for those who carry out services on the blockchain, they can read and write operations at any node in this system, and finally all the nodes around the world will complete synchronization again and again according to some mechanism, so as to realize the data of all the nodes in the blockchain network are exactly the same.

Currently, the Internet is booming and the traditional database operation is no longer suitable for the current decentralized industrialized ecological layout, therefore, the consensus mechanism of distributed nodes of blockchain is born. In addition, its decentralization, openness, traceability, immutability and anonymity also determine the future operation mode of "blockchain + self-media" will greatly solve the current dilemma of self-media infringement cases.

### 2.1 Decentralization

The biggest feature of blockchain is "decentralization". Decentralization means that all the operations of the actors are laid out on the distributed ledger, and no longer on the servers of centralized institutions or third-party organizations. In layman's terms, decentralization means that no third-party access is required, and anyone can access the entire database and its complete operation history, truly realizing person-to-person, peer-to-peer transactions and interactions.

### 2.2 Openness

Since blockchain is decentralized, all network nodes can participate in the record maintenance of the blockchain network data, which requires the blockchain network to be open. At the same time, the blockchain network system is open and transparent, and the data on the chain is transparent to all nodes of the network except for the private information of the transaction parties which is encrypted by the hash algorithm. Anyone or any participating node can query the data records of the blockchain through the public interface, which is the basis for the blockchain system to be trusted.

### 2.3 Traceability

The blockchain itself is a block-chain data structure, and the information on the chain is interlocked according to the chronological order, which makes the blockchain have the feature of traceability. When applied to life, it means that all information about the cultivation, production, transportation, sales and supervision of products are recorded on the blockchain. Once any problem occurs, it is possible to trace the source from the beginning and check each link in order to ensure the safety of the product.

### 2.4 Irreversible

Blockchain can be tamper-proof or very costly to tamper with because it uses cryptography to upload data, and the latter block contains the timestamp of the previous block and is sorted in chronological order. The irreversible feature means that once the data is written into the database of the blockchain, no one can easily change the data information without permission.

### 2.5 Anonymity

If decentralization is an important foundation of blockchain, then anonymity is an important reason for many people to choose blockchain. Blockchain uses cryptographic means such as hashing algorithms, asymmetric encryption, and private public keys to protect the privacy of individual nodes' transactions while achieving complete transparency of data.

However, the anonymity of blockchain has been repeatedly questioned because some unlawful elements use blockchain to carry out illegal acts such as money laundering and asset theft, but due to the anonymity of blockchain, the identity information of the unlawful elements cannot be known only through IP addresses, thus causing regulatory problems.

## 3 Copyright Dilemma: Self-Media Struggles to Maintain Rights

From blogs at the beginning of the 21st century to the launch of Sina Weibo in 2009, self-media has completed its original accumulation, and the introduction of WeChat in 2011 shifted the main battlefield of self-media to cell phones and other mobile terminals. In order to grab the first wave of traffic bonus, the major platforms have changed from "content is king" to "traffic is king", disregarding the copyright of the original creator's work, and the number of self-media infringement cases is increasing year by year.

### 3.1 Difficult to Confirm the Rights of Original Works

Self-media creators, also known as user-produced content or professional user-produced content, publish their works on third-party platforms, including written works, photo works and video works. With the gradual saturation of the self-media market, plagiarism, plagiarism, and illegal transportation have emerged one after another. Currently, the issue of copyright protection of original works is of great concern to all parties, and the scope of works covered is quite controversial. Some scholars believe that short videos in video works cannot be classified as works in the real sense because they are too short in length, which not only express limited meanings but also lower the character of the works; other scholars believe that no matter it is a written work, a picture work or a video work, as long as the work contains the original creator's originality and the emotion it wants to convey, it can constitute a work. From the perspective of legal elements, the Copyright Law stipulates that the composition of a work must meet three conditions: first, it comes from the field of literature, art or science and technology; second, it has a certain degree of

creativity; and third, it can be reproduced in tangible form. In addition, the revision of the Copyright Law in 2021 changed the term “cinematographic and electrotechnical works” to “audiovisual works”, expanding the scope of “cinematographic and electrotechnical works”, and short videos may may be classified as one of them [1].

In recent years, the competition among self-publishing creators active in major platforms has gradually become fierce, and the information and resources between different platforms are not synchronized, resulting in the lack of mutual consensus on common rights. In addition, both users and authors are faced with the overwhelming amount of information every day, and if the original creator has a relatively small fan base, he or she is likely to be unaware of the infringement even if it is “infringed”. Therefore, cross-platform infringement is very common. For the original creators, the biggest difficulty in terms of copyright is to confirm the right, that is, to determine the copyright of the work belongs to them. However, due to the current cumbersome and time-consuming process of confirming rights, in many cases, authors are often reluctant to assert their rights. In addition, the maturity of digital technology makes the uploading, reproduction and deletion of works a one-click operation, and the form of creation, creativity and content of works are easy to copy and paste, based on which the rights of original creators are “stuck in the mud”.

### 3.2 Infringement Occurred to Maintain the Rights of the Difficult

At present, infringement cases in China’s self-media market are frequent and difficult to defend rights mainly because traditional technical means are difficult to monitor. For textual works, the check condition for academic papers is determined as “duplicate” if 12 consecutive characters are the same, but except for academic articles, the monitoring condition for other textual works is not too harsh. With the development of Internet technology, the use of digital processing technology, convenient and efficient text processors provide a convenient way to modify the content of the text, and therefore also gave birth to the technical basis of “manuscript washing”. Moreover, it is difficult to leave any trace of the content modified by using text processors [2].

For picture works, the self-publishing author will convey novel ideas and inner feelings to the audience through pictures, although the author enjoys the copyright of the pictures, but the creativity and emotions are often subjective, so it is difficult for the platform machine to determine the ownership of the works, and if manual audit is used, it will also lead to high monitoring costs or unfairness caused by the subjective factors of the auditor and other consequences. For video works, video works can be simply divided into long videos and short videos, long videos are currently monitored using keyword extraction and video fingerprinting technology, which will not only compare frame by frame but also the name of the video, but this technology is only applicable to long videos and not short videos. First of all, short video editing, production is relatively simple, the author named randomly, and sometimes not even named, often "online surfing" bloggers will be more familiar with the major platforms of sensitive words, using pinyin, harmonics, symbols instead of the way to circumvent the platform’s technical monitoring. In addition, short video works are short, some even just a few seconds, the use of frame-by-frame comparison of technology for short video can be little room to play.

Whether it is a written work, a picture work or a video work, self-published authors still face the dilemma of difficulty in preserving evidence of infringement and the high cost of defending their rights. The decentralized nature of the Internet also determines the trend of unstructured communication, and the explosive growth of information makes it difficult for authors to determine whether the infringers are infringing, and even if they happen to find that their works are infringing, the infringers are likely to remove the infringing works quickly, and once the infringing works are removed, it is even more difficult for authors to obtain evidence. The dilemma of the high cost of defending the rights of self-media authors mainly lies in the fact that China does not have an online platform for defending rights, and currently the infringer can only take the form of litigation, while the self-media creator has limited funds and time, even if the court awards compensation, the amount obtained is difficult to cover the time and labor costs during the litigation period, and for the self-media creator, it is not worth the loss.

### 3.3 The Platform Identifies the Difficulty of Supervision

The major self-media platforms, as third-party institutions, have the obligation and responsibility to restrain every original creator of self-media and create a healthy and virtuous online community for users. However, there are many difficulties for the platform to determine the regulation. The first is reflected in the changing status of the self-publishing platform between content providers and technical service providers. When the platform only provides video storage services and video connectivity services, it belongs to the technical service providers; but when it provides content to users itself, it may become a content service provider [1]. The reason why platforms are clearly defined as either technology service providers or content service providers is that they straddle the "safe harbor rule" between the two. The "safe harbor principle" refers to the fact that in the event of copyright infringement cases, when an ISP only provides space services and does not produce web content, the ISP has the obligation to remove it if it is informed of the infringement, otherwise it is deemed to be infringing. If the infringing content is neither stored on the ISP's server nor informed which content should be removed, the ISP is not liable for infringement. Thus, the definition of the legal status of the platform also maps out the obligations it is supposed to fulfill.

In addition, although the platform will actively manage, but in fact, for the platform within the infringement, will only play the role of "notice – delete", the infringer and the infringer belong to the same platform, the platform as a third party is difficult to have a position to As a third party, it is difficult for the platform to "dispose" of a certain party. In the case of cross-platform infringement, it is even more difficult for the platform to take action. For one thing, the technical level and management ability of different platforms differ, so it is difficult for both platforms to be neutral and reach a consensus when cross-platform infringement cases occur; for another, platforms know that homogeneous video content and text content will make users resentful, but in order to increase station traffic and pop-up buzzwords, they usually turn a blind eye to users' "carrying behavior" Choose to turn a blind eye.

As you can see, whether it is the difficulty of defining the legal identity of the platform or its own self-interest based, it has brought a lot of trouble to the road of defending the rights of self-media creators.

# 4 Application Scenario: Blockchain Becomes the Key to Copyright Deposit and Protection

Because blockchain technology has stronger security, deeper anonymity and data integrity, coupled with the decentralized distributed model does not require a third-party copyright certification body to control the flow of information in it at all. Currently, blockchain technology can also play an irreplaceable role in copyright storage and protection based on the same approach in the field of self-publishing, where information is entered and uploaded to the blockchain in the three stages of confirmation, authorization and maintenance of rights, forming a new "copyright chain". The different attributes of blockchain technology can be used in different contexts to take advantage of the new technology and develop and innovate the management mechanism of the self-media industry.

## 4.1 Blockchain Facilitates Copyright Transactions

Blockchain technology is conducive to copyright deposition mainly because it can integrate all the scattered elements such as courts, parties and copyright certification bodies to form a blockchain consensus chain. In addition to the decentralized distributed model, the technical foundation of blockchain can ensure that the information is transparent in the process of data flow, and each node can realize self-verification, transmission and management of information, and will not be affected by the failure of one node. However, in the past copyright dispute cases, the disputing parties may affect the outcome of the copyright and copyright dispute due to the time, place and untimely communication. When the data uploaded to the blockchain by the original creator meets certain conditions, a smart contract similar to "electronic contract" can be automatically generated, which can play the role of copyright publicity, so that both parties to the transaction can understand the case information and progress at the first time. This way, both parties to the transaction can be informed of the case information and progress at the first time. Once the two parties have realized the cooperation relationship in the legal sense through the contract, it can also ensure the information symmetry relationship between the two parties of the copyright transaction and realize fair, transparent and standardized transactions.

## 4.2 Blockchain Helps to Ensure the Authenticity of Evidence

Hash algorithm is also called secure hash function also known as information summary. As we all know, the article summary is an overview summary of the content of the article, reading the article summary, we can understand most of the meaning of the article. Hashing algorithm also has this function, it can take an arbitrary set of information and describe it with very simple information. Unlike symmetric and asymmetric encryption, the hash function is a fast convergence algorithm that computes from input to output very fast and converges values quickly without consuming huge computational resources. Based on such excellent characteristics, hash functions are widely used.

The hash algorithm, as the technical foundation of the blockchain, plays a pivotal role. The reason why the information in the blockchain cannot be tampered with should

also be mainly attributed to the unidirectional and unique certainty of the hash value. Unidirectional means that the hash value cannot be verified in both directions to derive the original data, but only through a one-way forward operation to derive the data; unique certainty means that each hash value and the original data are the only correspondence, there is no two different data and the same hash value corresponds to the situation. This specific property of hash value can guarantee both the correctness and integrity of data, and also prevent the information from being maliciously tampered. From the time a work is registered and chained in the blockchain, the entire life cycle of the electronic data of that work runs in the chain of the blockchain depository system, and unless the infringer can grasp 51% of the nodes of the entire network of the blockchain, there is no chance that the whole process can be tampered with, so the hash algorithm has irreplaceable merits in the depository of the copyright of a work.

### 4.3 Blockchain Facilitates Clear Copyright Attribution Disputes

In September 2018, the "Provisions of the Supreme People's Court on Several Issues Concerning the Hearing of Cases in the Internet Courts" endorsed the legal effect of blockchain deposition: "The electronic data submitted by the parties can be proved through electronic signature, trusted timestamp, hash value verification, blockchain and other technical means of evidence collection, fixing and anti-tampering, or through authentication by electronic forensic deposition platform its authenticity, the Internet Court shall confirm it."

In the past copyright infringement cases, it was difficult to confirm the correlation between different documents due to the traditional electronic evidence, which led to unclear copyright attribution. There are two main reasons for this: first, as a new industry, the self-media industry has not yet formed a mature operation system, and many self-media workers are individual-based, and their awareness of copyright is not strong. Offline registration is hardly useful. Blockchain uses a full-record approach to store data information, so it can track all the links in the flow [3]. Based on this feature, blockchain technology can achieve the effects of confirming the ownership, anchoring the infringers and identifying the authenticity of copyright, which is the fundamental difference from the traditional copyright protection system.

In June 2018, the Hangzhou Internet Court supported the plaintiff's adoption of blockchain technology as a deposition method and found the infringement facts of the other party in a copyright dispute, and also recognized the legal validity of electronic data deposited with blockchain technology. This is the first blockchain deposition case in China.

## 5 Technology Cold Thought: Blockchain Brings Challenges to Self-Media Communication and Regulation

Neil Bozeman has said: Every technology is both a burden and a gift, not an either result, but a product of both pros and cons. Blockchain as a new technology, while bringing convenience to various fields of society, we should also think about the challenges that blockchain brings to the dissemination and regulation of self-publishing.

### 5.1 Technology and Law: Blockchain and Copyright Boundaries Overlap

One of the highly regarded technical features of blockchain is the immutability of data, however, this technical feature brings impact on the public's rights to writings and privacy. China's Copyright Law stipulates that authors enjoy a series of rights, including the right to copy, the right to modify, and the right to protect the integrity of the work, since the creation is completed [1]. However, in actual cases, authors do not have the right to make a final decision when they create their works. They may have new insights and ideas about their works as their life experiences increase and wish to reflect them in their works, or they may wish to delete their works for certain reasons. In the blockchain, the system is immune to changes in information, and neither human error, data leakage, nor fraud can affect the fact that the blockchain is immutable [4]. Authors who want to change or delete their works are subordinate to the scope of copyright protection, but the immutable technical feature of blockchain deprives citizens of the freedom to obtain the rights to their works, and the boundaries between the two overlap, leaving the rights in a "blind spot".

### 5.2 Algorithmic Black Box: Self-Media Communication is Caught in a "Trust Loop"

In an online world dominated by algorithms, the realistic landscape is being silently reconstructed by data and information. The logic of algorithm operation is opaque and "dark". The algorithmic black box is a metaphor for "opacity": human beings have many rights and obligations, and now, without understanding the logic behind the black box, they give their decisions about rights and obligations to the algorithm without reservation. Frank Pasquale brings this metaphor to life in "The Black Box Society," an attack on the incomprehensible state in which American society is being manipulated by the secret algorithms of the financial and technology industries [5]. The algorithmic black box is often associated with negative comments such as "algorithmic dictatorship" and "algorithmic monopoly," and the future presents a rather bleak picture: the human subjectivity in the real world is beginning to be hollowed out, and algorithms are reaching out from the black box as the masters of the digital world. The algorithm, as the master of the digital world, is extending its invisible hand from the black box to control people's lives [6].

The reason for the popularity of "blockchain+" is that there is a large amount of data and algorithm "backing", forming a distributed bookkeeping that is transparent, non-tamperable and traceable throughout. However, if the original data in the chain is not objective and true enough, once it is recorded in the chain, the "untrue" content will be "certified" as "true", which is contrary to the "trust system" built by blockchain. This is a paradox with the "trust system" built by blockchain [7]. In addition, the algorithm is cold and rational. When it finds a strong correlation between two data, the algorithm machine will determine that there is a causal relationship between them, and this "rational" determination is not accurate and reasonable even though it is supported by data. Therefore, the inaccuracy of the original data and algorithm of blockchain may lead to a secondary crisis in the field it covers.

With the chaos in the self-media market, platform operators have started to introduce blockchain into the communication environment, forming the communication pattern of “blockchain + self-media”. Blockchain has its specific technical foundation as the "backing". If the platform operators are used by "people with a heart" to disseminate remarks that endanger national sovereignty, split national unity and provoke gender confrontation through algorithmic black-box irregularities, it is likely that the public will If the platform operators are used by “people with a heart” to disseminate illegal algorithms that endanger national sovereignty, divide national unity and provoke gender stereotypes, it is likely that the public will fall into a “trust loop” and the self-media market will be a “natural disaster”.

### 5.3 Human-Machine Symbiosis: Blockchain Technology Weakens Humanistic Colors and Deviates from Mainstream Values

In the era of digital twin, new technologies such as artificial intelligence, 5G, Internet of Things and blockchain are continuously applied to the media ecological order, the whole Internet world is dominated by scientific rationalism, and the arrival of humanist crisis has become inevitable. in May 2018, the “right to be forgotten” came into force with the publication of the EU General Regulation on Data Protection. The “right to be forgotten” is officially recognized in the EU in the form of law. The technical characteristics of blockchain, such as precise timestamp, immutability, full traceability, network transparency and permanent storage, are obviously contrary to the citizens’ “right to be forgotten”. The existence of blockchain technology not only permanently records the “black history” of individual users, but also erases their chances of reformation, so that they no longer enjoy the right to be deleted and forgotten. In a media ecosystem where mistakes are not allowed or too costly, memory becomes the norm, forgetting becomes the exception, reason suppresses humanity, and rules dictate freedom [8].

Since the beginning of the “Diyi Newspaper”, newspapers have been responsible for the dissemination of information, and not only that, the contemporary media have to take into account the responsibility of spreading the mainstream values of society and guiding the benign development of public opinion. In the era of self-media, everyone has a communication channel, and media management has become very difficult. Traditional media management adopts a centralized management mechanism, and once there is content that deviates from the mainstream social values, violates the law, or is false and inaccurate, it can be removed, blocked, or banned to quickly curb the development. The decentralized and distributed operation mechanism of blockchain, on the other hand, has technical contradictions and conflicts for media management work. Once the content that contradicts the mainstream values and has a negative social impact is “chained”, it cannot be withdrawn or modified even if it is found; not only that, when there is false news, infringement of others’ rights and other illegal information, the spread of the content itself cannot be stopped except by issuing a statement to clarify the facts [9].

## 6 Conclusion

It can be seen that the development of new media technology brings infinite possibilities to people's lives, but also brings new challenges to the governance of the network environment. At present, the self-media market is booming, but there are a series of "secondary disasters" such as the difficulty to confirm the rights of original works, the difficulty to defend the rights of infringement, and the difficulty to identify and supervise the platform. Based on the technical features of decentralization, openness, traceability, immutability and anonymity of blockchain, it is conducive to promoting copyright transactions, ensuring data authenticity and clarifying copyright attribution, and has become the key to copyright storage and protection. At present, both academia and industry are starting from the technical perspective of blockchain, but before it is really put into practice, it is necessary to avoid overly optimistic thoughts such as "utopian" and always be vigilant and reflect on the technology. In conclusion, the communication mode of "blockchain + self-media" has a bright future, but the road is winding.

## References

1. Shuangge, Z., Ye, Y.: Advantages and limitations of blockchain technology applied to short video copyright protection. China Editorial **08**, 43–48 (2021)
2. Xiangen, L., Hongzhen, H.: Research on the prevention strategy of self-media "washing" in the era of big data. China Journalist **03**, 79–83 (2022)
3. Bohan, L.: The technical advantages and future applications of blockchain in copyright deposition–Insights from the first case of "Copyright Chain + Tianping Chain" collaborative forensics. J. Publishing **21**, 84–87 (2021)
4. Walch, A.: Blockchain's treacherous vocabulary: one more challenge for regulators. J. Internet Law 2017(2)
5. Pasquale, S.F.: The Black Box Society: The Secret Algorithms that Control Money and Information, Harvard University Press (2015)
6. Junlin, Y.: Law and accountability in the era of digital twins - a perspective on algorithmic black boxes through technical standards. Oriental Law **04**, 77–92 (2021)
7. Shanshan, Z., Wen, S.: A new path of blockchain application for short video dissemination and governance. Publishing Broad **10**, 37–39 (2021)
8. Wenjuan, Z., Chengbo, G.: Blockchain and the future media ecological order: a new view of thinking and methodology. News Lovers **03**, 12–14 (2020)
9. Yuan, G.: Cold thoughts on the blockchain news media fever. Audiovisual **11**, 200–201 (2020)

# A Novel Multi-Attribute Decision-Making Method Based on Linguistic Fermatean Fuzzy Sets and Power Average Operator

Xue Feng[1], Jun Wang[1], and Yuping Xing[2(✉)]

[1] School of Economics and Management, Beijing Jiaotong University, Beijing, China
19120604@bjtu.edu.cn, junwang_jg@mail.buct.edu.cn
[2] Glorious Sun School of Business and Management, Donghua University, Shanghai, China
ypxing@dhu.edu.cn

**Abstract.** This paper studies a novel tool for describing fuzzy information, called linguistic Fermatean fuzzy sets (LFFSs), in the process of multi-attribute decision-making (MADM). Compared to linguistic intuitionistic fuzzy sets and linguistic Pythagorean fuzzy sets, our LFFSs are more flexible and can depict more complicated decision-making information then the former two. In this study, we first introduce the notion of LFFSs. Afterwards, some other related concepts, such as operational rules, ranking methods as well as distance measure are interpreted. When considering aggregation operators for linguistic Fermatean fuzzy information, we generalize the classical power average (PA) operator into LFFSs and introduce the linguistic Fermatean fuzzy power average operator and its weighted form. Subsequently, a new MADM method based on LFFSs and their aggregation operator is developed. At last, an illustrative example is provided to show how our proposed method can be applied in solving realistic MADM problems.

**Keywords:** Fermatean fuzzy sets · linguistic Fermatean fuzzy sets · power average operator · linguistic Fermatean fuzzy power average operators · multi-attribute decision-making

## 1 Introduction

Multi-attribute decision-making (MADM) based on fuzzy information is an interesting and promising research topic, which has received much attention in the past decades. When considering fuzzy information based MADM methods, some researchers have focused on the issue of how decision makers' complicated and uncertain decision-making information can be effectively depicted. As an extension of intuitionistic fuzzy set [1], Pythagorean fuzzy set (PFSs) [2], originated by Dr. Yager, has been proved to be effective in depicting decision makers' fuzzy information [3]. Later on, Dr. Garg [4] extended PFSs and proposed the linguistic Pythagorean fuzzy sets (LPFSs), which use linguistic term to denote the membership and non-membership degrees. Compared to the linguistic intuitionistic fuzzy sets [5], LPFs can denote more complicated fuzzy decision-making information. Soon after its appearance, MADM based on LPFSs has been a hot research


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 30–42, 2023.
https://doi.org/10.1007/978-981-99-3618-2_4

topic. Liu et al. [6] investigated aggregation operators (AOs) and decision-making methods under LPFSs based on t-norm and t-conorm. Lin et al. [7] introduced interaction operations for LPFSs and based on which a novel partitioned Bonferroni mean was developed for LPFSs. Xu et al. [8] generalized LPFSs into cubic LPFSs and under which a power Hamy mean based decision-making method was presented. Sarkar and Biswas [9] investigated novel MADM method in LPFSs under Einstein t-norm and t-conorm.

LPFS is a powerful tool in MADM, however, its disadvantage is still obvious. The constraint of LPFS is that the square sum of the scripts of linguistic membership and non-membership degrees should not exceeds the pre-defined linguistic term set. As a matter of fact, this constraint cannot be always satisfied in actual decision-making situations, which motivates us to investigate novel fuzzy information representation tool. The Fermatean fuzzy set (FFS) [10], originated by Yager, is with the constraint that the cubic sum of membership and non-membership degrees is less than or equal to one. Obviously, compared with PFSs, FFSs have larger information space and can depict more complicated decision-making information. Hence, FFSs have been widely applied in solving realistic MADM problems [11–13]. The powerfulness of FFS motivates us to investigate its extension to accommodate more complicated decision-making environment. Just like extending PFSs to LPFs, this paper extends FFSs to linguistic Fermatean fuzzy sets (LFFSs), which satisfy the condition that the cubic sum of script of linguistic membership and non-membership degrees should be less than or equal to the pre-predefined linguistic term set. Hence, LFFS is more powerful and flexible than LPF. In the paper, we further study properties and AOs of LFFSs. Afterward, we use LFFS to propose a new MADM method.

The rest of this paper is organized as follows. Section 2 reviews some basic concepts that will be used in the following sections. Section 3 proposes the concept of LFFSs and investigate their desirable properties. Section 4 presents some AOs for linguistic Fermatean fuzzy information and discuss their properties. A novel MADM method is presented in Sect. 5. An illustrative example is provided in Sect. 6.

## 2 Basic Notions

### 2.1 Fermetean Fuzzy Sets

**Definition 1** [10]. Let $X$ be a given fixed set, then a Fermatean fuzzy set (FFS) defined on $X$ is expressed as.

$$F = \{\langle x, \alpha_F(x), \beta_F(x)\rangle | x \in X\}, \tag{1}$$

where $\alpha_F(x)$ and $\beta_F(x)$ denote the membership and non-membership degrees of element $x \in X$ to the set $F$, such that $0 \leq \alpha_F(x), \beta_F(x) \leq 1$ and $0 \leq (\alpha_F(x))^3 + (\beta_F(x))^3 \leq 1$. For convenience, the ordered pair $f = (\alpha, \beta)$ is called a Fermatean fuzzy number (FFN).

Basic operational rules for FFNs are presented as follows.

**Definition 2 [10].** Let $f = (\alpha, \beta), f_1 = (\alpha_1, \beta_1)$, and $f_2 = (\alpha_2, \beta_2)$ be any three FFNs, and $\lambda$ be a positive real number, then.

(1) $f_1 \oplus f_2 = \left(\left(\alpha_1^3 + \alpha_2^3 - \alpha_1^3\alpha_2^3\right)^{1/3}, \beta_1\beta_2\right)$;
(2) $f_1 \otimes f_2 = \left(\alpha_1\alpha_2, \left(\beta_1^3 + \beta_2^3 - \beta_1^3\beta_2^3\right)^{1/3}\right)$;
(3) $\lambda f = \left(\left(1 - \left(1 - \alpha^3\right)^{\lambda}\right)^{1/3}, \beta^{\lambda}\right)$;
(4) $f^{\lambda} = \left(\alpha^{\lambda}, \left(1 - \left(1 - \beta^3\right)^{\lambda}\right)^{1/3}\right)$;

### 2.2 Power Average Operator

**Definition 3** [14]. Let $x_i (i = 1, 2, \ldots, n)$ be a collection of positive real numbers, then the power average operator is defined as.

$$PA(x_1, x_2, \ldots, x_n) = \frac{\sum_{i=1}^{n}(1 + T(x_i))x_i}{\sum_{i=1}^{n}(1 + T(x_i))}, \tag{2}$$

where $T(x_i) = \sum_{j=1;j\neq i}^{n} Sup\left(x_i, x_j\right)$ and $Sup\left(x_i, x_j\right)$ denotes the support degree for $x_i$ from $x_j$, satisfying the following properties:

(1) $Sup\left(x_i, x_j\right) \in [0, 1]$;
(2) $Sup\left(x_i, x_j\right) = Sup\left(x_j, x_i\right)$;
(3) $Sup\left(x_i, x_j\right) \geq Sup(x_m, x_n)$ if and only if $\left|x_i - x_j\right| \leq |x_m - x_n|$

## 3 Linghuistic Fermatean Fuzzy Sets and Their Properties

In this section, the definition of LFSs and their properties. First of all, the definition of LFS are presented. Then, some other notions, such as operational rules, distance measure, comparison method, etc.

### 3.1 Definition of Linghuistic Fermatean Fuzzy Sets

**Definition 4**. Let $X$ be a given fixed set and $\tilde{S} = \{s_\alpha | 0 \leq \alpha \leq l\}$ be a continuous linguistic term set with odd cardinality. A LFFS $A$ definition on $X$ is expressed as.

$$A = \{(x, s_a(x), s_b(x)) | x \in X\}, \tag{3}$$

where $s_a(x), s_b(x) \in \tilde{S}$ denote the linguistic membership and non-membership degree of the element $x$ to the set $A$, respectively, such that $a^3 + b^3 \leq l^3$. The linguistic hesitant degree of $x$ to $A$ is expressed as $\pi = s_{\sqrt[3]{l^3 - a^3 - b^3}}$. The ordered pair $\gamma = (s_a, s_b)$ is called a LFFN.

**Remark 1.** It is noted from Definition 4 that all LIFSs and LPFSs are LFFSs. In other words, LIFSs and LPFSs are two special cases of LFFSs.

### 3.2 Operational Rules of Linguistic Fermatean Fuzzy Numbers

Based on the operational of FFNs and considering the basic operations of LIFNs and LPVs, the operational rules for LFFNs are proposed as follows.

**Definition 5.** Let $\gamma_1 = \left(s_{a_1}, s_{b_1}\right)$, $\gamma_2 = \left(s_{a_2}, s_{b_2}\right)$ and $\gamma = (s_a, s_b)$ be any three LFFNs defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \le \alpha \le l\}$, and $\xi$ be a positive real number, then.

(1) $\gamma_1 \oplus \gamma_2 = \left(s_{\left(a_1^3+a_2^3-a_1^3a_2^3/l^3\right)^{1/3}}, s_{b_1b_2/l}\right)$;

(2) $\gamma_1 \otimes \gamma_2 = \left(s_{a_1a_2/l}, s_{\left(b_1^3+b_2^3-b_1^3b_2^3/l^3\right)^{1/3}}\right)$;

(3) $\xi\gamma = \left(s_{l\left(1-\left(1-a^3/l^3\right)^{\xi}\right)^{1/3}}, s_{l(b/l)^{\xi}}\right)$;

(4) $\gamma^{\xi} = \left(s_{l(a/l)^{\xi}}, s_{l\left(1-\left(1-b^3/l^3\right)^{\xi}\right)^{1/3}}\right)$.

**Theorem 1.** Let $\gamma_1 = \left(s_{a_1}, s_{b_1}\right)$, $\gamma_2 = \left(s_{a_2}, s_{b_2}\right)$ and $\gamma = (s_a, s_b)$ be any three LFFNs defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \le \alpha \le l\}$, then.

(1) $\gamma_1 \oplus \gamma_2 = \gamma_2 \oplus \gamma_1$;
(2) $\gamma_1 \otimes \gamma_2 = \gamma_2 \otimes \gamma_1$;
(3) $\xi(\gamma_1 \oplus \gamma_2) = \xi\gamma_1 \oplus \xi\gamma_2, \xi > 0$;
(4) $\gamma(\xi_1 + \xi_2) = \gamma\xi_1 \oplus \gamma\xi_2, \xi_1, \xi_2 > 0$;
(5) $(\gamma_1 \otimes \gamma_2)^{\xi} = (\gamma_1)^{\xi} \otimes (\gamma_2)^{\xi}, \xi > 0$;
(6) $(\gamma)^{\xi_1} \otimes (\gamma)^{\xi_2} = (\gamma)^{\xi_1+\xi_2}, \xi_1, \xi_2 > 0$

### 3.3 Distance Measure Between Any Two Linguistic Fermatean Fuzzy Numbers

**Definition 6**. Let $\gamma_1 = \left(s_{a_1}, s_{b_1}\right)$ and $\gamma_2 = \left(s_{a_2}, s_{b_2}\right)$ be any two LFFNs defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \le \alpha \le l\}$, then the normalized Hamming distance between $\gamma_1$ and $\gamma_2$ is defined by.

$$dis(\gamma_1, \gamma_2) = \frac{\left|a_1^3 - a_2^3\right| + \left|b_1^3 - b_2^3\right| + \left|\pi_1^3 - \pi_2^3\right|}{2l^3}. \tag{4}$$

### 3.4 Comparison Method of Linguistic Fermatean Fuzzy Numbers

**Definition 7**. Let $\gamma = (s_a, s_b)$ be a LFFN defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \le \alpha \le l\}$, then the score function of $\gamma$ is defined as.

$$S(\gamma) = \left(\frac{l^3 + a^3 - b^3}{2}\right)^{1/3}, \tag{5}$$

and the accuracy function of $\gamma$ is defined as.

$$H(\gamma) = \left(a^3 + b^3\right)^{1/3}. \tag{6}$$

For any two LFFNs $\gamma_1 = \left(s_{a_1}, s_{b_1}\right)$ and $\gamma_2 = \left(s_{a_2}, s_{b_2}\right)$ defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \leq \alpha \leq l\}$, then.

(1) If $S(\gamma_1) > S(\gamma_2)$, then $\gamma_1 > \gamma_2$;
(2) If $S(\gamma_1) = S(\gamma_2)$, then

If $H(\gamma_1) > H(\gamma_2)$, then $\gamma_1 > \gamma_2$;
If $H(\gamma_1) = H(\gamma_2)$, then $\gamma_1 = \gamma_2$.

# 4 Power Average Operators for Linguistic Fermatean Fuzzy Numbers

In this section, the classical PA operators is extended into the LFSs and some new AOs for LFNs are proposed. In addition, significant properties of these AOs are also investigated.

## 4.1 The Linguistic Fermatean Fuzzy Power Average Operator

**Definition 8**. Let $\gamma_i = \left(s_{a_i}, s_{b_i}\right)(i = 1, 2, \ldots, n)$ be a collection of LFFNs defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \leq \alpha \leq l\}$, then the linguistic Fermatean fuzzy power average (LFFPA) operator is expressed as

$$LFFPA(\gamma_1, \gamma_2, \ldots \gamma_n) = \frac{\oplus_{i=1}^{n}(1 + TT(\gamma_i)\gamma_i)}{\sum_{i=1}^{n}(1 + T(\gamma_i))}, \tag{7}$$

where $T(\gamma_i) = \sum_{j=1;j\neq i}^{n} Sup\left(\gamma_i, \gamma_j\right)$ and $Sup\left(\gamma_i, \gamma_j\right)$ denotes the support degree for $\gamma_i$ from $\gamma_j$, satisfying the following properties:

(4) $Sup\left(\gamma_i, \gamma_j\right) \in [0, 1]$;
(5) $Sup\left(\gamma_i, \gamma_j\right) = Sup\left(\gamma_j, \gamma_i\right)$;
(6) $Sup\left(\gamma_i, \gamma_j\right) \geq Sup(\gamma_m, \gamma_n)$ if and only if $\left|\gamma_i - \gamma_j\right| \leq |\gamma_m - \gamma_n|$

;

If we assume

$$\varphi_i = \frac{(1 + T(\gamma_i))}{\sum_{i=1}^{n}(1 + T(\gamma_i))}, \tag{8}$$

then Eq. (7) can be

$$LFFPA(\gamma_1, \gamma_2, \ldots \gamma_n) = \oplus_{i=1}^{n}\varphi_i\gamma_i, \tag{9}$$

where $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_n)^T$ is called the power vector weight, such that $0 \leq \varphi_i \leq 1$ and $\sum_{i=1}^{n}\varphi_i = 1$.

Based on the operational rules that presented in Definition 5, the following theorem is derived.

**Theorem 2.** Let $\gamma_i = \left(s_{a_i}, s_{b_i}\right)(i = 1, 2, \ldots, n)$ be a collection of LFFNs defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \le \alpha \le l\}$, the aggregated value by the LFFPA operator is still a LFFNs and

$$LFFPA(\gamma_1, \gamma_2, \ldots, \gamma_n) = \left( s_{l\left(1-\prod_{i=1}^{n}\left(1-a_i^3/l^3\right)^{\varphi_i}\right)^{1/3}}, s_{l\prod_{i=1}^{n}(b_i/l)^{\varphi_i}} \right), \tag{10}$$

**Proof.** When $n = 1$, the LFFPA operator is

$$LFFPA(\gamma_1) = \varphi_1\gamma_1 = \left( s_{l\left(1-\left(1-a_1^3/l^3\right)^{\varphi_1}\right)^{1/3}}, s_{l(b_1/l)^{\varphi_1}} \right)$$

Suppose that when $n = k$, then the LFFPA operator is

$$LFFPA(\gamma_1, \gamma_2, \ldots, \gamma_k) =$$

$$\left( s_{l\left(1-\prod_{i=1}^{k}\left(1-a_i^3/l^3\right)^{\varphi_i}\right)^{1/3}}, s_{l\prod_{i=1}^{k}(b_i/l)^{\varphi_i}} \right)$$

When $n = k+1$, the LFFPA operator is

$$LFFPA(\gamma_1, \gamma_2, \ldots \gamma_{k+1}) = \oplus_{i=1}^{k}\varphi_i\gamma_i \oplus \varphi_{k+1}\gamma_{k+1}$$

$$= \left( s_{l\left(1-\prod_{i=1}^{k}\left(1-a_i^3/l^3\right)^{\varphi_i}\right)^{1/3}}, s_{l\prod_{i=1}^{k}(b_i/l)^{\varphi_i}} \right) \oplus \left( s_{l\left(1-\left(1-a_{k+1}^3/l^3\right)^{\varphi_{k+1}}\right)^{1/3}}, s_{l(b_{k+1}/l)^{\varphi_{k+1}}} \right)$$

$$= \left( s_{l\left(1-\prod_{i=1}^{k}\left(1-a_i^3/l^3\right)^{\varphi_i} * \left(1-a_{k+1}^3/l^3\right)^{\varphi_{k+1}}\right)^{1/3}}, \right.$$

$$\left. s_{\left(l\prod_{i=1}^{k}(b_i/l)^{\varphi_i}\right) * l(b_{k+1}/l)^{\varphi_{k+1}}} \right)$$

$$= \left( s_{l\left(1-\prod_{i=1}^{k+1}\left(1-a_i^3/l^3\right)^{\varphi_i}\right)^{1/3}}, s_{l\prod_{i=1}^{k+1}(b_i/l)^{\varphi_i}} \right)$$

Therefore, the proof of the Theorem 2 is complete.

**Theorem 3** (Idempotency). Let $\gamma_i = \left(s_{a_i}, s_{b_i}\right)(i = 1, 2, \ldots, .n)$ be a collection of LFFNs defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \le \alpha \le l\}$, if $\gamma_i = \gamma = (s_a, s_b)$ for all *i,* then.

$$LFFPA(\gamma_1, \gamma_2, \ldots, \gamma_n) = \gamma \tag{11}$$

**Proof:** According to the aggregations of According to $\gamma_i = \gamma = (s_a, s_b)$, we can have $Sup(\gamma_i, \gamma_j) = 1$ for all $i, j = (1, 2, 3, \ldots, n)$, $(i \neq j)$, then $\varphi_i = \frac{1}{n}$ hold for all $i$. Then, we can get.

$$LFFPA(\gamma_1, \gamma_2, \ldots, \gamma_n)$$

$$= \left( s_{l\left(1-\prod_{i=1}^{n}\left(1-a_i^3/l^3\right)^{\varphi_i}\right)^{1/3}}, s_{l\prod_{i=1}^{n}(b_i/l)^{\varphi_i}} \right)$$

$$= \left( s_{l\left(1-\prod_{i=1}^{n}\left(1-a^3/l^3\right)^{\frac{1}{n}}\right)^{1/3}}, s_{l\prod_{i=1}^{n}(b/l)^{\frac{1}{n}}} \right)$$

$$= \left( s_{l\left(1-\left(1-a^3/l^3\right)^{\frac{1}{n}*n}\right)^{1/3}}, s_{l(b/l)^{\frac{1}{n}*n}} \right)$$

$$= \left( s_{l(a^3/l^3)^{1/3}}, s_{l(b/l)^{\frac{1}{n}*n}} \right)$$

$$= (s_a, s_b) = \gamma$$

**Theorem 4** (**Boundedness**). Let $\gamma_i = \left(s_{a_i}, s_{b_i}\right)(i = 1, 2, \ldots, n)$ be a collection of LFFNs defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \leq \alpha \leq l\}$, and $\gamma^- = \min(\gamma_1, \gamma_2, \ldots, \gamma_n)$ and $\gamma^+ = \max(\gamma_1, \gamma_2, \ldots, \gamma_n)$, then.

$$\gamma^- \leq LFFPA(\gamma_1, \gamma_2, \ldots, \gamma_n) \leq \gamma^+. \tag{12}$$

**Proof**. According to the Definition 5, we can obtain

$$\varphi_i \gamma_i \leq \varphi_i \gamma^+$$

and

$\sum_{i=1}^{n} \varphi_i \gamma_i \leq \sum_{i=1}^{n} \varphi_i \gamma^+$.

which means that $LFFPA(\gamma_1, \gamma_2, \ldots, \gamma_n) \leq \gamma^+$.

Similarly, we can also prove that $\gamma^- \leq LFFPA(\gamma_1, \gamma_2, \ldots, \gamma_n)$. Thus the proof of Theorem 4 is completed.

### 4.2 The Linguistic Fermatean Fuzzy Power Weighted Average Operator

**Definition 9.** $\gamma_i = \left(s_{a_i}, s_{b_i}\right)(i = 1, 2, \ldots, n)$ be a collection of LFFNs defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \leq \alpha \leq l\}$ and the corresponding weight vector be $\varpi = (\varpi_1, \varpi_2, \ldots \varpi_n)^T$, such that $\sum_{i=1}^{n} \varpi_i = 1$ and $0 \leq \varpi_i \leq 1$. Then the linguistic Fermatean fuzzy power weighted average (LFFPWA) operator is expressed as.

$$LFFPWA(\gamma_1, \gamma_2, \ldots, \gamma_n) = \frac{\sum_{i=1}^{n} \varpi_i (1 + T(\gamma_i)) \gamma_i}{\sum_{i=1}^{n} \varpi_i (1 + T(\gamma_i))}, \tag{13}$$

where $T(\gamma_i) = \sum_{j=1;j\neq i}^{n} Sup(\gamma_i, \gamma_j)$ and $Sup(\gamma_i, \gamma_j)$ denotes the support degree for $\gamma_i$ from $\gamma_j$, satisfying the following properties presented in Definition 8.

If we assume

$$\psi_i = \frac{\varpi_i(1 + T(\gamma_i))}{\sum_{i=1}^{n} \varpi_i(1 + T(\gamma_i))}, \tag{14}$$

then Eq. (10) can be.

$$LFFPWA(\gamma_1, \gamma_2, \ldots, \gamma_n) = \sum\nolimits_{i=1}^{n} \psi_i \gamma_i, \tag{15}$$

where $\psi = (\psi_1, \psi_2, \ldots, \psi_n)^T$ is called the power weight vector, such that $\sum_{i=1}^{n} \psi_i = 1$ and $0 \leq \psi_i \leq 1$.

**Theorem 5**. Let $\gamma_i = \left(s_{a_i}, s_{b_i}\right)$ be a collection of LFFNs defined on a continuous linguistic term set $\tilde{S} = \{s_\alpha | 0 \leq \alpha \leq l\}$, the aggregated value by the LFFPWA operator is still a LFFNs and

$$LFFPA(\gamma_1, \gamma_2, \ldots, \gamma_n) = \left( s_{l\left(1-\prod_{i=1}^{n}(1-a_i^3/l^3)^{\psi_i}\right)^{1/3}}, s_{l\prod_{i=1}^{n}(b_i/l)^{\psi_i}} \right), \tag{16}$$

The specific proof process of Theorem 5 is omitted here because it is similar to the proof of Theorem 2. In addition, it is easy to prove that the LFFPWA operator has the property of boundedness.

## 5 The Procedure of A novel Multi-Attribute Decision-Making Method

In this part, based on LFFSs and their AOs, a novel MADM method is presented. Assume that, there are $m$ alternative, which can be denoted by $A = \{A_1, A_2, \ldots, A_m\}$. The $m$ alternatives are evaluated under $n$ attributes, which can be denoted by $C = \{C_1, C_2, \ldots, C_n\}$. The weight of the attributes is $\varpi = (\varpi_1, \varpi_2, \ldots, \varpi_n)^T$, such that $0 \leq \varpi_i \leq 1$ and $\sum_{i=1}^{n} \varpi_i = 1$. Suppose $\tilde{S} = \{s_\alpha | 0 \leq \alpha \leq l\}$ be a continuous linguistic term set with odd cardinality, and decision makers use LFFNs defined on $\tilde{S}$ to express their evaluation opinion over alternatives. In the following, we present the main steps of our proposed MADM method.

*Step 1*: Collect decision makers' evaluation opinion. Based on the pre-defined linguistic term set $\tilde{S}$, decision makers use a LFFN $\gamma_{ij} = \left(s_{a_{ij}}, s_{b_{ij}}\right)$ express their evaluation value for attribute $C_j (j = 1, 2, \ldots, n)$ of alternative $A_i (i = 1, 2, \ldots, m)$. Hence, a linguistic Fermatean fuzzy decision matrix is $\mathrm{D} = \left(\gamma_{ij}\right)_{m\times n}$ is constructed.

*Step 2:* Normalize the original decision matrix according to the following formula.

$$\gamma_{ij} = \begin{cases} \left(s_{a_{ij}}, s_{b_{ij}}\right), C_i \in I_1 \\ \left(s_{b_{ij}}, s_{a_{ij}}\right), C_i \in I_2 \end{cases} \tag{17}$$

where $I_1$ and $I_2$ denote benefit type attributes and cost type attributes, respectively.

*Step 3:* Calculate the $Sup(\gamma_{ip}, \gamma_{iq})$ as follows.

$$Sup(\gamma_{ip}, \gamma_{iq}) = 1 - dis(\gamma_{ip}, \gamma_{iq}), \tag{18}$$

where $dis(d_{ip}, d_{iq})$ denotes the Hamming distance between $\gamma_{ip}$ and $\gamma_{iq}$, and $p, q = 1, 2, 3, \ldots, n (p \neq q)$,

*Step 4*: Compute the $T(\gamma_{ij})$ according to

$$T(\gamma_{ij}) = \sum\nolimits_{p=q=1, p\neq q}^{n} Sup(\gamma_{ip}, \gamma_{iq}), \tag{19}$$

*Step 5*: Compute the power weight $\psi(d_{ij})$ associating with the LFFN $\gamma_{ij}$ according to

$$\psi(\gamma_{ij}) = \frac{\varpi_i(1 + T(\gamma_{ij}))}{\sum_{i=1}^{n} \varpi_i(1 + T(\gamma_{ij}))}, \tag{20}$$

*Step 6*: Obtain the overall values of the alternatives $\{A_1, A_2, \ldots, A_m\}$ by using the LFFPWA operator, i.e.,

$$\gamma_i = LFFPWA(\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{in}). \tag{21}$$

*Step 7*: Calculate the score values $\gamma_i (i = 1, 2, \ldots, n)$ according to Eq. (5).

***Step 8***: Determine the optimal alternative according to the score values of each alternatives.

We provide the following flowchart to between demonstrate the calculation steps of our proposed MADM method (Fig. 1).

## 6 An Illustrative Example

This section applies our proposed method in a real MADM method to show its effectiveness.

*Example*: Let' consider a software systems selection problems. An enterprise is now considering to select a suitable software system, and there are four alternatives ($A_1$, $A_2$, $A_3$, and $A_4$). The four software systems are evaluated by a group of experts under for attributes, i.e., contribution about organization performance ($C_1$), effort to transform from current system ($C_2$), benefit of the software system ($C_3$), and outsourcing software developer reliability ($C_4$). The weight vector of attributes is$\varpi = (0.2, 0.3, 0.25, 0.25)^T$. Let $S = \{s_0 =$ "Extremely poor", $s_1 =$ "Very poor", $s_2 =$ "Poor", $s_3 =$ "Slightly poor", $s_4 =$ "Fair", $s_5 =$ "Slightly good", $s_6 =$ "Good", $s_7 =$ "Very good", $s_8 =$ "Extremely good"$\}$ be a linguistic term. Decision makers use LFFNs defined on $S$ to express their evaluation values. In the following, we use our propose method to help the enterprise to determine the best software system.

*Step 1*: Decision makes use a LFFN $\gamma_{ij} = (s_{a_{ij}}, s_{b_{ij}})$ to denote the evaluation value of the attribute $C_j (j = 1, 2, 3, 4)$ of alternative $A_i (i = 1, 2, 3, 4)$, and the original decision matrix is constructed, which is listed in Table 1.

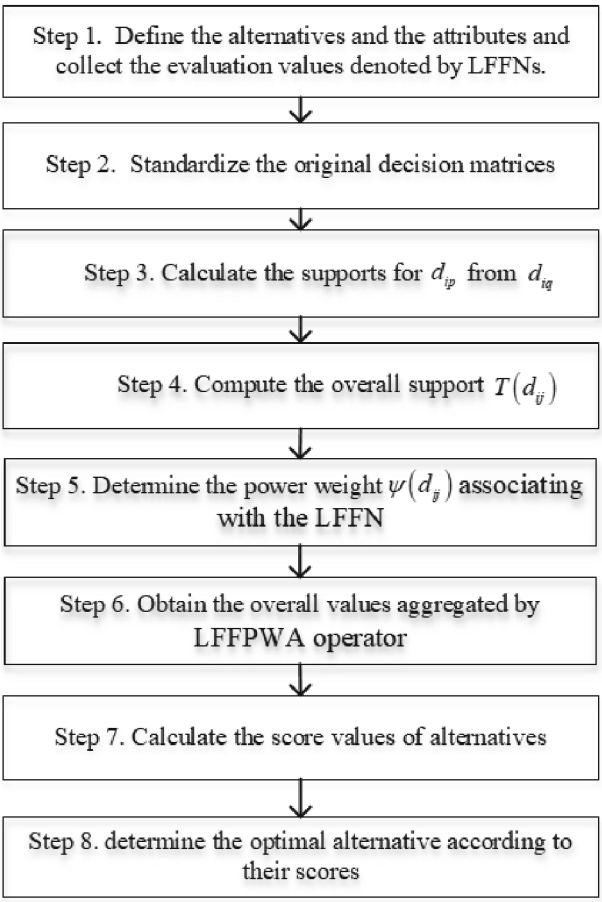


**Fig. 1.** The schedule of the proposed method

**Table 1.** The original decision matrix denoted by LFFNS.

| | A1 | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| $C_1$ | $(s_3, s_2)$ | $(s_1, s_3)$ | $(s_2, s_4)$ | $(s_1, s_4)$ |
| $C_2$ | $(s_4, s_3)$ | $(s_2, s_4)$ | $(s_3, s_4)$ | $(s_0, s_3)$ |
| $C_3$ | $(s_3, s_1)$ | $(s_2, s_3)$ | $(s_1, s_2)$ | $(s_2, s_1)$ |
| $C_4$ | $(s_4, s_1)$ | $(s_3, s_2)$ | $(s_2, s_1)$ | $(s_2, s_4)$ |

*Step 2:* Normalize the original decision matrix. It is obviously that all the attributes are benefit and hence the original decision matrix does not need to be normalized.

*Step 3:* Calculate the $Sup(\gamma_{ip}, \gamma_{iq})$ according to the Eq. (18). For convenience, we use $S^{pq}$ to denoted the $Sup(\gamma_{ip}, \gamma_{iq})$. Then, we can obtain

$$S^{12} = S^{12} = (0.5520, 0.6480, 0.8480, 0.6960)$$

$$S^{13} = S^{31} = (0.9440, 0.9440, 0.4960, 0.4960)$$

$$S^{14} = S^{41} = (0.7040, 0.7920, 0.4960, 0.9440)$$

$$S^{23} = S^{32} = (0.4960, 0.7040, 0.3440, 0.7920)$$

$$S^{24} = S^{42} = (0.7920, 0.5520, 0.3440, 0.6400)$$

$$S^{34} = S^{43} = (0.7040, 0.8480, 0.9440, 0.4960)$$

*Step 4:* Compute the $T(\gamma_{ij})$ according to the Eq. (19) and have

$$T = \begin{bmatrix} 2.2000, 1.8400, 2.1440, 2.2000 \\ 2.3840, 1.9040, 2.4960, 2.1920 \\ 1.8400, 1.5360, 1.7840, 1.7840 \\ 2.1360, 2.1280, 1.7840, 2.0800 \end{bmatrix}$$

*Step 5*: Compute the power weight $\psi(\gamma_{ij})$ by Eq. (20)

$$\psi = \begin{bmatrix} 0.2079, 0.2768, 0.2554, 0.2599 \\ 0.2102, 0.2706, 0.2714, 0.2478 \\ 0.2088, 0.2796, 0.2558, 0.2558 \\ 0.2069, 0.3095, 0.2296, 0.2540 \end{bmatrix}$$

*Step 6:* Aggregate the evaluation values and obtain the overall evaluation values $\gamma_i (i = 1, 2, \ldots, n)$ of alternatives

$$\gamma_1 = (s_{1.8339}, s_{4.1326}); \gamma_2 = (s_{0.5952}, s_{4.4910});$$

$$\gamma_3 = (s_{1.1968}, s_{4.7724}); \gamma_4 = (s_{0.5921}, s_{4.7744})$$

*Step 7:* Calculate the score values of alternatives and we have

$$S(\gamma_1) = 3.1173; S(\gamma_2) = 2.5871;$$

$S(\gamma_3) = 2.0808; S(\gamma_4) = 2.0155.$

**Step 8**. According to the score values shown in Step 7, we can get the ranking orders of alternatives, i.e., $A_1 \succ A_2 \succ A_3 \succ A_4$ and $A_1$ is the optimal alternative.

## 7 Conclusions

This paper introduced a novel MADM method for decision makers to determine the ranking order of alternatives. In order to this, we first proposed the concept of LFFS, which is a powerful tool to depict decision makers' complicated evaluation values. LFFS absorbs the advantages of FFS and linguistic term set and hence it has obvious superiorities over LPFS and LIFS. Afterwards, some new AOs for LFFSs based on PA were presented and their important properties are investigated in detail. Furthermore, based on LFFSs and their AOs, a novel MADM method was originated and its calculation steps were clearly presented. Finally, our method was applied in solving a real case to demonstrate its effectiveness. Our future research directions include the following two aspects. First, shall consider MADM methods under LFFSs based on consensus to make the final decision-making results more acceptable and reliable. Second, we will consider large-scale group decision-making problems based on LFFSs.

**Acknowledgment.** This work was supported by Shanghai Science and Technology Development Funds (Yang Fan Program, Grant no. 22YF1401400).

## References

1. Atanasso, K.T.: Intuitionistic fuzzy sets. Fuzzy Set Syst. **33**, 37–46 (1989)
2. Yager, R.R.: Pythagorean membership grades in multicriteria decision making. IEEE Trans. Fuzzy Syst. **22**, 958–965 (2013)
3. Peng, X., Selvachandran, G.: Pythagorean fuzzy set: state of the art and future directions. Artif. Intell. Rev. **52**(3), 1873–1927 (2017). https://doi.org/10.1007/s10462-017-9596-9
4. Garg, H.: Linguistic Pythagorean fuzzy sets and its applications in multiattribute decision-making process. Int. J. Intell. Syst. **33**, 1234–1263 (2018)
5. Zhang, H.: Linguistic intuitionistic fuzzy sets and application in MAGDM. J. Appl. Math., vol. 2014 (2014)
6. Liu, Y., Qin, Y., Xu, L., Liu, H., Liu, J.: Multiattribute group decision-making approach with linguistic pythagorean fuzzy information. IEEE Access. **7**, 143412–143430 (2019)
7. Lin, M., Wei, J., Xu, Z., Chen, R.: Multiattribute group decision-making based on linguistic pythagorean fuzzy interaction partitioned bonferroni mean aggregation operators. Complexity **2018**, 1–24 (2018)
8. Xu, W., Shang, X., Wang, J.: Multiple attribute group decision-making based on cubic linguistic Pythagorean fuzzy sets and power Hamy mean. Complex Intell. Syst. **7**(3), 1673–1693 (2021). https://doi.org/10.1007/s40747-020-00255-z
9. Sarkar, B., Biswas, A.: Linguistic Einstein aggregation operator-based TOPSIS for multicriteria group decision making in linguistic Pythagorean fuzzy environment. Int. J. Intell. Syst. **36**, 2825–2864 (2021)
10. Senapati, T., Yager, R.R.: Fermatean fuzzy sets. J. Ambient. Intell. Humaniz. Comput. **11**(2), 663–674 (2019). https://doi.org/10.1007/s12652-019-01377-0
11. Senapati, T., Yager, R.R.: Fermatean fuzzy weighted averaging/geometric operators and its application in multi-criteria decision-making methods. Eng. Appl. Artif. Intell. **85**, 112–121 (2019)

12. Senapati, T., Yager, R.R.: Some new operations over Fermatean fuzzy numbers and application of Fermatean fuzzy WPM in multiple criteria decision making. Informatica **30**, 391–412 (2019)
13. Aydemir, S.B., Gunduz Yilmaz, S.: Fermatean fuzzy TOPSIS method with dombi aggregation operators and its application in multi-criteria decision making. J. Intell. Fuzzy Syst. **39**, 851–869 (2020)
14. Yager, R.R.: The power average operator. IEEE Trans. Syst. Man Cybern. **31**, 724–731 (2001)

# Establishment of Evaluation System to Ensure the Safety of Environmental Industry in China Under the New Development Pattern

Weizhen Gao[1,2,3], Menggang Li[1,2], Xiaojun Jia[1,2(✉)], Yuejiao Zhao[1,2], and Chengzhang Zou[4]

[1] National Academy of Economic Security, Beijing Jiaotong University, Beijing 100044, China
{gaowz,mgli1}@bjtu.edu.cn, 13810582343@163.com
[2] Beijing Laboratory of National Economic Security Early-Warning Engineering, Beijing Jiaotong University, Beijing 100044, China
[3] Postdoctoral Programme of China Centre for Industrial Security Research, Beijing Jiaotong University, Beijing 100044, China
[4] Beijing Normal University, Beijing 100091, China
11132019066@bnu.edu.cn

**Abstract.** Under the new development pattern, this article modifies the index weights of the analytic hierarchy method using the entropy weight method and studies the evaluation system to ensure the safety of China's environmental industry. The results show that from 2014 to 2016, the range of environmental and industrial safety scores in China is 33.4–37.2, and its status is unsafe, but the overall situation continues to improve. The competitiveness of the environmental industry in the international market is improving, greatly increasing the viability of the environmental industry. Independent innovation has made great breakthroughs, showing an obvious improvement trend from 2016 to 2018. In 2017–2021, China's environmental industrial safety score was 41.8–56.7, which improved the industrial safety status and showed a general improvement trend. Foreign-funded enterprises have accounted for a large proportion of China's environmental industry; the environmental industry will continue this trend for a long time, and China should speed up the environmental industry development process. The state should formulate relevant policies in terms of the financial environment, market access, foreign mergers, acquisitions, etc., to ensure China's control over important environmental industries, improve the competitiveness of environmental industries, strengthen the control of important environmental industries and achieve environmental industry safety.

**Keywords:** environmental industry · analytic hierarchy method · safety evaluation · entropy weight method · system construction


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 43–52, 2023.
https://doi.org/10.1007/978-981-99-3618-2_5

## 1 Introduction

Environmental safety means that the general environment for human existence is free from crisis, and its ideal state is objectively derived from the absence of environmental threats. Environmental problems do not cause fear, and sustainable and harmonious development can be ensured [1]. Environmental safety is a typical survivability threat caused by non-military force, characterized by globality, primacy, intergenerational urgency, etc. [2–4]. With the increasing level of global economic development and the development of regional economic integration, all countries in the world pay more attention to the issue of industrial environmental safety [5]. Domestic researchers have thoroughly studied and analyzed the industrial safety evaluation system and constructed the industrial safety evaluation model, which includes an explicit safety model and implicit safety model, whose internal and external influencing factors need to be evaluated and analyzed in the dynamic open environment [6–8]. Through the evaluation of industrial competitiveness, the evaluation of the domestic environment of the industry, and the evaluation of industrial control, Y.Jing pointed out that internal and external factors include market environmental performance, performance structure, government regulatory environment, country concentration, and industrial control, production factor environment, industrial soft environment, finance, industrial performance, market demand, foreign-invested industrial control, industrial structure, foreign capital country concentration and other influencing factors to construct an industrial security assessment model, and pointed out that the evaluation of industrial security status requires qualitative and quantitative analysis combined with the study [9]. Ertao Cao and others proposed the principles, methods, and contents of geological environmental safety evaluation by discussing the interaction between human activities and geological environment systems and constructed an urban geological environment safety evaluation index system. Taking the Yangtze River Delta region as the research area, the rationality of the index system was verified by using the comprehensive index model and GIS spatial analysis methods [10]. Based on the construction of the "pressure-state" industrial safety evaluation model, Li Bin et al. constructed an industrial safety evaluation index system from the perspectives of industrial competitiveness, external shock risks, external industrial dependence, industrial control, and sustainable industrial development. Further, the efficacy coefficient method is used to evaluate the safety status of my country's industrial industry. The results show that the safety status of the industrial industry gradually decreases from 2015 to 2019, which is close to the lower limit of the degree of light police [8], in which some indexes may repeat, influencing the scientific nature of the evaluation results. Based on the research and analysis, Menggang Li proposed the industrial safety evaluation index system, fully integrating the domestic environment index of the first level industry and the industrial policy environment evaluation index and improving the design structure of the index system, but the weight assignment contradiction remained [11]. Although domestic scholars have carried out constructive research in designing the evaluation system, the concept of industrial security is unclear. Industrial security should include development security and practical security. The evaluation index system only reflects practical security [12]; the indicators only look at industrial control and industrial competition, government, industrial coordination, regulation and other issues. This is very important for industrial security but has been ignored [13]. The new

development pattern features dual circulation, in which domestic and overseas markets reinforce each other, with the domestic market as the mainstay. Based on this, this article studies the system of evaluation of environmental industrial safety in China under the new development pattern.

## 2 Industrial Safety and Its Influencing Factors

Industrial safety is the level of competition in a country's important industries, as well as its control, autonomy, and development rights to those industries, its ability to deal with potential threats in international industrial competition, and its status and ability to improve the corresponding industrial rights and interests are all indicators of industrial safety. Both external and internal factors will affect industrial safety. Table 1 shows the main factors that affect industrial safety.

**Table 1.** Factors Affecting Industrial Safety

| No | Factors Affecting Industrial Safety |
|---|---|
| 1 | International Direct Investment |
| 2 | Competition Environment in the Market |
| 3 | Trade Environment |
| 4 | Factors of the Financial Environment |
| 5 | Industrial Policy |
| 6 | International Competitiveness of the Industry |

From Table 1 we see that the factors affecting industrial security mainly include international direct investment, competition environment market, trade environment, factors of the financial environment, industrial policy, and international competitiveness of the industry. In Table 2, it can be seen that international direct investment will directly impact industry safety, development, and upgrading. For further development, multinational corporations have introduced their own advanced technology and capital experience in the host country, which will effectively promote the structural upgrading and technical level of the domestic host industry and help maintain industrial safety. When foreign capital enters a host country that has not yet established a sound economic system, it may directly control some important industries and directly threaten the safe development of national industries in the host country. If the host country wants to achieve industrial expansion, it must have a strong capacity to expand the industrial market. Otherwise, due to the weak protection of the domestic market, the domestic market will be occupied by foreign products, resulting in the outflow of resources, which will have a great impact and a negative impact on national industrial safety. Under the opening-up condition, industrial development can be greatly influenced due to factors such as enhanced international trade protection and the further opening of domestic markets, together with differentiated external competitive advantage and dependence on

industrial trade. The financial environment can directly affect the risks and costs of industrial financing. Through industry policy, the government can intervene in some industrial activities, resulting in the formation of a stronger industrial competitive advantage through adjustments such as inter-industry allocation and resources. For the international economic circle, the international competitiveness of a country's industry represents its competitiveness in a certain field, which will directly impact industrial safety.

## 3 Construction of a New System Model of Industrial Safety Index Environmental

The main components of the four-factor model to develop an environmental industrial safety index system, according to this paper, are the ecological industrial environment, industrial competitiveness, industrial productivity generation capacity, and industrial control power. The core position of the ability to generate industrial competitiveness is beyond doubt for industrial safety. Industrial control power is directly related to the degree of impact of foreign investment on industrial safety, can play a guaranteed role in the control of the industrial safety process, and the survival of an industry cannot be separated from the support of the domestic environment, while industrial competition is also an important, indispensable index. In this model, the general index is environmental industrial safety. There are 4 first-level indexes and 19 s-level indexes. The new evaluation model of the environmental industrial safety evaluation index system is shown in Table 2.

Based on research on existing environmental industrial safety and related indexes at home and abroad, a new evaluation model is constructed for the environmental industrial safety evaluation index system adopting a series of means, such as reference, screening, utilization, and adjustment. In the model, industrial environmental safety refers to ensuring the security of the current situation while continuing to confirm its security basis. If there are political and economic risks, the foreign side imposes an economic blockade on our country, making it impossible for us to continue providing high-quality services and products. The environmental industry is insecure. Enterprises integrate, select, transform, and use advanced technology, cutting-edge knowledge, and market resources to further realize the generation capacity of environmental industry productivity, during which the process requires continuous rise and mutation to enhance the competitiveness level of the environmental industry further.

## 4 Empirical Study on the New System Model of Environmental Industrial Safety Index

### 4.1 Index Boundary Setting and Score Mapping

The safety status levels of industry indexes are A (very safe), B (safe), C (critical state), and D (insecure), and the corresponding scoring ranges are [80,100], [40,60], [20,40], [0,20]. Based on the index values of the environmental industry and relevant foreign index data from 2014 to 2021, the safety status limits of the index are established, which are detailed in Table 3.

**Table 2.** The New Evaluation Model of Environmental Industrial Safety Evaluation Index System

| Evaluation Index System of Environmental Industry Safety | | | |
|---|---|---|---|
| 1.1.Industrial Productivity Generation Capacity | 1.2. Industrial control Power | 1.3. Industrial Ecological Environment | 1.4.Industrial Competitiveness |
| 2.11Technological innovation capacity | 2.21.Technology Control Rate by Foreign Capital | 2.31.industrial policy | 2.41.Profit Margin |
| 2.12. Management innovation capacity Capability | 2.22.Share of foreign brands | 2.32.Capital return rate | 2.42.Labor productivity |
| 2.13.R&D Input | 2.23.External Dependence of Industrial Capital | 2.33.Capital efficiency | 2.43.Dominant Comparative Advantage Index |
| 2.14.Absorption learning ability of technology | | 2.34. Degree of Industrial Concentration | |
| 2.15.Technical Competitiveness | | 2.35. Industrial Clustering Degree | |
| 2.16.Technology External Dependency | | 2.36. Human capital | |
| | | 2.37.Domestic demand growth rate | |

**Table 3.** Safety Status of Environmental Industry Indexes

| Safety Status | Very safe | Safe | Critical State | Insecure |
|---|---|---|---|---|
| Safety Status Score Range | [80,100] | [40,60] | [20,40] | [0,20] |
| Safety Level | A | B | C | D |

## 4.2 Empirical Methods

The efficacy coefficient method is applied in the process of evaluating the safety status of China's environmental industry. Through the goal programming principle, the impermissible and satisfactory values of each evaluation index are determined, and the impermissible and satisfactory values are taken as the existence of the lower limit and the upper limit of the index, respectively. Then, the degree of satisfaction of the index is calculated, and the score of the index is determined. After the comprehensive weighted average value is obtained, the status of the research object can be thoroughly evaluated.

### 4.3 Determination of Index Weight

This paper sets the index weight based on the analytic hierarchy process. This research method may be quite affected by subjective factors to some extent. As a result, this study incorporates the entropy weight method to modify the index weight, which can effectively improve the advantages of the weight structure. It is assumed that there are $m$ items in safety evaluation indexes and data sources for $n$ years, and there is a data matrix $X = (x_{ij})_{n\times m}$. When there is a large difference among the index values $x_{ij}$ of an index, the more important the index in the comprehensive evaluation. An index, if its values $x_{ij}$ are all the same, indicates that the index is not functional. As shown in formula (1).

$$H(x) = -\sum_{i=1}^{n} f(x_i) Inf(x_i) \tag{1}$$

The entropy weight method is used to determine the weight of each second-level index. In this case, the first step is to define $f_{ij}$ under the $j$ th index of matrix $X$. The proportion of the index value of the $i$ th evaluated object, such as in Formula (2).

$$f_{ij} = x_{ij} \Big/ \sum_{i=1}^{n} x_{ij} \tag{2}$$

Step two is to set the entropy value of item $j$ as $e_j$, and there is $e_j = -k \sum_{i=1}^{n} f_{ij} \times Inf_{ij}$ (in which $k = 1/Inn$). Then calculate the weight of each index through formula $w_j = (1 - e_j) \Big/ \sum_{j=1}^{n} (1 - e_j)$, when $w_j$ is larger, the index is more important.

The third step is to determine the final weight $a_j$ of each index, i.e. $a_j = w_j \Big/ \sum_{j=1}^{m} w_j$, and based on the above steps, each index weight is determined. Table 4 shows the weights of the environmental industrial safety evaluation index in China, we find that the largest weight is the industrial productivity generation capacity of the first-level indexes, which is 0.302, and the smallest weight is the industrial control power, which is 0.213.

### 4.4 Result Analysis

The power function transforms the evaluation score to ensure the measurability of the results, and then the formula is used: comprehensive efficacy coefficient = Σ individual efficacy factor × index weight, which is used to calculate the final measurement data of China's environmental industrial safety from 2014 to 2021 (Table 5).

Figure 1 reflects the results of the measurement of industrial environmental safety in China. It can be seen that the score ranges from 33.4 to 37.7 from 2014 to 2016 and its status is uncertain. From 2017 to 2021, the range of environmental industrial safety scores in China is 41.8–56.9. The industrial safety status has advanced, showing an overall trend of improvement, but it is still in a critical state. As a result, the safety of China's environmental industry must be further improved. Figure 2 reflects an analysis

**Table 4.** Weight of the Evaluation Index for Environmental Industrial Safety in China

| General Goal | First-Level Indexes and Weights | Second-level indexes and weights |
|---|---|---|
| Safety Status of Environmental Industry | Industrial Productivity Generation Capacity (0.302) | Technological Innovation Capability(0.156) |
| | | Management Innovation Capability (0.129) |
| | | Research and Development Input(0.134) |
| | | Absorption and Learning capacity of Technology(0.204) |
| | | Technical Competitiveness(0.211) |
| | | Technology External Dependency(0.167) |
| | Industrial control power (0.213) | Technology Control Rate by Foreign Capital (0.386) |
| | | Foreign Brand Share(0.325) |
| | | External Dependency of Industrial Capital(0.289) |
| | Industrial Ecological Environment (0.2226) | Industrial Policy(0.162) |
| | | Capital Return Rate(0.097) |
| | | Capital Efficiency(0.102) |
| | | Industrial Concentration Degree(0.139) |
| | | Industrial Clustering Degree(0.181) |
| | | Human Capital(0.137) |
| | | Growth rate of domestic demand (0.182) |
| | Industrial Competitiveness(0.259) | Profit Margin(0.334) |
| | | Labor Productivity(0.297) |
| | | Dominant Compaative Advantage Index(0.369) |

of the reasons for the low degree of industrial environmental safety in China from the perspective of first-level indexes.

**Table 5.** Calculation Results of the Efficiency Coefficient of Second-Level Index of Environmental Industrial Safety in China

| Second-level Indexes and Weights | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|---|
| Technological Innovation Capability | 19.7 | 20.8 | 21.9 | 24.7 | 26.8 | 32.9 | 35.6 | 39.8 |
| Management Innovation Capability | 27.5 | 28.3 | 29.2 | 30.3 | 33.4 | 36.3 | 39.7 | 44.2 |
| Research and Development Input | 21.2 | 27.4 | 27.1 | 26.2 | 29.9 | 33.5 | 32.1 | 32.3 |
| Absorption and Learning ability of Technology | 22.3 | 24.4 | 25.7 | 29.7 | 33.4 | 39.9 | 45.8 | 51.3 |
| Technical Competitiveness | 16.4 | 17.2 | 17.9 | 20.8 | 23.4 | 25.4 | 28.5 | 31.7 |
| Technology External Dependency | 39.6 | 41.8 | 50.6 | 43.3 | 45.9 | 47.6 | 48.8 | 50.2 |
| Technology Control Rate by Foreign Capital | 27.7 | 30.8 | 35.9 | 37.7 | 35.3 | 38.1 | 40.2 | 41.3 |
| Foreign Brand Share | 31.2 | 31.3 | 29.8 | 33.2 | 35.5 | 36.3 | 41.3 | 45.8 |
| External Dependency of Industrial Capital | 18.3 | 21.2 | 19.7 | 23.6 | 27.4 | 31.2 | 37.3 | 39.3 |
| Industrial Policy | 31.2 | 33.3 | 34.4 | 36.3 | 37.2 | 38.8 | 40.9 | 43.2 |
| Capital Return Rate | 41.3 | 42.4 | 43.4 | 45.7 | 47.8 | 49.3 | 50.4 | 53.3 |
| Capital Efficiency | 31.3 | 32.4 | 31.9 | 35.3 | 36.3 | 37.5 | 39.9 | 41.3 |
| Industrial Concentration Degree | 14.3 | 17.2 | 15.5 | 15.3 | 16.2 | 19.4 | 21.4 | 24.5 |
| Industrial Clustering Degree | 17.3 | 15.4 | 18.3 | 19.2 | 20.8 | 23.3 | 21.9 | 24.8 |
| Human Capital | 41.2 | 43.4 | 44.5 | 43.2 | 45.7 | 46.9 | 48.1 | 50.1 |
| Growth Rate of Domestic Demand | 79.5 | 76.7 | 73.4 | 71.2 | 67.8 | 66.2 | 68.4 | 63.9 |
| Profit Margin | 24.3 | 27.4 | 26.3 | 29.5 | 31.2 | 33.4 | 36.5 | 36.9 |
| Labor Productivity | 31.2 | 34.3 | 33.1 | 34.7 | 34.8 | 36.9 | 38.3 | 40.4 |
| Dominant Comparative Advantage Index | 9.2 | 10.4 | 12.4 | 13.2 | 13.1 | 16.4 | 19.3 | 23.9 |

Source: Calculated from relevant data from the China Scientific and Technological Statistics Yearbook, the China Industrial Enterprise Database and the China Environmental Statistics Yearbook from 2014 to 2021

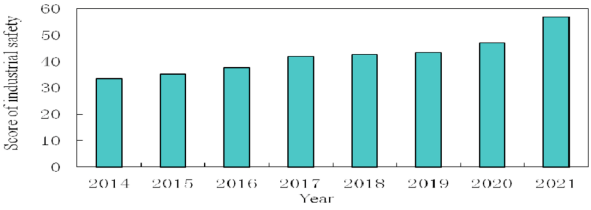


**Fig. 1.** The Measurement of Industrial Environmental Safety in China.

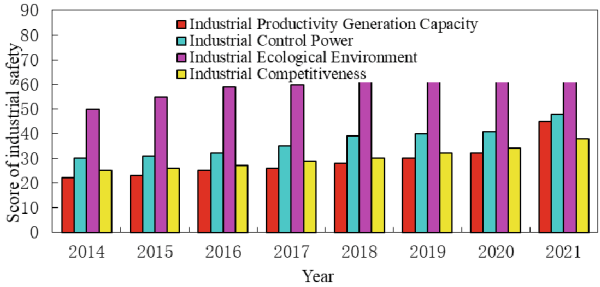


**Fig. 2.** Change in First-Level Indexes of Industrial Safety of Environmental Products.

As shown in Fig. 2, as far as the index of industrial productivity generation capacity is concerned, China's environmental industry as a whole shows a rapid developmental trend. In the index forecast, China will continue to take the productivity generation capacity of the environmental industry as a target of improvement in the medium and long term. This is due to the accelerated growth of scientific and technological personnel and the ability to manage innovation. In contrast, the level of investment in research and development and growth of technological competitiveness is relatively less rapid, directly impacting the growth of industrial productivity generation capacity. At the same time, the entire country needs to consider improving independent innovation ability, application and promotion of innovation achievements, and increase investment in research and development to continuously improve the conditions for independent innovation of enterprises. China has made great breakthroughs in independent innovation in the industrial control index, showing an obvious positive trend from 2016 to 2018. After many years of development, foreign companies have formed a large part of China's environmental industry. Therefore, our country should formulate policies and regulations on foreign mergers and acquisitions, market access, and financial environment to ensure control over important environmental industries, enhance their competitiveness, and promote the rapid implementation of environmental industrial safety.

Regarding the environmental index of the development of the environmental industry, the environmental industry will keep this trend for a relatively long time. China needs to accelerate the development of the environmental industry. With the continuous optimization and improvement of human capital, labor force, industrial policies and trade protection, demand growth rate decreases. At this stage, our government should play a more important macrocontrol function in industrial development, improve resource allocation efficiency and market environment optimization, and adopt systematically optimized financial, fiscal, and tax mechanisms to construct a complete core industrial chain for construction and development. In terms of the index of industrial competitiveness, from 2014 to 2021, the index status is not safe, but the overall situation continues to improve. In other words, the competitiveness of the environmental industry in the international market is increasing, greatly enhancing the viability of the environmental industry. However, existing problems require more attention. It is required to control the level of important environmental industries, improve industrial competitiveness, and promote the rapid implementation of environmental industrial safety.

## 5 Conclusions

In this paper, under the new development pattern, a study is carried out on the evaluation system to ensure the safety of China's environmental industry, and the following conclusions are drawn.

(1) The new environmental industrial safety evaluation model is made up of four first-level indexes: industrial control, industrial productivity generation capacity, industrial ecological environment, and industrial competitiveness. There are 19 s-level indexes in the model.
(2) From 2014 to 2016, the range of environmental industrial safety scores for China is between 33.4 and 37.7, which is unsafe. From 2017 to 2021, the range of environmental industrial safety scores in China is 41.8 to 56.9, indicating an improved state, indicating an overall trend of improvement, but still in a critical state.
(3) China needs to formulate policies and regulations on foreign mergers and acquisitions, market access, and the financial environment to ensure control over important environmental industries. It is also required to vigorously improve the level of control over important environmental industries, enhance their industrial competitiveness, and promote the rapid implementation of environmental industrial safety.

## References

1. Zhu, J., Wei, D.: Re-building of evaluation indicator system for Chinese industry security and empirical research. Sci. Res. Manage., 146–153 (2013)
2. Liu, L., Zheng, K., Liu, L.: A discussion on some key concepts of industrial security. J. Beijing Jiaotong Univ. (Soc. Sci. Edn.) **14**, 33–39 (2015)
3. Bao, Q.: Analysis on the security status of China's core industrial chain under the new development pattern. Trade Fair Econ., 90–92 (2022)
4. Tan, F., Zhang, L., Li, M.: Construction of industrial safety index system in China from the perspective of low-carbon economy. Stat. Decis., 57–60 (2016)
5. Song, L.: Construction of enterprise safety production evaluation system based on 4M1E. Inner Mongolia Coal Econ., 10, 14 (2016)
6. Zhang, Z.: Establishment and application of coal mine safety evaluation system based on analytic hierarchy process. Energy Conserv., 209–210 (2022)
7. Zhang, Y.: Connotation and evaluation system of industrial chain security. China Econ. Trade Herald, 55–59 (2021)
8. Li, B., Gao, S.: Evaluation system and empirical research on China's industry security under the new external shocks. Mod. Ind. Econ. Inf. **12**, 185–189 (2022)
9. Jing, Y.: Research on evaluation index system of industrial safety. Economist, 70–76 (2006)
10. Cao, E., Zhang, C., Zhang, H.: Study on safety evaluation index system of urban geological environment. Railway Invest. Surv. **46**, 59–64 (2020)
11. Li, M.: Research on the classification of industrial security. Prod. Res., 190–191,222 (2006)
12. Li, Z., Wang, S.: The theoretical logic and realization path of state-owned enterprises to improve the modernization level of industrial chain. Study Explor., 112–120 (2021)
13. Li, Y.: Innovation ecosystem under evaluation system and empirical research on China's manufacturing industry security. Forum Sci. Technol. China, 22–30 (2018)

# Research on Copyright Protection of Digital Publications Based on Blockchain Technology

Ning Wang[1] and Liang Wang[2(✉)]

[1] School of Journalism and Publishing, Beijing Institute of Graphic Communication, Beijing, China

[2] School of Economics and Management, Beijing Institute of Graphic Communication, Beijing, China

wangliang@bigc.edu.cn

**Abstract.** The development of the Internet has made digital publishing an inevitable trend for the transformation and development of the publishing industry, and with it comes a series of problems: easy access within, easy copying and long time to confirm the rights, high transaction and maintenance costs, which makes the copyright issue always a bottleneck for the development of the digital publishing industry. This paper attempts to explore the application path of copyright deposit, copyright review, copyright maintenance, copyright transaction and copyright supervision from the features of distributed ledger and impenetrability of blockchain technology to promote the high-quality development of China's digital publishing industry.

**Keywords:** digital publishing · Blockchain · Copyright protection · Infringement

## 1 Introduction

On March 29, 2022, the website of the Hubei Provincial Copyright Protection Center announced a case of copyright infringement. Since 2018, the defendant DingYue Company, under the management or participation of 12 defendants including Mr. Qin, used web crawler technology to crawl genuine electronic books and display them in its "Hong Yan Chuan Shu" and "TXT Free Novels" apps for others to access and download without the permission of Palm Reader Technology Co. "Ltd. And Beijing Fantasy Vertical Network Technology Co. The defendant Mr. Qin and others 12 people were arrested and brought to justice in March 2019. The court of first instance ruled that DingYue company and the 12 defendants, including Mr. Qin, were all guilty of copyright infringement and sentenced DingYue to a fine of 1.5 million yuan, while the other 12 suspects were sentenced to prison terms ranging from three years to one year and three months and a number of fines, as appropriate. After the first trial, all parties accepted the verdict and did not appeal.

This work is partially supported by National Social Science Fund of China under contract with No. 21BXW037.


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 53–62, 2023.
https://doi.org/10.1007/978-981-99-3618-2_6

All digital cultural works with creative nature we can define it as digital publications. As a kind of digital publication, the piracy of e-books is endless and repeatedly prohibited. It is difficult for copyright owners to get the profit they deserve, and it is difficult for digital publications to achieve effective development and sustainable development through the old model of copyright confirmation. The market is chaotic and development is hampered. Therefore, digital publishing must find a way out that not only meets the requirements of the development of the times but also helps to transform and upgrade traditional publishing, and the application of blockchain technology in the field of digital publishing provides the possibility to solve the existing problems and break through the bottleneck.

## 2 Copyright Protection Issues and Industry Transformation Dilemma of Digital Publishing Development

Digital copyright protection has been the biggest obstacle for digital publishing on the road to development. Copyright, also known as copyright, refers to the rights that the copyright holder has under the law to the work, including payment for the manuscript and attribution. It is also a restricted right that other people can use the author's work by obtaining authorization and paying for the work, and the author can get economic and other rights and interests as a result [1]. In the field of digital publishing, plagiarism, piracy and other infringements occur frequently and in huge quantities. Traditional copyright registration takes a long time, usually one to three months; the copyright registration process is complicated, and the registration fee price is high, such as some text and oral works charge from 100 yuan, and from 300 yuan for more than 10,000 words. In addition, the field of digital publishing also faces problems such as low cost of piracy, high cost of rights protection, long cycle time, high energy consumption, and difficulty in obtaining evidence after infringement.

According to the official website of the National Copyright Administration, the total number of copyright registrations nationwide reached 4,186,549,000 in 2019, up 21.09% compared with 3,457,338,000 in 2018. Such a massive demand for copyright confirmation is already difficult to be met quickly by the traditional registration system alone. In addition to the difficulty of registration, the number of copyright infringement cases is also on the rise. More than 480,000 new cases of various types of intellectual property rights were received by the people's courts in 2019, and more than 310,000 civil cases involving copyright. From these figures, we can see that the long-term development of digital publishing is constrained by copyright issues. There are two main reasons for this: on the one hand, there is a lack of perfect legal and institutional protection in the field of digital publishing; on the other hand, while digitalization brings convenience to people, it also makes works with greater probability and risk of infringement. All these make the interests of publishers, creators, copyright holders and other groups cannot be guaranteed, and become the biggest challenge facing the digital transformation and development of the publishing industry.

# 3 Blockchain Technology in the Field of Digital Publishing

## 3.1 The Advantages of Blockchain Technology in Digital Publishing

### 3.1.1 Timestamp: Copyright Confirmation, Proof of Rights

Using blockchain technology, a corresponding timestamp will be formed when operating each data block, which can record the immutable time when the digital material appears. From the perspective of copyright, the earliest recorded person is undoubtedly the owner of the copyright, and even the author who prefers to remain anonymous can confirm the copyright ownership through the timestamp [2]. This has strong proof power in resolving copyright disputes, proof of rights and recourse. In addition, in terms of the loss caused by infringement, blockchain can record all the operations that are modified against the data content on the chain, and in case of infringement, the loss caused by infringement can be accurately and credibly evaluated through algorithms, so as to guarantee the legitimate interests of copyright right holders to the greatest extent [3].

### 3.1.2 Consensus Mechanism: Preventing Copyright Tampering

Consensus mechanism means that consensus can be reached among all nodes in the blockchain to finally determine the validity of a record, and the consensus mechanisms currently used include proof of workload, proof of equity, and proof of share authorization [4]. This consensus mechanism can well play a role in eliminating forgery. Taking Bitcoin as an example, it uses proof of workload, and it is only possible to forge a non-existent record if it controls more than 51% of the bookkeeping nodes in the whole network, which is basically impossible when there are enough nodes joining the blockchain. Similarly, when any node in the blockchain wants to change it, it must be agreed by more than 51% of the nodes, which is extremely costly. Therefore, the consensus mechanism can well prevent the copyright content from being tampered with, which plays a great protective role.

### 3.1.3 Digital Signatures: Verify that Copyright Information Has not Been Tampered with

A digital signature is a message to be sent followed by another paragraph that serves as proof of the originality of the sender of the message and guarantees that the message has not been tampered with. The sender checks the content to be sent through a hashing algorithm to get a hash value, then encrypts the hash value with a private key to get a digital signature, and then bundles the two together to send, and the receiver receives it by decryption the digital signature with a public key to restore the hash value, and then compares the restored hash value with the verified hash value through a hashing algorithm to see if it is consistent. This verifies whether the information comes from the sender and whether it has been tampered with, so as to protect copyright [5].

### 3.1.4 Distributed Storage: Decentralized and Efficient

The distributed network formed by blockchain has decentralized characteristics, which makes the same data content stored equally and openly and transparently in each node.

The database built by blockchain technology is global in scope, where all copyright trading activities are completed and all data are complete and clear [6]. In addition, each node in the decentralized blockchain system is supervised and restricted, and with the dual protection of de-trust mechanism and asymmetric encryption, anyone and any organization can conveniently register copyright. Compared with the cost of copyright registration in the traditional system, copyright owners only need to spend a very low cost for quick registration, which greatly simplifies the process of copyright registration; both sides of the transaction can also realize the transaction without Both sides of the transaction can also realize the transaction without knowing each other, which not only saves the intermediary cost but also reduces the operation and maintenance cost of the system while ensuring information security.

## 3.2 Blockchain Technology in Practice in the Field of Digital Publishing

### 3.2.1 Establishment of an Integrated Platform for the Confirmation + authorization + maintenance of Rights

Digital publications can be removed at any time after uploading on the platform, and defenders who find infringement are likely to be removed by users or the platform if they do not have timely access to evidence. Content providers cooperate with platforms or technology parties, and licensing also has problems such as opaque valuation. Based on this, blockchain technology can be used to protect the copyright of digital publications. Blockchain technology is equivalent to a distributed ledger, where each part is relatively independent and the copyright information recorded in real time cannot be tampered with, and thus is very reliable as evidence. In addition, blockchain technology can also monitor and manage digital publications in real time using smart contracts, which can automatically complete the work of digital publication copyright and platform authorization, as well as prevent infringement through early warning procedures.

At present, a one-stop service platform for electronic rights protection and authorization has been established at home and abroad, with functions including online authorization, network-wide monitoring, one-click rights protection, etc., to form an effective copyright protection system, improve the original's awareness of rights protection, realize an integrated platform that can quickly take evidence and solidify evidence, optimize and simplify copyright protection and rights protection procedures, and protect copyright protection for digital publishing (Fig. 1).

The first domestic "blockchain + copyright" protection platform is the "original" original content trading and protection platform, which was launched in May 2017. The platform can guarantee that all data will not be tampered with and damaged through the formation of digital "DNA identification", and the original content after "DNA identification" can be selected on the platform for commercialized paid reprint or free knowledge sharing agreement, which is convenient for authors to maintain their rights at a later stage [7].

The "China Copyright Chain" service platform was officially released by the China Copyright Association in June 2021, which can provide the whole process of copyright protection services such as copyright deposit, infringement monitoring, online forensics, issuance of letters and shelving, copyright mediation, and rights protection litigation for

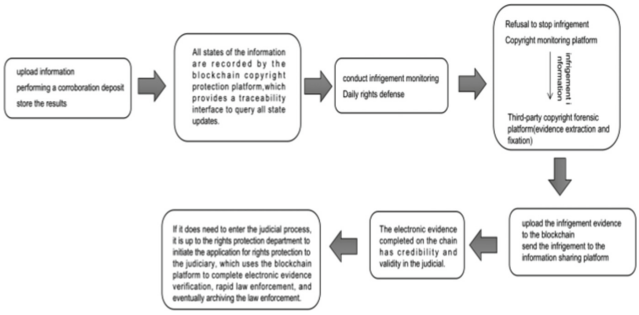


**Fig. 1.** Blockchain technology-based business process design for copyright protection of digital publications

rights holders [8]. In addition, Copyright Home's copyright blockchain platform has been connected to the Beijing Internet Court's Tianping Chain system in March 2019, and while users conduct deposition, the hash value of their deposition will be stored in the Beijing Internet Court's Tianping Chain system synchronously, which greatly improves the efficiency of court trials to check evidence [9].

### 3.2.2 "Blockchain + Copyright" Model: Giving Digital Publications a Unique Identity

At present, the "blockchain + copyright" model is emerging in the publishing industry, mainly using blockchain technology to generate unique digital credentials corresponding to specific works and artworks, and to achieve authentic and credible digital distribution, purchase, collection and use on the basis of protecting their digital copyrights, with the characteristics of openness, uniqueness and non-tamperability.

Tianjin Yangliuqing Woodblock Print Museum's digital collection of the classic "The Lady Traveling in Spring" was launched on the "Time Collection" digital copyright collection platform, and the first limited edition of 9,999 copies was snapped up in just a few seconds, which was beyond imagination. This is the first attempt of the publishing house of Tianjin Publishing Media Group in "blockchain + copyright".

At present, the most popular book in the publishing industry is the digital book collection, which is in the form of blockchain limited collection of physical books plus blockchain limited digital book collection tickets, the first to realize the combination of physical books and digital publications. It issues a unique, non-tamper able digital "ID card" - "digital voucher" for each limited edition book with different numbers and collection value. The digital "ID card" is anchored and uniquely corresponds to the digital book stamps and digital book covers that also have the value of collection [10].

Digital collections are digital cultural products derived from the physical publication of books, which are publications with independent value and authorized by the publisher. Unlike e-books, which are mostly unlimited and even able to be copied without restriction and without physical objects. Digital book collection, an innovative representative of new industry, new technology, new scene, and new product of blockchain technology application in publishing industry, realizes the innovation of copyright protection, digital

publishing, book distribution and value transformation, allowing the publishing industry to cross to the digital copyright and digital asset field of blockchain.

The design of book collection ticket in digital book collection refines the content and core ideas of books and fully integrates the content of physical books, including the elements of cover and illustration, and uses blockchain technology in the generation process, which not only increases the content of physical books, but also builds a brand new industrial model for publishers. In addition, this model builds a new scenario for publishing and distribution. The publisher offers a limited number of physical books plus a limited number of digital book collection tickets online, with a clear transaction process, a large distribution volume, and a good effect, which can promote the publicity and promotion of new books.

Collecting knowledge and transmitting value are the connotation of digital collections. The current uncertainty of digital collections lies in regulation, the country has not yet issued relevant transaction regulation and supervision policies, and the relevant laws are yet to be improved. The emergence of digital collection publishing form not only protects intellectual property rights and publication copyright protection, but also will show a wonderful digital cultural world for readers based on new consumer demand and after continuous iterative development and maturity.

## 4 Problems and Solutions of Blockchain Applied to Copyright Protection of Digital Publishing

### 4.1 Issues

#### 4.1.1 Legal Recognition of Copyright Certification

First of all, in terms of copyright tracking, whether the copyright proof audited and certified by the blockchain platform is supported by the law, for example, the Digital Millennium Copyright Act enacted by the United States in 1998 has played a great role in digital copyright protection, but whether the blockchain platform works are applicable to the law and are The primary challenge is whether blockchain platform works are compatible and recognized by existing laws.

Secondly, blockchain copyright certification needs to be supported by the government and the promulgation of new legal provisions to clarify the boundary of ownership of self-published works.

Finally, there are still loopholes in confirming rights according to copyright law. For example, there are already a certain number of blockchain content distribution platforms in the U.S. If self-published authors chain the same work on different platforms, they will get different public and private keys of the work, and which version is the standard for copyright authentication and maintenance becomes another problem that needs to be solved. Authentication work [11].

#### 4.1.2 Technical Security Issues

The core technology to ensure information security in blockchain is asymmetric encryption technology, which is not unbreakable. With encryption technology, there is naturally

decryption technology, and the continuous development of cryptography, information science and other discipline technologies will form a certain threat to asymmetric encryption technology. In addition, decryptors are likely to expose their public key IP address when decrypting with the public key, thus disclosing personal privacy, and there may also be the risk of infringement of the decrypted information content.

### 4.1.3 Absolute Control of Copyright is not Conducive to the Expression of Public Opinion

The initial purpose of using blockchain technology in digital publishing is for digital copyright protection, but imperfect technology can also impose limitations on copyright play. Usually, blockchain technology can instantly protect the digital works of the original creator, while the process of truth discovery is very long. Many people may have similar views on the same topic, but the public expression of similar views and topics is greatly restricted because one person quickly confirms the right with a timestamp.

Therefore, if the real-time use of nodality technology cannot effectively protect ideas, then its will cause curbs on the spread of inspiration and ideas. In addition, the use of blockchain technology for the confirmation of rights is not clearly stipulated in the law, and part of the public does not have the awareness of using blockchain technology for the confirmation of rights, based on which the problem of technical monopoly of ideas will greatly affect the public's desire for expression and is detrimental to the development of the industry. At the same time, the process of blockchain protecting copyright may stiffen the circulation of digital works, leading to the absolute control of technology combined with individuals and seriously affecting the sustainable development of the copyright market.

## 4.2 Solution

### 4.2.1 Improve the Legal Regulatory System

In order to make smart contracts develop better and smoothly in the field of digital copyright, they need to be moderately regulated, and the law will have to respond to blockchain technology and provide an effective regulatory system. Therefore, it is important to clarify the legal status of smart contracts and accelerate the construction of a legal regulatory system, so that the regulation can be based on the law and the risks brought by the new technology can be reduced, so that it can be better adapted to the laws and regulations of China. We can mainly start from the following two aspects.

### 4.2.2 The Executive Branch to Introduce Relevant Regulatory Laws and Regulations

As one of the important components in the field of regulating digital copyright, the administrative departments should promulgate administrative rules, regulations and measures as soon as possible to clarify the specific illegal acts and regulatory sanctions related to digital copyright smart contracts, which can be upgraded to "laws" when the conditions are ripe at the end of the transition period. At the same time, according to the national promulgation of special regulatory laws, localities should update and adjust their own

rules and regulations in a timely manner to maintain the consistency of the lower law with the higher law.

### 4.2.3 Building an Authentication Mechanism for the Subject of Smart Contracts

The decentralized and anonymous characteristics of smart contracts in blockchain make it difficult to identify the real parties in the transaction process, and the validity and relief of digital copyright contracts will be affected to a certain extent. Therefore, it is suggested that the regulatory backstage of blockchain technology should be managed in a hierarchical manner, with real-name authentication according to certain standards, and it is sufficient to adopt low standards for ordinary users and higher standards for enterprises and creators, etc. Such hierarchical management It is conducive to the efficiency of transactions and the allocation of regulatory resources. Moreover, this information can only be managed by authoritative institutions to prevent leaks and eliminate the concerns of the parties concerned about privacy and security. When infringement is reported or suspected infringing works are detected, the verification of the real name information of the parties concerned can only be initiated through layers of approval to pursue the legal responsibilities of the relevant personnel.

### 4.2.4 Optimize Blockchain Technology and Enhance the Confidentiality of Personal Information

In the era of mobile internet, personal information leakage has brought a lot of troubles to the people. If the leakage and illegal collection of personal information are allowed, the development of digital economy will hardly be healthy and may even become a fatal flaw in the future. The suspected risk of leakage of blockchain technology will inevitably affect the use of the technology by enterprises and individuals. Therefore, the government and society level need to further strengthen the confidentiality strength of blockchain technology by investing more funds in technology research and development, while how to attract more funds to invest needs to further optimize the trustworthiness of the block data itself and the stability of related equipment, so as to enhance its investment value.

### 4.2.5 Strictly Stipulate the Procedure of "RIght to Amend"

In order to prevent technical monopoly of ideas, blockchain can be appropriately granted certain modification rights to protect the right of as many members of the public as possible to fully express their opinions on a certain topic and to confirm their rights. Moreover, blockchain technology is not widely used in the field of copyright protection at present and is still under development. From a regulatory point of view, making blockchain technology applicable to existing laws seems to be the most efficient option, but it can also cause a great deal of problems. Blockchain technology has authority in the field of copyright protection because of its "immutability". If the "right of amendment" can be exercised on the blockchain, it will certainly lead to substantial mutation of blockchain technology.

The reason why blockchain has been able to attract many high technologies is due to its special trust mechanism. If the contents recorded in the blockchain can be changed at

will, there is no substantial difference with ordinary platforms, and the unique advantages are lost. In order to retain this characteristic, some people will propose that the law can be modified to adapt to the development of blockchain technology, but the law has a lagging nature and cannot keep up with the speed of technological development, and friction between them is inevitable.

In order to maximize the distance between law and technology, we can start from two aspects together. On the one hand, if operationally feasible, we can appropriately change the inherent properties of blockchain and make it have "modification" procedures.

On the other hand, it is necessary to set up a national-level regulator,, to strictly review the procedure and authority of "modification" and introduce the concept of "super user", i.e., when the author needs to modify the work, the regulator will approve the authentication, including identity information and face recognition. After the supervisor approves the work, the actual right holder is granted the right to "modify" the work on the blockchain. This not only preserves the "immutability" of the blockchain to a certain extent, but also protects the rights of the author, plus the introduction of state supervision to better maintain the rule of law environment of the blockchain.

## 5 Conclusion

The application of blockchain technology in the field of digital copyright has put online copyright protection into the fast lane of technology, making "publishing is depositing" possible, and the copyright of the right holder is protected more stably than before. However, new technologies are developing so fast that people's acceptance of new things and the speed of legal regulation of new technologies often lag behind, which will disrupt the social order to a certain extent, and for a long-term centralized management of social structure, the decentralized blockchain technology may cause some impact on it. The industry governance structure of "blockchain + copyright" still needs to be continuously adapted and adjusted. Along with the arrival of the 5G era, more digital content products will come into being, and blockchain copyright protection will be more deeply integrated with industry and technology, and its future development prospects are a long way off.

## References

1. Hong, L.: Digital copyright protection management scheme based on user behavior: an example of "YueYueYueXue" system of Electronic Industry Publishing House. Publishing Reference, (No. 9) (2016)
2. Shen, J.: Blockchain technology and its applications in digital content copyright protection. In: Yuan, C., Li, X., Kent, J. (eds.) Proceedings of the 4th International Conference on Economic Management and Green Development. AEPS, pp. 18–25. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-5359-9_3
3. . Lai, L., Li, Y.: Opportunities, challenges and development path of digital copyright protection under blockchain technology. Rule of Law Res. (04), 127–135 (2020)
4. Yan, Z., Guohua, G., Di, D., Fei Fei, J., Aiping, C.: Research on security in blockchain key technologies. Inf. Secur. Res. **2**(12), 1090–1097 (2016)
5. Lv, Y., Shen, X., Yang, J., Zang, W.: Summary of digital copyright protection based on blockchain technology. Remote Sens. Sci., 7 (2019)

6. Jian, W., Li, G., Jingning, Z.: Digital copyright protection based on blockchain technology. Radio Telev. Inf. **07**, 60–62 (2016)
7. CSN: The first domestic blockchain+copyright protection platform goes online [EB/OL]. https://www.cs.com.cn/sylm/jsbd/201705/t20170526_5299229.html
8. Beijing Daily Client: China Copyright Chain Goes Online, Exploring Technology for Copyright Management and Protection [EB/OL]. https://baijiahao.baidu.com/s?id=1701413557665919662&wfr=spider&for=pc
9. Sohu.com: "Tianping chain" the first case of Annie shares copyright blockchain evidence was recognized [EB/OL]. https://www.sohu.com/a/307313380_100021129
10. Lei, S.: Digital collections, the new money-making windfall in publishing?. Yangcheng Evening News, 17 July (2022)
11. Chen, X., Fu, Y.: A study of blockchain applications in the U.S. digital publishing industry. SE Commun., (09), 90–93 (2021). https://doi.org/10.13556/j.cnki.dncb.cn35-1274/j.2021.09.026

# Importance Evaluation of Factors for the Railway Accidents Based on TF-K

Dan Chang[1], Min Zhang[2(✉)], and Daqing Gong[1]

[1] Beijing Logistic Informatics Research Base, Beijing Jiaotong University, Beijing, China
{dchang,dqgong}@bjtu.edu.cn
[2] School of Economics and Management, Beijing Jiaotong University, Beijing, China
21120625@bjtu.edu.cn

**Abstract.** Rail accidents cause casualty and financial loss to society. In order to extract and identify the key factors from the accident reports more accurately, this study added the word frequency-correlation importance evaluation function(TF-K*) based on complex network on the basis of text mining, and built an importance evaluation model of factors for the railway accidents. When evaluating the importance of factors, the word frequency and the correlation between factors can be considered simultaneously. In this study, 213 railway accident reports from China and Britain were collected to analyze the cause of the accident, and the final results also verified the validity of the model.

**Keywords:** text mining · Complex network · Association feature · Accident factor · Importance evaluation

## 1 Introduction

In recent years, trains (especially high-speed trains) have become more and more important, but derailment and collision accidents still occur, such as China's D2809 train accident in 2022 and Egypt's Derailment accident on April 18, 2021, which have caused heavy casualties and huge economic losses. The accident investigation report is an unstructured text, if the key factors of the accident can be extracted and identified from that, and then putting forward countermeasures to prevent or reduce similar accidents, railway safety management can be better carried out. The *TF-IDF* is commonly used to extract feature on the basis of text segmentation. This method ranks the key factors according to the word frequency and document frequency, but ignores the relationship between factors. As a method for abstracting and modeling complex systems based on graph theory, complex network is often used to explore the interrelationships between factors. In this study, the causes of railway accidents were extracted based on text mining, and complex networks were introduced to represent the correlation features among factors, so as to build an importance evaluation model of factor for the railway accident. At the same time, the investigation reports of railway accidents in Britain and China were collected, and after the cause analysis of accidents was done by using the model, the suggestions on railway safety management would be given and the validity of the model would be verified.


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 63–76, 2023.
https://doi.org/10.1007/978-981-99-3618-2_7

## 2 Literature Review

### 2.1 Cause Analysis of Railway Accidents

In order to clarify the content of railway safety management, carry out risk prevention and control in advance and emergency management after the event, many scholars focus on railway accident analysis and accident cause research. Building accident cause model is a common analysis method. Classic accident cause models include Domino Theory Model [1], Swiss Cheese Model [2], STAMP (Systems ta effect Accident Model And TA effect Processes) accident analysis model [3–5], HFACS model [6], fault tree model [7], system thinking method [8] and Bayesian network (BN) [9]. Complex network theory is also an important method of accident cause research. Chen Y and Deng Y (2021) decomposed traffic accidents into several key nodes according to the sequence of occurrence, each of which was abstracted as a risk, and formed an accident chain one by one in chronological order. All the accident chains were combined to form a triggering network of risks, and key factors were identified by the nature of complex networks [10]. Lv L (2020) builds a double weighted and directed complex network model for railway accidents, in which the second weight is based on the degree of damage of accidents. The results show that the double weighted complex network can better identify the key factors causing high-speed railway accidents [11]. Therefore, it is a good research idea to use complex network to identify the key cause, and appropriate weight assignment can highlight the importance of accident cause.

### 2.2 Application of Text Mining in Accident Cause Analysis

Text mining techniques can also be used for causation analysis by processing incident reports. Goh et al. (2017) combined text mining and machine learning algorithms to achieve the classification of all kinds of accidents [12]. Zhong et al. (2020) used LDA topic model to classify high-dimensional word segmentation results and summarized risk factors on the basis of document word segmentation. However, most studies focus on the extraction of key causes, lacking the evaluation of the importance of factors [13]. Na XU et al. (2021) proposed a text mining method based on information entropy weighted word frequency *(TF-H)* on the basis of *TF-IDF* weight calculation, which can better evaluate the importance of word items [14], but the importance evaluation based on text word segmentation ignores the connection between factors.

Therefore, based on the $TF\text{-}K^*$ algorithm evaluation function [15] proposed by Chen Zhiyuan et al. (2022), this study combines text mining technology to build an evaluation model for the importance degree of railway accident causes, and at the same time brings the word frequency of accident cause factors and the strength of correlation characteristics between factors into the importance degree consideration range.

## 3 Theory of Complex Network

Complex network as a method of abstracting and modeling complex systems based on graph theory, uses nodes to represent elements within the system, and edges to represent the relationships between elements. Some metrics in complex networks are as follows:

**Node.** As the complex network is the abstraction of the complex system, the nodes in the complex network correspond to each entity in the complex system.

**Edge.** An edge is the relationship between nodes in a complex network, which means the relationship between different entities in a complex system. Edges are weighted to indicate how close the connection is. There are two types of edges: undirected edges and directed edges, corresponding to undirected complex networks and directed complex networks.

**Degree of the Node.** The degree of a node refers to the number of all nodes adjacent to the node, that is, the number of edges connecting the node. Degree is an important parameter to characterize the importance of a node and its influence on surrounding nodes.

According to the definition of complex network, some nodes that play key roles in the structure and function of the network can be discovered through the degree of nodes. And the edge weight can reflect the correlation between nodes. The occurrence of railway accidents is the result of the interaction and co-action of various influencing factors, so complex network model could be introduced to reflect the interaction between the factors of railway accidents.

## 4 Importance Evaluation Model for the Railway Accidents' Factor

Based on the text mining method and complex network theory, this study constructed an evaluation model for the significance of railway accident cause, and analyzed the text information of railway accident reports. Figure 1 is the framework of the model, which is mainly divided into two modules:1) accident cause factor extraction module based on text mining; 2) Accident cause-importance evaluation module based on complex network and improved word frequency-correlation importance evaluation function *(TF-K*)*.

**Factor Extraction Module:**
When using text mining technology to extract accident factors from the obtained railway accident report, it mainly includes the following steps: text pre-processing, construction of customized word segmentation thesaurus and stop word thesaurus, word segmentation and word frequency statistics.

(1) Text pre-processing: choose the cause-summary in the accident report and convert it into Chinese, and remove the repeated and worthless text data to obtain a text data set.
(2) Build a customized thesaurus: in order to keep the original semantic information, combine the Sougou thesaurus and the description of the phrases in the accident report to build a special corpus for railway accident reports.
(3) Text segmentation and keyword extraction: use the Jieba word segmentation package in python and import the customized thesaurus to perform a segmentation process, and then extract high-frequency keywords. The causative factor framework of railway accidents including human factors, technical factors, environmental factors, and management factors [16]. The factors are finally obtained.

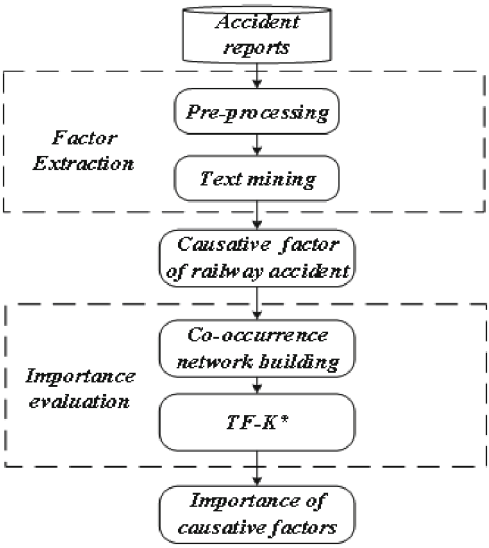


**Fig. 1.** Model framework of cause-importance assessment of railway accidents

**Factor Importance Evaluation Module:**
After extracting the causal factors of railway accidents, this study introduces a complex network to obtain the correlation degree between the factors, and brings it to the function of word frequency-correlation relationship evaluation, which is called *TF-K*$^*$ [15] based on the research of Chen Zhiyuan (2022). The specific operations are as follows:

(1) Construct a co-occurrence matrix according to the co-occurrence of accident causes in the accident report, and generate an undirected weighted network of accident causative factors;
(2) On the basis of *TF-IDF*, the network edge weights are used to calculate the importance evaluation function based on the word frequency-association relationship, and finally the importance of the causative factors of the railway accident is obtained. The formula is as follows (1):

$$TF - K^*(i) = TF \times K^*(i) \tag{1}$$

Among them, *TF* represents the word frequency in the accident reports. $K^*$(i) represents the correlation strength of the accident factor I, refers to the number of factors strongly associated with the accident factor i. Formula is calculated as follows (2):

$$K_{\mathrm{i}}^* = \sum_{i,j=1}^{N} a_{ij}^* = \sum_{i,j=1}^{N} a_{ji}^* \tag{2}$$

$N$ represents the total number of accident factors, and $a_{ij}^*$ represents the correlation edge weights of nodes $i$ and $j$ in the network and it can be calculated through (3):

$$a_{ij}^* = \begin{cases} 1\ ,\ w_{ij} \geq \overline{w}_{ij} \\ 0\ ,\ w_{ij} < \overline{w}_{ij} \end{cases} \tag{3}$$

In Formula (3), $w_{ij}$ is the edge weight of node $i$ and $j$ of the complex network, which is determined by the co-occurrence frequency of $i$ and $j$ in the accident report in this study. $\overline{w}_{ij}$ is the average edge weight of the network. The importance of the cause of the accident is obtained by combining the above formula. The larger the $TF\text{-}K_{(i)}{}^{*}$ value is, the greater the influence of the cause of the accident on the railway accident $i$, and the more important it is to carry out safety management and prevention.

## 5 Application

### 5.1 Data Collection

This study collected 213 railway accident reports from China's State Railway Administration and the UK website: https://www.gov.uk/raib-reports. Most of the accidents occurred during 2011 to 2021. The main types of accidents included serious accidents such as train derailment, collision, fire, explosion, etc.. The distribution of accident type is shown in Fig. 2.

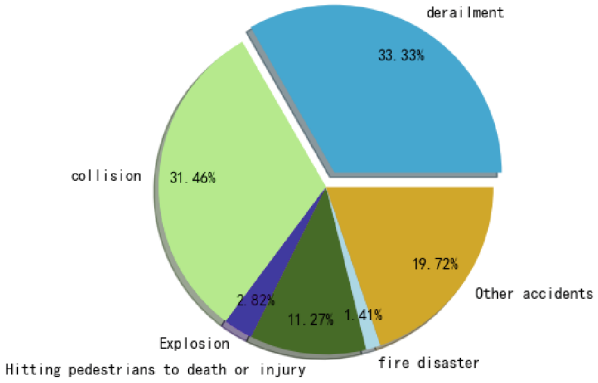


**Fig. 2.** Distribution of railway accidents

### 5.2 Extraction of Accident Factors

(1) Data pre-processing: The collected 213 accident reports contained a large amount of irrelevant information, including the specific situation of the accident, the investigation process and the division of responsibility, etc. The worthless information was removed, and the reason-summary in the accident report was extracted as the corpus of text mining.

(2) Construction of customized thesaurus: In order to reduce the noise vocabulary as much as possible, the vocabulary of railway safety field and some common phrases in sougou thesaurus is combined, as well as the words in accident report to construct thesaurus of segmentation. And in order to divide texts more accurately, this study added irrelevant words in accident reports, such as "Britain", "country", "federal",

"Gansu Province", into the stop words database based on HIT University stop words database.

Then, Jieba word package was imported into Pycharm and the customized word segmentation and stop word segmentation were called for word segmentation to obtain the key words in Python3.

(3) Word frequency statistics and extraction of key factors: After word segmentation, 2455 keywords were obtained, among which 155 were of greater frequency than 10. The word cloud map of accident cause was drawn as shown in Fig. 3.

By analyzing the description of keyword in the text, it is possible to judge the accident factors that it may correspond to. And from the four aspects of human factors, technical factors, environmental factors and management factors, the medium and high frequency causative words are classified and summarized, and finally 30 key causative factors of railway accidents are extracted, as shown in Table 1.


**Fig. 3.** Word Cloud Map of Causes of Railway Accidents

### 5.3 Importance Evaluation of Accident Factors

(1) Network construction: Draw a co-occurrence matrix according to the co-occurrence of accident causative factors in the accident report, and then take the accident causative factors as network nodes, and the co-occurrence frequency of factors as the edge weights of the network to generate the co-occurrence network of the accident causal factors as shown in Fig. 4.

(2) importance calculation: The co-occurrence network reflects the correlation characteristics between the causative factors of the accident, and can initially show the importance of the causative factors. For further evaluation, the evaluation function is calculated, and the results in Table 2 are finally obtained.

# 6 Result Analysis

## 6.1 Importance Analysis of the Accidents' Factors

As can be seen from Table 2, different factors have different impacts on railway accidents with different importance. The following is an analysis.

1) According to the results, A1, A2, A5, A8, A15 and A30 have the greatest influence on railway accidents.

Among them, A1, A2, A5 and A8 are human factors, and A2 has the highest impact, indicating that incomplete fault inspection, illegal operation, lack of communication or ineffective communication, and weak personal safety awareness are the main causes of railway accidents. This reflects a fluke mentality and a weak sense of safety in some operators, as well as accident victims. Therefore, it is important for the next step of safety management to carry out systematic training for relevant personnel from the aspects of safety awareness and professional skills, and strengthen the management of personnel.

A15 is a technical factor and A30 is a management factor. Design defects are fatal factors in railway accidents, so stricter technical review and more careful risk assessment must be carried out from the management level to the technical level to ensure that all designs do not go wrong.

It can also be inferred that compared with the existing technical defects, the management and audit departments should be more responsible for the occurrence of the accident for the value of A30 in the table is greater than A15, which also shows that there are still large loopholes in the safety audit work, and manage rectification. Needs to be done immediately.

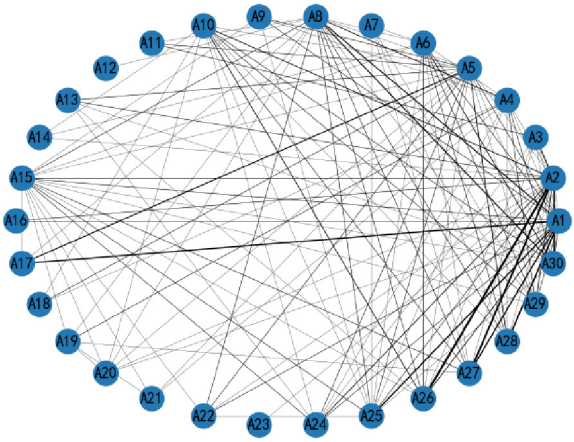


**Fig. 4.** Co-occurrence network of accident factors

2) The accident factors whose importance is in the second gradient include A4, A10, A24, A26, A27, A28, and A29. Among them, A24, A26, A27, A28 and A29 belong to

**Table 1.** Factors of railway accidents

| Category | Code | Accident Causation Factor | Word Segmentation |
|---|---|---|---|
| Human factor | A1 | violation operation | Violation operation, overspeed, execution, violation, according to regulations |
| | A2 | Incomplete fault diagnosis | Inspection, undetected, faulty, not in place, defect, not found |
| | A3 | Careless maintenance | Repair, maintain, keep, protect, update, replace |
| | A4 | response delay | Time, immediate, braking, untimely, reaction, too late |
| | A5 | Lack of effective communication | Communicate, contact, request, call, respond and authorize |
| | A6 | distractibility | Distraction, attention, distractibility, concentration, phenomenon, observation, forgetting |
| | A7 | Tiredness | Tiredness, action, reaction, state, sleep, prescription drugs |
| | A8 | Lack of safety awareness | Unprepared, safe, conscious, weak, think, casual, neglect |
| Technical factor | A9 | faulty signal equipment | Signals, signal lights, signal systems, annunciators, signal sources, interrupts |
| | A10 | Damage of the rail | Rail, broken rail, gauge, sleeper, unreinforced, twisted, widened, road damaged |
| | A11 | Faulty component | Split, loosen, come out, fall off, wear, parts |
| | A12 | Faulty connection component | Connecting rod, connecting, sealing, bridging, connecting and wiring |
| | A13 | Faulty braking system | Brake, performance, cock, variable resistance, cotter pin, resistance |
| | A14 | Faulty protective equipment | Protection, guardrail, buffer, block, road protection, slope protection, railing, isolation |

(*continued*)

**Table 1.** (*continued*)

| Category | Code | Accident Causation Factor | Word Segmentation |
|---|---|---|---|
| | A15 | Design defect | Capacitor, vertical track geometry, deterioration, manufacture, design, defect, installation |
| | A16 | Faulty monitoring device | Monitoring, invisible, interruption, extension, power, interlocking, display and control system |
| | A17 | incomplete information | Information, integrity, clarity, ambiguity and monitoring |
| Environmental factor | A18 | Bad geography | Slope, landslide, cut slope, slope, environment, geographical location |
| | A19 | Bad line condition | Line, congestion, stability, main line, side line, drainage |
| | A20 | bad weather | High temperature, heavy rain, fog, snow melting, thawing, extreme weather |
| | A21 | natural disaster | Mountain, landslide, flood, debris flow, inundation, erosion |
| | A22 | Foreign body intrusion | Obstacle, block, debris, car, stop, invade, approach |
| | A23 | Overloaded vehicles | Load, overrun, on-board, load, loading, carrying |
| Management factor | A24 | Incomplete quality inspection | Control, Standards, Evaluation, Specification, Infrastructure, Compliance |
| | A25 | Inadequate safety monitoring | Unattended, identification, misunderstanding, supervision, comprehensive, misjudgment |
| | A26 | Incomplete security checks | Safety, undetected, ensured, omitted, recorded, tested, degraded |
| | A27 | Incomplete rectification of hidden dangers | Postpone, update, repair, risk, introduce, implement, solve |

(*continued*)

**Table 1.** (*continued*)

| Category | Code | Accident Causation Factor | Word Segmentation |
|---|---|---|---|
| | A28 | The lack of training | Consensus, training, operation manual, familiar, unclear, inexperienced, management, notice copy |
| | A29 | Lack of emergency preparedness measures | Emergency, delay, prevention, remedial measures, emergency, unspecified |
| | A30 | Careless planning review | Plan, review, guidance, standards, processes, rules, arrangements |

management factors, indicating that there are many loopholes in the railway management department, such as the incomplete safety inspection, incomplete implementation of hidden danger rectification, lack of training for personnel, and insufficient emergency prevention measures. So, risk managing and controlling needs to be further strengthened.

A4 belongs to human factors and A10 belongs to technical factors. It shows that the untimely response operation in the accident and the damage of the rail are also the major causes of the accident. Therefore, it is also necessary to strengthen the investigation of hidden dangers of railway infrastructure.

**Table 2.** Importance degree calculation of accident factors

| Code | Causation Factor | *TF* | IDF | K* | *TF-IDF* | *TF-K** |
|---|---|---|---|---|---|---|
| A1 | violation operation | 0.097 | 0.650 | 19 | 0.063 | 1.847 |
| A2 | Incomplete fault diagnosis | 0.162 | 0.747 | 19 | 0.121 | 3.078 |
| A3 | Careless maintenance | 0.080 | 0.923 | 7 | 0.074 | 0.563 |
| A4 | response delay | 0.076 | 0.981 | 13 | 0.074 | 0.983 |
| A5 | Lack of effective communication | 0.056 | 0.747 | 19 | 0.042 | 1.072 |
| A6 | distractibility | 0.038 | 0.981 | 12 | 0.038 | 0.461 |
| A7 | Tiredness | 0.028 | 1.525 | 5 | 0.042 | 0.138 |
| A8 | Lack of safety awareness | 0.118 | 0.680 | 17 | 0.080 | 1.999 |

(*continued*)

**Table 2.** (*continued*)

| Code | Causation Factor | *TF* | IDF | K* | *TF-IDF* | *TF-K** |
|---|---|---|---|---|---|---|
| A9 | faulty signal equipment | 0.108 | 0.923 | 5 | 0.100 | 0.540 |
| A10 | Damage of the rail | 0.060 | 0.923 | 12 | 0.055 | 0.720 |
| A11 | Faulty component | 0.038 | 1.127 | 3 | 0.043 | 0.115 |
| A12 | Faulty connection component | 0.018 | 1.826 | 1 | 0.033 | 0.018 |
| A13 | Faulty braking system | 0.052 | 1.127 | 5 | 0.058 | 0.258 |
| A14 | Faulty protective equipment | 0.025 | 1.349 | 3 | 0.034 | 0.076 |
| A15 | Design defect | 0.066 | 0.712 | 17 | 0.047 | 1.122 |
| A16 | Faulty monitoring device | 0.061 | 1.224 | 2 | 0.075 | 0.122 |
| A17 | incomplete information | 0.030 | 0.785 | 4 | 0.024 | 0.120 |
| A18 | Bad geography | 0.066 | 1.224 | 3 | 0.081 | 0.198 |
| A19 | Bad line condition | 0.073 | 1.127 | 6 | 0.083 | 0.439 |
| A20 | bad weather | 0.040 | 0.981 | 7 | 0.039 | 0.277 |
| A21 | natural disaster | 0.049 | 1.224 | 4 | 0.060 | 0.197 |
| A22 | Foreign body intrusion | 0.088 | 0.872 | 6 | 0.076 | 0.526 |
| A23 | Overloaded vehicles | 0.053 | 1.826 | 1 | 0.096 | 0.053 |
| A24 | Incomplete quality inspection | 0.106 | 0.712 | 8 | 0.075 | 0.845 |
| A25 | Inadequate safety monitoring | 0.040 | 0.547 | 12 | 0.022 | 0.475 |
| A26 | Incomplete security checks | 0.100 | 0.446 | 9 | 0.044 | 0.896 |
| A27 | Incomplete rectification of hidden dangers | 0.109 | 0.504 | 9 | 0.055 | 0.983 |
| A28 | The lack of training | 0.095 | 0.596 | 8 | 0.056 | 0.758 |
| A29 | Lack of emergency preparedness measures | 0.068 | 0.622 | 9 | 0.043 | 0.616 |
| A30 | Careless planning review | 0.109 | 0.464 | 11 | 0.051 | 1.201 |

3) The importance score of A18-A23, which belong to environmental factors, is low, indicating that although harsh environments are unavoidable, we can reduce the occurrence of railway accidents through effective management and technical improvement.

But at the same time, the bad influence of the harsh environment cannot be ignored, and certain measures need to be taken to control the environment.

### 6.2 Validation of Model's Validity

In order to verify the validity of the importance evaluation model of factors for the railway accidents based on *TF-K*, t*his study focuses on comparing the importance calculation results of the two evaluation functions: *TF-K** and *TF-IDF*. As can be seen from the table, in most cases, the factor importance calculated by *TF-K** are consistent with those calculated by *TF-IDF*, but the former has a more obvious performance. However, in some special cases, the results calculated by TF-K are more accurate. The data of A6 and A11 in Table 2 are selected for comparison. It can be seen from the table: $TF_{A6} = TF_{A11}$, indicating that factor A6 and factor A11 have the same word frequency, but $IDF_{A6} < IDF_{A11}$, $K^*_{A6} > K^*_{A11}$, indicating that the document frequency of factor A6 is higher, the distribution of A6 in the survey report is more even and the correlation with other accident factors is higher, so in terms of the impact of the accident, The A6 factor is more important [15], this analysis corresponds to the result $TF - K^*_{A6} > TF - K^*_{A11}$, while the result of $TF - IDF_{A6} < TF - IDF_{A11}$ does not match the analysis. Therefore, it is verified that *TF-K** algorithm is better than *TF-IDF* in the evaluation of factor importance. Furthermore, the validity of the importance evaluation model of factors for the railway accidents based on *TF-K** has been verified.

## 7 Conclusion and Discussion

### 7.1 Conclusions

Based on the improved *TF-K** importance assessment function, this study constructed the importance assessment model of railway accident cause, and completed the extraction of accident key causation factors and importance assessment from the text report of railway accident. The study finally concludes that among the numerous causes of railway accidents, human factors and management factors are more closely related to accident occurrence. Therefore, the risk of railway accidents can be reduced to a certain extent by strengthening the training of safety awareness and relevant skills of relevant personnel, and by reforming the management system and carrying out railway safety management with a stricter, more thorough and more serious attitude. At the same time, text mining method is used to extract the key factors of accident cause, and the complex network is introduced to calculate the correlation between factors. The results also verify that the model has more advantages than the traditional *TF-TDF.*

### 7.2 Innovation and Reflection

Based on Chen Zhiyuan's importance evaluation function TF-K*, this study constructed the importance evaluation model of railway accident factors, and used the model to analyze the important influencing factors of railway accidents. There are two major innovations:

(1) The co-occurrence matrix and co-occurrence graph of railway accident factors are constructed by using complex network, and the relationship among factors is obtained to analyze the factors' importance.
(2) Based on TF-K*, the importance evaluation model of railway accident factors is constructed and applied to realize the innovation of application. This paper makes an innovation in application.

Here are the shortcomings and reflections of the paper:

(1) On the basis of word segmentation, this paper uses word frequency statistics to get key words, and summarizes the key factors. However, the keyword extraction based only on word frequency statistics has a lot of noise interference, so it can carry out the operation of noise reduction.
(2) Due to the limitation of sample data volume, the results of model validation have some errors, so we can consider expanding the sample data volume. At the same time, this study collected the accident adjustment reports of China and Britain. Due to the differences of railway conditions in different countries, certain comparison can be made in the next step.

**Acknowledgment.** This paper was supported by Beijing Logistic Informatics Research Base, China. We appreciate their support very much.

## References

1. Katsakiori, P., Sakellaropoulos, G., Manatakis, E.: Towards an evaluation of accident investigation methods in terms of their alignment with accident causation models. Saf. Sci. **47**(7), 1007–1015 (2009)
2. Reason, J.: Human error. Cambridge University Press, Cambridge (1990)
3. Leveson, N.: A new accident model for engineering safer systems. Saf. Sci. **42**(4), 237–270 (2004)
4. Song, T., Zhong, D., Zhong, H.: A STAMP analysis on the China-Yongwen railway accident. In: Ortmeier, F., Daniel, P. (eds.) SAFECOMP 2012. LNCS, vol. 7612, pp. 376–387. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33678-2_32
5. Dong, A.: Application of CAST and STPA to railroad safety in China. Diss. Massachusetts Institute of Technology (2012)
6. Shappell, S.A., Wiegmann, D.A.: A human error approach to accident investigation: the taxonomy of unsafe operations. Int. J. Aviat. Psychol. **7**(4), 269–291 (1997)
7. Liu, P., et al.: Fault tree analysis combined with quantitative analysis for high-speed railway accidents. Safety Sci. **79**, 344–357 (2015)
8. Fan, Y., et al.: Applying systems thinking approach to accident analysis in China: case study of "7.23" Yong-Tai-Wen High-Speed train accident. Safety Sci. 76, 190–201 (2015)
9. Özaydın, E., et al.: A hybrid model for marine accident analysis based on Bayesian Network (BN) and Association Rule Mining (ARM). Ocean Engineering 247, 110705 (2022)
10. Chen, Y., Deng, Y.: Traffic accident risk factor identification based on complex network. IOP Conf. Ser.: Earth Environ. Sci. Publishing **719**(3), 032074 (2021)
11. Lv, L.: High-speed rail safety analysis based on dual-weighted complex network. Diss. New Jersey Institute of Technology (2020)

12. Goh, Y.M., Ubeynarayana, C.U.: Construction accident narrative classification: An evaluation of text mining techniques. Accident Anal. Prev. **108**, 122–130 (2017)
13. Zhong, B., et al.: Hazard analysis: a deep learning and text mining framework for accident prevention. Adv. Eng. Inform. 46, 101152 (2020)
14. Na, X.U., et al.: An improved text mining approach to extract safety risk factors from construction accident reports. Saf. Sci. **138**, 105216 (2021)
15. Chen, Z., Wang, T.: Evaluation method of accident causes importance based on text mining and complex network: a case study of larger and above accidents in housing and municipal engineering. China Safety Prod. Sci. Technol. 18, 224–230 (2022)
16. Hua, L., Zheng, W.: Research on causation of railway accidents based on complex network theory. China Safety Sci. J. 29.S1, 114 (2019)

# The Contradiction Between Private Tutoring Industry and Capitalization

Lu Yu(✉)

School of Economics and Management, Beijing Jiaotong University, Beijing, China
18811591360@163.com

**Abstract.** Private tutoring industry is a supplement to national basic education. The service provided by this industry is to strengthen the test-taking ability of customers and provide knowledge supply service in line with the examination selection system. By analyzing the characteristics of education industry and the purpose of industry capitalization, and combining with the empirical data of uneven distribution of educational resources in China, this paper illustrates the essential contradiction between private tutoring industry and industry capitalization.

**Keywords:** Private tutoring Industry · Capitalization

## 1 Introduction

The private tutoring industry is an integral part of the education industry. Any national training program for personal ability outside of compulsory education can be regarded as an private tutoring industry. In the narrow sense, Off-campus training of exam ability for exam-oriented education subjects belongs to belongs to the private tutoring industry. The concept of private tutoring industry mentioned in this paper is a narrow concept. It should be noted that two specific types of training are beyond the scope of this article. One is training for personal interests, such as driving or gardening. The second is professional training for some professional projects, such as chef training (about cooking) or large machinery operation training.

This paper analyzes the characteristics of education industry, combined with the fact that China's education resources distribution difference, to explain the characteristics of private tutoring industry requires that the industry does not need capitalization.

## 2 Data

To observe the uneven distribution of educational resources in China, the following two data can be used to prove it (Tables 1, 2 and 3).

According to the data listed, we can see the specific situation of the difference in the distribution of educational resources in China. In general, the differences are manifested in the uneven distribution of individual family income and regional education funds. Specifically, there are three points:


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 77–84, 2023.
https://doi.org/10.1007/978-981-99-3618-2_8

**Table 1.** The scale of basic education in China

| Number of students (Ten thousands of people) | |
|---|---|
| Northern China | 1862 |
| Northeast China | 846 |
| Eastern China | 4614 |
| Central China | 2991 |
| South China | 2321 |
| Southwest China | 2641 |
| Northwest China | 1294 |
| Average school attendance | |
| Northern China | 645 |
| Northeast China | 569 |
| Eastern China | 809 |
| Central China | 619 |
| South China | 797 |
| Southwest China | 664 |
| Northwest China | 528 |
| The region represents the provinces and cities | |
| Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia | Northern China |
| Heilongjiang, Jilin, Liaoning | Northeast China |
| Shanghai, Shandong, Jiangsu, Anhui, Jiangxi, Zhejiang, Fujian | Eastern China |
| Hubei, Hunan and Henan | Central China |
| Guangdong, Guangxi, Hainan | South China |
| Chongqing, Sichuan, Guizhou, Yunnan, Tibet | Southwest China |
| Shaanxi, Gansu, Ningxia, Xinjiang, Qinghai | Northwest China |

Source: CHINA CITY STATISTICAL YEARBOOK (2006–2017) and Education statistics yearbook of China (2006–2017) [1]

**Table 2.** Education expenditure in China

| Unit: Ten thousand (Yuan) | | | |
|---|---|---|---|
| Year | Total | State financial funds for education | |
| | | Funds for education | General public budget for education |
| 1992 | 8670491 | 7287506 | 5649364 |
| 1995 | 18779501 | 14115233 | 10929473 |

(*continued*)

**Table 2.** (*continued*)

| Unit: Ten thousand (Yuan) | | | |
|---|---|---|---|
| Year | Total | State financial funds for education | |
| | | Funds for education | General public budget for education |
| 2000 | 38490806 | 25626056 | 21917652 |
| 2005 | 84188391 | 51610759 | 49460379 |
| 2006 | 98153087 | 63483648 | 61353481 |
| 2007 | 121480663 | 82802142 | 80943369 |
| 2008 | 145007374 | 104496296 | 102129675 |
| 2009 | 165027065 | 122310935 | 119749753 |
| 2010 | 195618471 | 146700670 | 141639029 |
| 2011 | 238692936 | 185867009 | 178217380 |
| 2012 | 286553052 | 231475698 | 203141685 |
| 2013 | 303647182 | 244882177 | 214056715 |
| 2014 | 328064609 | 264205820 | 225760099 |
| 2015 | 361291927 | 292214511 | 258618740 |
| 2016 | 388883850 | 313962519 | 277006325 |
| 2017 | 425620069 | 342077546 | 299197838 |
| 2018 | 461429980 | 369957704 | 319927298 |
| 2019 | 501781166 | 400465452 | 346485685 |
| The central budget | 41674035 | 28213048 | 17927053 |
| Local funds | 460107132 | 372252405 | 328558633 |

Source: National Bureau of Statistics

**Table 3.** The distribution of students' family background (%)

| Family Education background | | Family professional Background | | Family income background | | Family area Background | |
|---|---|---|---|---|---|---|---|
| Education background | % | Professional | % | Income (Yuan) | % | The administrative level | % |
| Graduate | 2.6 | Management of public institutions | 14.6 | 3001 of the following | 19.3 | Provincial capitals and Municipalities | 17.8 |
| Undergraduate | 14.3 | The enterprise management | 6.4 | 3001–5000 | 19.2 | Prefecture | 26.1 |

(*continued*)

**Table 3.** (*continued*)

| Family Education background | | Family professional Background | | Family income background | | Family area Background | |
|---|---|---|---|---|---|---|---|
| Education background | % | Professional | % | Income (Yuan) | % | The administrative level | % |
| Junior | 9.8 | Professional and technical | 7.8 | 5001–10000 | 20.1 | County or county-level city | 27.9 |
| Senior and Secondary | 28.6 | Staff | 2.9 | 10001–20000 | 15.0 | The towns | 7.3 |
| Middle | 29.8 | Business | 10.4 | 20001–50000 | 13.5 | Villages | 20.9 |
| Primary school | 12.4 | Agriculture | 28.2 | More than 50000 | 12.8 | | |
| Illiteracy | 2.6 | Factory and Transport | 5.2 | | | | |
| | | Migrant workers | 15.0 | | | | |
| | | Unemployedor Underemployed | 9.5 | | | | |
| Total | 100 | Total | 100 | Total | 100 | Total | 100 |

Source: China Statistical Yearbook (2010) andNational sample survey on population change (2009) [2].

1. The difference between the total number of basic schools (primary and secondary schools) and the average number of students is obvious.
2. According to China's National Bureau of Statistics, the proportion of government spending on education has leveled off. However, it should be pointed out that since 2015, China has begun to improve the guarantee mechanism of local compulsory education in terms of compulsory education, but in the statistical report of 2021, China's local education expenditure still occupies an absolute high position.
3. Different conditions for running schools in urban and rural areas. In 2015, the Ministry of Finance and the Ministry of Education issued the Notice on Further Improving the Funding Guarantee Mechanism for Urban and Rural Compulsory Education in order to improve school conditions in rural areas and promote educational equity. In 2017, the Ministry of Education issued the Supervision and Evaluation Measures for the Balanced Quality Development of Compulsory Education at County level, which proves that the problem of education equity at all levels in China still needs to be solved.

## 3 Analysis

The above data are mainly used to show that the unfair distribution of educational resources in China is still widespread. The differences are evident both in the macro data and in household income differences. From the perspective of education, income

inequality leads to unfair distribution of individual education resources, and the development degree of local finance determines the amount of education resources in a region. Therefore, in the topic of educational resources, regional resource inequality and family resource inequality exist simultaneously.

We can conclude from the above data:

1. From the total number of schools distributed by region to the relationship between family education and education investment, the situation of educational inequality is very obvious. "At the same time, the decrease of educational inequality within the group does not lead to the decrease of educational inequality as A whole" [3].
2. The unbalanced distribution of urban and rural educational resources is still a serious problem. As scholars say: "At present, there are some problems: the development of education in rural areas is lagging behind; the development of urban and rural education is imbalanced; The unenthusiastic distribution of educational resources is still serious" [4].

Of course, it has been a consensus to compensate for the imbalance of educational resources through fiscal subsidies. [5, 6] The main means is also to emphasize policy adjustment [7]. A relatively new point of view is to use information technology to solve the problem of unbalanced educational resources [8]. At the same time, the fairness of cost sharing in the education industry is also a problem [9]. But that's not the focus of this article.

It should be emphasized that this paper does not discuss how to solve the problem of unfair education resources at the national level. The purpose of this paper is to analyze the basic characteristics of the education industry, so as to conclude that restricting the development of private tutoring industry is conducive to the development of education and the realization of education goals.

The premise is that the characteristics of the education industry itself include:

1. Positive externalities are huge.
2. Human capital industry.
3. The period is relatively constant, and the content is relatively stable.

The private tutoring industry as a supplement to the education industry, is not only the orientation of the industry, but also the derivative characteristics of the industry compared with the education industry. Therefore, the private tutoring industry has the characteristics of education industry itself. In addition, its complementary features are as follows:

1. As a supplement to the education industry, its purpose of business behavior should not be inconsistent with the purpose of educational behavior.
2. It is a way to pursue outcome justice and make up for the unfairness of school education to a certain extent.

Therefore, the private tutoring industry cannot replace the education industry to carry out basic education, and few large capital can undertake such a large long-term investment as regional education. Secondly, the purpose of the private tutoring industry is to complete the trainees' learning of basic knowledge and develop their ability to take exams. In other words, the main business of the industry is to provide education services

to those who are willing to learn the basic knowledge of school, but need additional education because of various subjective and objective circumstances. This industry cannot compensate for the inequality of family income and the inequitable investment of family education resources. In terms of costs and benefits, service buyers pay money and time (study time). The benefits of this activity include knowledge and reflect the individual learning ability, obtain higher competition opportunities and benefits with other competitors in the examination.

Therefore, it can be seen that there are two preconditions for the development of this industry:

1. Only under the premise of the selection mechanism of exam-oriented education can there be a foothold for this industry. Only under the premise of timing and quantification, namely, stable preparation time (generally 3–6 years) and relatively stable amount of knowledge contained in the exam-oriented examination, and the mode of testing and selecting individual ability by means of examination, training and education can provide services corresponding to industrial needs. As a lifelong behavior, education and learning itself leave the field of training industry and are not the scope of this paper.
2. The industry can only play a complementary role if educational resources are relatively equitable. The industry's role is to provide supplementary education on a small scale and on a relatively equitable basis. However, it is impossible to remedy the inequality of education resources at the national level. Therefore, the greater the scope of development of this industry, the stronger the industrial concentration, but will aggravate the inequality of education.

To sum up, the way to develop private tutoring industry should be to carry out the characteristic development mode of small regions.The establishment of monopolistic competition or oligopoly in a city - wide market.In this way, targeted services can be provided to the regional market.However, the homogeneous products provided by the large-scale perfect competition market conflict with the unbalanced distribution of educational resources.The more homogeneous education services are, the more time costs are consumed and the results of unfair education resources are stabilized.At the same time, consumers have limited choice time, because the time from the beginning of learning basic knowledge to the selection education is relatively stable, and the total amount of services that the supplier can provide is also limited by all the knowledge contained in basic education.Therefore, in order to save transaction costs and reduce consumers' choice time, monopolistic competition should be the best market model.

Take the formula of Coase theorem as an example, because:

$$E = \frac{U}{TC + PC}$$

In this model, E is efficiency, U is utility, TC is transaction cost, and PC is production cost.Because the production cost and the utility value are relatively constant, the higher the transaction cost, the lower the efficiency situation.

It can be seen that, even if macro educational purposes at the national level are not considered, the contradiction of the development of this industry is that if the PCM market form is forcefully pursued, the prerequisite of "providing similar products" cannot

be guaranteed. On the contrary, if the regional market is established according to the specific needs of regionalization from the empirical perspective, excessive suppliers will increase transaction costs and damage efficiency. Therefore, the goal of industrial capitalization itself - to improve the efficiency of resource allocation has caused a fundamental contradiction, which can not be realized.

Therefore, even without considering the ideal individual educational development purpose, only with concrete benefits.In China's college entrance examination education system, the capitalization of private tutoring industry is also an inefficient operation mode.The industrial characteristics of long-term investment and strong externality contradict the benign operation mode of capital.Because the capitalized industrial model requires accurate and timely income return, investment return and industrial development cannot be balanced without internalization of externalities. In the personal education and training service defined in this paper, its ultimate purpose is only to assist the demander to complete the higher education selection examination.Therefore, under the selection mechanism of regional proportional enrollment, excessive cost input will eventually become redundant input.Under the premise of relatively stable knowledge points and the same educational resources, individual education and training will only change in individual ranking, thus resulting in the overall constant benefit of families of the same class who try to increase education costs in a short period of time (3–6 years) in order to compete for ranking.

## 4 Conclusions

The purpose of industrial capitalization should be to optimize the allocation of resources, improve the value-added rate of assets and increase investors' income.

And educational work because of its high externality and a number of social purpose characteristics. Before the internalization of externalities is completed, the development of private tutoring industry is bound to face a dilemma:

1. To achieve the purpose of capital operation, strengthen investor returns. Therefore, due to the lack of accurate income calculation model, social benefits must be abandoned.
2. If we want to achieve the social purpose of education and strengthen its positive externalities, it will be difficult to achieve the profit goal of investors.

In other words, the long-term nature of educational investment and the efficiency of industrial capitalization are a pair of inherent contradictions.

By analyzing the characteristics of education industry and private tutoring industry, this paper points out that private tutoring industry, as a derivative industry of education industry, has the same characteristics and can not solve the above contradiction. Therefore, at the present stage, it is a reasonable choice to restrict the capitalization of private tutoring industry.

## References

1. Chaoqian, C.: Research on the Spatial and Temporal Differentiation of Chinas Basic Education Resources and its Influence Factors. Xiangtan University (2019)

2. Yan, S., Kai, Z.: Impacts of family background on entrance opportunity of higher education in china: a national survey. Educ. Res. Monthly (05), 30–36 (2015)
3. Meng, D.-H., Xu, C.-X.: Education expansion and educational inequality in China: based on education Gini coefficient investigation. J. Hangzhou Normal Univ. [Humanit. Soc. Sci.] **44**(01), 50–60 (2022)
4. Chen, T.: The research of the equity between urban and rural education under the background of New Era. J. Zhongzhou Univ. **37**(03), 105–109 (2020)
5. Shuanghua, J.: The policies of fiscal expenditure for education and income distribution. Finance Trade Econ. (01), 58–62+97 (2003)
6. Yan, L.: Research on countermeasures to eliminate urban and rural education in China. J. Liaoning Radio Telev. Univ. (04), 49–50 (2011)
7. Wentao, W., Meiling, Z.: On the establishment of equitable urban and rural education mechanism. J. Inner Mongolia Agric. Univ. (Soc. Sci. Ed.) (05), 28–30 (2007)
8. Hongguang, P., Junfen, L.: Mechanism and Strategy of Promoting balanced Development of compulsory education by informatization. China Electron. Educ. (10), 33–39 (2010)
9. Gangmin, G.: Efficiency and equity analysis of Higher education cost sharing in China. Acad. China (03), 255–260 (2009)


# Literature Review of Audio-Driven 2D Avatar Video Generation Algorithms

Yuxuan Li, Han Zhang(✉), Shaozhong Cao, Dan Jiang, Meng Wang, and Weiqi Wang

Beijing Institute of Graphic Communication, Beijing, China
{zhanghan,caoshaozhong,jiangdan}@bigc.edu.cn

**Abstract.** Audio-driven 2D avatar video generation algorithms have a wide range of applications in the media field. The technology of generating 2D avatar videos with only the input of compliant audio and images has been a positive boost to the development of online media and other fields. In such generation algorithms, the accurate coupling of speech audio and appearance changes such as faces and gestures in subtle movements has been a point of continuous improvement, with appearance changes moving from an early focus on matching speech content only to starting to incorporate human emotions expressed by speech. There has been a significant improvement in fidelity and synchronization compared to the early experimental results of the study, and the behavioral performance of the 2D avatars in the generated videos is getting closer to that of humans. This paper provides an overview of existing audio-driven 2D avatar generation algorithms and classifies their tasks into two categories: talking face generation and co-speech gesture generation. Firstly, the article describes the task specifically and describes its application areas. Secondly, we analyze the core algorithms in order of technological advancement and briefly describe the performance effects of the methods or models. Thirdly, we present common datasets for both types of tasks as well as evaluation metrics and compare the performance metrics of some recently proposed algorithms. Finally, the paper discusses the opportunities and challenges faced by the field and gives future research directions.

**Keywords:** audio-driven · 2D avatar · talking face generation · co-speech gesture generation · deep learning

## 1 Introduction

As the intelligence of computer technology continues to increase, it is better applied to more areas. It is clear that the application of artificial intelligence technology to image and character recognition has increased people's productivity and improved information security. At the same time, the development of AI technology has also enriched the way people entertain themselves, playing a very important role in film production, computer

This subject was supported by the Research Project of Education Commission of Beijing (KM202110015003), Initial funding for the Doctoral Program of BIGC, and Innovation team project of BIGC (Eb202103).


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 85–96, 2023.
https://doi.org/10.1007/978-981-99-3618-2_9

games, and virtual reality. It is worth noting that AI technology cannot be studied without a large amount of data, including text, images, audio, and video. This multimodal data underpins the various directions of the research field. As research techniques based on a single modality matured, researchers began to take on the challenge of more challenging techniques -- research based on multiple modalities. We know that the face is the most representative attribute of a person's identity. And voice carries a great deal of information in everyday life, including the timbre of the speaker and the mood of the speech. This has facilitated the creation of a 2D avatar video generation task. In this task, a given audio clip and an image of a person or a picture of an avatar are used as input to generate an audio-driven video in which the changes in the face or gestures are well synchronized with the audio.

Among the common audio-driven 2D avatar video generation tasks are talking face generation, lip-syncing, and co-speech gesture generation. Of these, audio-driven talking face generation is one of the more important tasks in the field of cross-modal generation. In this task, given an audio clip and a face image, a video of a speaker with a more natural head and eye pose is generated, corresponding to the mouth movements corresponding to the audio. This task is widely used in areas such as virtual anchoring, teleconferencing, game development, and film and TV post-production. As early as 2017, creative videos of virtual cartoon images were posted by creators on the YouTube platform and generated a large number of subscriptions from viewers. It is worth mentioning that several media outlets, such as Xinhua News Agency and China Central Radio and Television, have also made use of virtual anchors to broadcast news.

Co-speech gesture generation is part of the speech body movement generation task. The simultaneous synthesis of spontaneous body movements from the input speech audio is the aim of this task. During human expression, information related to speech is often conveyed through gestures. Speakers, in particular, are adept at conveying more valuable information to their audience through gestures. In the field of social robotics and the digital body, this task has a decisive role to play. It is interesting to note that common speech gesture generation differs from the talking head generation [1] task. The lip synchronization [2] task in the talking head task has been extensively studied, with research scholars using deterministic mappings to couple lip movements to audio. Gestures, however, have no specific linguistic meaning and are implicit human behavior. It is very difficult for people to find a mapping relationship between speech and gesture.

In this paper we present an extensive survey of audio-driven 2D avatar generation algorithms, with a particular focus on deep learning-based techniques. Our goal is to help the reader navigate through the different approaches in the literature. In the first part, we present the mainstream generation algorithms for audio-driven 2D avatars. Then, we present the talking face generation method and co-speech gesture generation method in Sect. 2 and Sect. 3, respectively, and summarize their common datasets and evaluation metrics. At the end of the article, we conclude the article and propose future research directions.

## 2 Methods of Audio-Driven Talking Face Generation

The naturalness of facial expressions and contextualization of emotional expressions in the task of speech face generation has been a point of concern for research scholars. In recent years, deep learning has played an important role in the field of computer vision and natural language processing. Among the generative class of algorithms, Goodfellow et al. proposed generative adversarial networks (GAN) [3] in 2014 to accomplish a breakthrough in generating high-quality images and data enhancement. To improve the quality of generated images, GANs are widely used for cross-modal learning between audio and images.

GAN-based methods have achieved significant performance improvements compared to the initial traditional methods [4, 5] based on Hidden Markov Models (HMMs) [6]. Vougioukas et al. [7] used temporal generative adversarial networks to discriminate true from false at the video frame level and true from false at the video sequence level, and Song et al. [8] used conditional recurrent adversarial networks to encapsulate audio and images in recurrent units to achieve smooth transitions for facial and mouth movements.

In the task of talking face generation, the processing methods fall into two main categories: generating video frames directly from audio and using intermediate modal representations to connect audio input and video output.

### 2.1 Direct Mapping from Audio to Talking Face

In the early days of audio-driven talking face generation tasks, a technical approach used a model that enables direct relational mapping of audio to talking-face. Initially, research work explored this on a dataset for a single task. Suwajanakorn et al. [2] used the Obama video dataset. Firstly, the audio signal was converted to mouth sparse key points, and a sparse mouth shape was output for each output video frame. Secondly, facial texture synthesis is performed, i.e. the texture of the lower region of the face is synthesized based on the sparse mouth pattern. Finally, the video is modified and retimed to make the head movements look natural and match the given input speech.

However, in terms of societal needs, it is not enough to implement a single identity of the speaking face. Therefore most of the research work has started to explore the possibility of implementing speaking faces with multiple personas.

Chung et al. [9] designed the Speech2Vid model, containing an audio encoder, an identity image encoder, a decoder, a speaker face image decoder, and a deblurring module. Then, the method was extended by Zhou [10] through adversarial learning. Chen et al. [11] designed a method to generate multiple mouth image fusions by embedding audio sequences and identity images. Also, the authors introduced derivatives of audio features in the temporal dimension and optical flow between lip motion frames to maximize the mutuality between the two through cosine similarity loss. Zhu et al. [12] proposed a model containing a face generator, an Asymmetric Mutual Information Estimator (AMIE), and a frame discriminator incorporating cross-modal consistency. Zhou [13] et al. consider that speech can also be used as one of the tools for biometrics, and achieve higher quality results for speech face generation by decoupling the approach to

obtain pure data features. Vougioukas et al. [14] proposed a temporal generative adversarial network with Gaussian noise added to the input side of the model. This Gaussian noise is represented by a single layer of Gate Recurrent Unit (GRU) [15]. This potential representation is used to model spontaneous movements such as die blinks and eyebrow movements.

### 2.2 Connecting Audio and Output Using Intermediate Modal Representation

As it is difficult to directly map information such as head pose and facial expressions from speech signals, good results can be achieved by using intermediate modal representations to connect audio input and video output. Intermediate modal representations are generally 2D and 3D landmark representations of the face. These sets of landmarks are of lower dimensionality than the images and can better find mapping relationships with the lower dimensional audio. In summary, this type of generation algorithm often consists of two parts: prediction using audio to landmarks and landmark to face image generation.

In recent years, such algorithms are mainly based on long short-term memory (LSTM), convolutional neural network (CNN), and GAN. Talking head videos based only on LSTM or CNN models are blurred and unrealistic, and those based solely on GAN are not as good as they could be due to the unstable nature of the network output. Therefore, more and more researchers are combining LSTM, CNN, and GAN to improve the quality of the generated videos.

Eskimez et al. [16, 17] use deep neural networks to generate simultaneous facial landmarks with robust noise recovery that can generate key points of a talking face from the speech in real-time. This system uses an LSTM network and automatically extracts facial landmarks after training on 27 different frontal videos of speakers. Subsequently, Chen et al. [18] devised a GAN-based multi-stage approach to generate speaker face sequences. The method was used in the stage of audio prediction landmark, and the face key points features of the identity image that were downscaled by Principal Component Analysis (PCA) [19] were input into LSTM [20] together with the audio features to obtain the face key points features corresponding to the audio. In the landmark-to-speaker face generation stage, two attention masks are obtained by convolution and transposed convolution, enabling the generative network to learn stable pixels and move pixels to mitigate the incoherence of pixels appearing in the same region between adjacent frames. The attention mask is used to focus on the most variable parts of the face such as the lips. Greenwood et al. [21] jointly learn facial expressions and head pose landmarks from a bifurcated bidirectional LSTM network.

These audio-driven talking face generation efforts above focus on matching speech content, and identity information is usually ignored due to pattern collapse or averaging during the training process. However, different people have different facial expressions and head movement postures when speaking the same passage, although their mouth patterns should be consistent. The approach proposed by Zhou et al. [22] separates speech content features and speaker identity information features from the input audio signal and drives the capture of landmarks of speaker-related dynamics. The method utilizes 68 face landmarks as intermediate modalities to link speech to 2D animation mapping.

The method uses the speech transformation neural network AutoVC proposed by Qian et al. [23] to decompose speech content and identity information. The speech content animation module models the sequential dependencies of the output landmarks based on the following transformations:

$$c_t = LSTM_C(A_t \rightarrow_{t+\tau}; w_{lstm,c}), \tag{1}$$

$$\Delta q_t = MLP_C(c_t, q; w_{mlp,c}), \tag{2}$$

$$p_t = q + \Delta q_t, \tag{3}$$

where $w$ is the network learnable parameter. In this model, the LSTM has three layers of cells, each with an internal hidden state vector of size 256. The decoder MLP network has three layers with internal hidden state vectors of sizes 512, 256 and 204 ($68 \times 3$) respectively.

As it is very unnatural to move only the lips in facial expressions, head movements such as eyebrows are also needed to create subtle associations with the mouth. In the speaker-aware animation model structure, landmarks are perturbed to match the head movements and personalized expressions observed during training. The expressions for this part are as follows:

$$h_t = Attn_s(\tilde{c}_{t\rightarrow t+\tau'}, s; w_{attn,s}), \tag{4}$$

$$\Delta p_t = MLP_s(h_t, q; w_{mlp,s}), \tag{5}$$

$$y_t = p_t + \Delta p_t, \tag{6}$$

$\{w_{attn}, s, w_{mlp}, c\}$ is a trainable parameter for the self-attentive encoder and MLP decoder, $p_t$ and $y_t$ are landmarks that capture the speech content and the identity of the speaker for each frame.

MakeItTalk outputs video at a resolution of 256*256, which limits the scenarios in which it can be used. And this method does not separate the portrait from the rear view, and the background will follow as it is driven. In the case of complex portrait backgrounds, wireframes around the head will be evident in the generated video.

Other approaches using facial landmarks have been used in recent research [24]. Next, inspired by the development of neural rendering techniques, a NeRF [25] based talking head generation technique was proposed. Notably, Wav2Lip [26] synchronizes mouth and audio for internally-coated reconstruction, and PC-AVS [27] learns pose and lip reconstruction through implicit modulation. FOMM [28] extended Monkey-Net by a first-order local affine transformation. Recent work has also used motion flow fields [29] for topic-independent methods. Monkey-Net [30] proposes a network that transfers the deformation of motion flows from sparse to dense. In summary, audio-driven talking face generation tasks seek to generate high-quality video footage and near-realistic facial expressions and head movements, and current research methods, while gradually improving, still have many limitations.

### 2.3 Dataset

Most of the current datasets associated with the talking face task are multi-person datasets. Table 1 compares the number of categories, duration, and time of presentation for each dataset.

The LRW dataset is a large outdoor audiovisual dataset proposed by Chung et al. [31]. The dataset was collected from BBC television footage and contains 1000 h of video files with clear spoken words, over 1000 different words, and over 1000 speaking characters. The authors selected the 500 most frequent words between 5 and 10 characters in length as dictionary labels and divided the dataset into three parts: a training set, a validation set, and a test set. The VoxCeleb dataset, a large outdoor audiovisual dataset proposed by Nagrani et al. [33], covers various ethnicities, regions, and age groups to ensure sample diversity in the dataset. Also, the scenes in which the speaking characters are located vary. The HDTF dataset [] is a collection of speaker videos from the last two years on Youtube, the videos themselves are at 720P or 1080P resolution and cropped to 512 * 512 faces.

**Table 1.** Summary of Data Sets Related to the Talking Face Task

| Dataset | Concrete information | | |
|---|---|---|---|
| | *Number of categories* | *Duration* | *Time of presentation* |
| GRID[32] | 34 | 27.5 | 2006 |
| LRW[31] | >1000 | 173 | 2016 |
| ObmaSet[2] | 1 | 14 | 2017 |
| VoxCeleb[33] | 1251 | 352 | 2017 |
| VoxCeleb2[34] | 6112 | 2400 | 2018 |
| RAVDESS[35] | 24 | 7 | 2018 |
| LRS2-BBC[36] | >500 | 224.5 | 2018 |
| LRS3-TED[37] | >5000 | 438 | 2018 |
| HDTF[29] | 362 | 15.8 | 2021 |
| GLips[38] | 250,000 | - | 2022 |

### 2.4 Evaluation Metric

In order to assess how effective the overall head movement, facial expressions, and dynamics are, a quantitative evaluation of head posture prediction was performed using the following metrics proposed by Zhou [22]:

- Landmark distance (D-L): the average Euclidean distance between all predicted facial landmark locations and the reference location.
- Landmark velocity difference (D-V): the average Euclidean distance between the reference landmark velocity and the predicted landmark velocity.

- Head rotation and position difference (D-Rot/Pos): The average difference between the reference and predicted head rotation angles and the head position.

In the field of 2D talking face generation, researchers have the choice of SSIM and FID to measure the image-level visual quality of synthetic video frames and CPBD to assess sharpness:

- SSIM is a perception-based model that treats image degradation as a perceptual change in structural information, while also incorporating important perceptual phenomena such as luminance masking and contrast masking.
- Heusel et al. [39] introduced FID, which uses a specific layer of InceptionNet [40] to compute latent features of the input image. FID measures the distance between the synthetic data distribution and the real data distribution, in line with the results of the human perceptual evaluation.
- In order to measure the loss of sharpness during generation, we use a non-reference measure, cumulative probabilistic blur detection (CPBD) [41]. CPBD integrates the concept of cumulative probabilistic blur detection with the concept of visible-only blur into a probability and model to assess the sharpness of an image from a perceptual perspective.

Several methods from recent years are described below using the above three evaluation metrics under the LRW dataset. It is better if the SSIM and CPBD are high, and the FID is low (Table 2).

**Table 2.** Performance Comparison of Several Methods on the LRW Dataset

| Method | Evaluation on LRW | | |
|---|---|---|---|
| | *SSIM*↑ | *CPBD*↑ | *FID*↓ |
| Wiles et al.[42] | 0.63 | 0.22 | 206.16 |
| Wang et al.[43] | 0.35 | 0.28 | 239.40 |
| Chen et al.[18] | 0.38 | 0.07 | 189.59 |
| Zakharov et al.[45] | 0.42 | 0.11 | 105.97 |
| Jamaludin et al.[9] | 0.34 | 0.21 | 197.05 |

## 3 Methods of Co-speech Gesture Generation

In an audio-driven 2D avatar task, a good co-speech gesture generation algorithm will make the avatar's behavior seem more natural. However, the need for research scholars to establish a mapping between abstract communication intentions and the physical realization of gestures is not an easy task. In our everyday communication, different people gesture differently when we say the same sentence, and not every sentence has to be accompanied by a gesture, and even the same sentence can be gestured differently in different scenarios. Therefore, the input audio does not provide enough information to

specifically identify a gesture sequence. How to generate natural and stylized gestures has become the focus of research scholars.

### 3.1 Research Progress

Early common speech gesture synthesis tasks were mainly rule-based implementations [45–47], but since gestures are implicit human behavior, the results are very unnatural. The intuitive approach does not capture micro-scale motion and cross-modal information. Moreover, inconsistent biased mappings between gestures and the input audio are produced, and in recent approaches, adversarial learning methods [48–50] are often used to narrow the gap between generated and real by using discriminators. Later, research scholars used deep networks to infer body signs directly from the input audio in order to reduce the impact of such non-deterministic mappings on the task [1, 48–50, 52–54].

Among such algorithms, some methods rely on a set of predefined gestures [52] and others on individual style-specific conditions [48, 50], and Qian et al. [53] remove the ambiguity of collaborative speech gesture synthesis through the learning of template vectors. However, these solutions have limitations in terms of movement diversity and fidelity. Therefore, it is quite important to develop algorithms that model the non-deterministic mapping between speech audio and gestures.

The current method proposed by Qian et al. [53] enhances the fidelity and diversity of gesture synthesis without sacrificing synchronization quality. They first converted the audio clip into a Mel spectrogram following previous methods [48, 55]. We then sent it to an audio encoder to obtain audio features $\mathrm{A} \in \mathbb{R}^{256\times F}$, $F$ as frames. As another input, the network receives the template vector and stacks copies of it into the template feature $\mathrm{T} \in \mathbb{R}^{C\times F}$, with the aim of aligning the timeline of the audio features. The complete input to our model is therefore $[\mathrm{A}|\mathrm{T}] \in \mathbb{R}^{(256+C)\times F}$, the stitching of the audio feature with the template feature T.

The output is a sequence of gestures $G \in \mathbb{R}^{2K\times F}$, Where $2K$ corresponds to the 2D coordinates of the $K$ upper body key points in a frame. The algorithm applies the $L_1$ regression loss on the regression gesture sequence $G$ with the following expression:

$$L_{reg} = \frac{1}{F}\sum_{i=1}^{F}\left\|G^{(i)} - \hat{G}^{(i)}\right\|_1, \tag{7}$$

This equation defines $G^{(i)}$ as the predicted vector for frame i, and $\hat{G}^{(i)}$ as the ground truth vector.

### 3.2 Dataset

The PATS dataset [49] consists of diverse and heavily aligned pose, audio and text recordings to help generate natural and relevant gestures for virtual agent technology. The dataset contains up to 251 h of audio in 25 different styles. The Speech2Gesture dataset [48] contains speaker-specific videos of television anchors. However, most of the videos are television programs and have noisy background sounds. Body movements are also limited as the speaker is often sitting in a chair or leaning on a table. The Ted Gesture dataset [51] contains videos of multiple speakers with simpler gestures.

### 3.3 Evaluation Metric

The synchronization between the generated gestures and the speech audio is one of the key factors in measuring how well the model captures the relationship between speech and action. Landmark velocity difference (LVD) [22] is used here to assess the synchronization with speech. To measure the quality of the generated gestures, Xu et al. [56] trained a binary classifier to distinguish between real and fake samples and used the prediction scores on the test set as the quality of the generated movements [57]. Table 3 compares the performance of recent models on the Speech2Gesture dataset.

**Table 3.** Performance Comparison of Models on the Speech2Gesture Dataset

| Model | Evaluation on Speech2Gesture | |
|---|---|---|
| | *LVD*↓ | *Quality*↑ |
| Audio2Body[58] | 14.1 | 0.375 |
| TriCon[50] | 7.8 | 0.278 |
| S2G[48] | 7.2 | 0.163 |
| Mix-StAGE[49] | 10.6 | 0.382 |
| Tmpt[53] | 7.3 | 0.178 |
| FreeMo[56] | 6.7 | 0.502 |

In addition, in the approach of Qian et al. [53], conducted an extensive human study to perceptually compare their method with a baseline model. They used different methods to generate gesture sequences from the same speech clips and then published them as an online questionnaire for human assessment. After viewing each video, those who participated in the scoring were asked to select the video with the best quality of gesture synchronization and the video with the most natural gestures.

## 4 Conclusion and Future Work

This paper presents a study related to audio-driven 2D avatar generation algorithms. In the talking face generation task, there has been a move from direct mapping of audio to the talking face to the use of intermediate modalities to make the output video more natural. In the co-speech gesture generation task, there has been a move from the early days when only a single gesture could be generated to the present day when gestures can be generated with variety and fidelity. The performance of the algorithms for tasks related to virtual humans has been continuously optimized and improved. However, a number of issues and challenges remain, and researchers continue to pursue the generation of higher resolution videos with natural facial expressions and a variety of smooth gestures. In recent years, 2D-avatar-related technology based on the research, generating 3D avatars related technology gradually appears, and gradually has the advantage of generating natural facial expressions. This is also one of the future research directions. Today,

audio-driven avatar technology is widely used in various fields, making sessions such as book sharing and program interaction interesting. This technology will have more scope for development in the future.

## References


1. Chen, L., Cui, G., Kou, Z., Zheng, H., Xu, C.: What comprises a good talking-head video generation?: a survey and benchmark. arXiv preprint arXiv:2005.03201 (2020)
2. Suwajanakorn, S., Seitz, S.M., Kemelmacher-Shlizerman, I.: Synthesizing Obama: learning lip sync from audio. ACM Trans. Graph. (ToG) **36**(4), 1–13 (2017)
3. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: International Conference on Learning Representations (2016)
4. Xie, L., Liu, Z.-Q.: A coupled HMM approach to video-realistic speech animation. Pattern Recogn. **40**(8), 2325–2340 (2007)
5. Simons, A.D.: Generation of mouthshape for a synthetic talking head. Inst. Acoust. (1990)
6. Rabiner, L.R., Juang, B.-H.: An ntroduction to hidden Markov models. IEEE ASSP Mag. Citeseer **3**(1), 4–16 (1986)
7. Vougioukas, K., Petridis, S., Pantic, M.: End-to-end speech-driven facial animation with temporal GANs. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 37–40 (2019)
8. Song, Y., Zhu, J., Li, D.: Talking face generation by conditional recurrent adversarial network. In: International Joint Conference on Artificial Intelligence, pp. 919–925 (2019)
9. Chung, J.S., Jamaludin, A., Zisserman, A.: You said that?: synthesising talking faces from audio. Int. J. Comput. Vision **127**, 1767–1779 (2017)
10. Zhou, H., Liu, Y., Liu, Z., Luo, P., Wang, X.: Talking face generation by adversarially disentangled audio-visual representation. In: AAAI (2019)
11. Chen, L., Li, Z., Maddox, R.K., Duan, Z., Xu, C.: Lip movements generation at a glance. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 538–553. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_32
12. Zhu, H., Zheng, A., Huang, H.: High-resolution talking face generation via mutual information approximation. arXiv preprint arXiv:1812.06589 (2018)
13. Zhou, H., Liu, Y., Liu, Z.: Talking face generation by adversarially disentangled audio- visual representation. In: AAAI Conference on Artificial Intelligence, vol. 33, pp. 9299–9306 (2019)
14. Vougioukas, K., Petridis, S., Pantic, M.: Realistic speech-driven facial animation with GANs. Int. J. Comput. Vision **128**(5), 1398–1413 (2020)
15. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
16. Eskimez, S.E., Maddox, R.K., Xu, C., Duan, Z.: Generating talking face landmarks from speech. In: Proceedings of LV A/ICA (2018)
17. Eskimez, S.E., Maddox, R.K., Xu, C., Duan, Z.: Noise-resilient training method for face landmark generation from speech. IEEE/ACM Trans. Audio Speech Lang. Process. (2019)
18. Chen, L., Maddox, R.K., Duan, Z., Xu, C.: Hierarchical cross-modal talking face generation with dynamic pixel-wise loss. In: Proceedings of CVPR (2019)
19. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemom. Intell. Lab. Syst. **2**(1–3), 37–52 (1987)
20. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 1735–1780 (1997)
21. Greenwood, D., Matthews, I., Laycock, S.: Joint learning of facial expression and head pose from speech. In: Interspeech (2018)

22. Zhou, Y., Han, X., Shechtman, E., Echevarria, J., Kalogerakis, E., Li, D.: Makelttalk: speaker-aware talking-head animation. TOG (2020)
23. Qian, K., Zhang, Y., Chang, S., Yang, X., Hasegawa-Johnson, M.: AUTOVC: zero-shot voice style transfer with only autoencoder loss. In: Proceedings of ICML, pp. 5210–5219 (2019)
24. Song, L., Wu, W., Fu, C., Qian, C., Loy, C.C., He, R.: Everything's Talkin': pareidolia face reenactment. arXiv preprint arXiv:2104.03061 (2021)
25. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRf: Representing scenes as neural radiance fields for view synthesis. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12346, pp. 405–421. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58452-8_24
26. Prajwal, K.R., Mukhopadhyay, R., Namboodiri, V.P., Jawahar, C.V.: A lip sync expert is all you need for speech to lip generation in the wild. In: ACM Multimedia (2020)
27. Zhou, H., Sun, Y., Wu, W., Loy, C.C., Wang, X., Liu, Z.: Pose-controllable talking face generation by implicitly modularizedaudio-visual representation. In: CVPR (2021)
28. Siarohin, A., Lathuilière, S., Tulyakov, S., Ricci, E., Sebe, N.: First order motion model for image animation. In: NIPS (2019)
29. Zhang, Z., Li, L., Ding, Y., Fan, C.: Flow-guided one-shot talking face generation with a high-resolution audio-visual dataset. In: CVPR (2021)
30. Siarohin, A., Lathuilière, S., Tulyakov, S., Ricci, E., Sebe, N.: Animating arbitrary objects via deep motion transfer. In: CVPR (2019)
31. Chung, J.S., Senior, A.W., Vinyals, O., et al.: Lip reading sentences in the wild. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3444–3453 (2017)
32. Cooke, M., Barker, J., Cunningham, S., et al.: An audio-visual corpus for speech perception and automatic speech recognition. Acoust. Soc. Am. **120**, 2421–2424 (2006)
33. Nagrani, A., Chung, J.S., Zisserman, A.: Vox Celeb: a large-scale speaker identification dataset. Comput. Speech Lang. 101027 (2017)
34. Chung, J.S., Nagrani, A., Zisserman, A.: Vox Celeb2: deep speaker recognition. In: Interspeech, pp. 1086–1090 (2018)
35. Livingstone, S.R., Russo, F.A.: The ryerson audio-visual database of emotional speech and song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in North American English. PloS ONE e0196391 (2018)
36. Afouras, T., Chung, J.S., Senior, A., et al.: Deep audio-visual speech recognition. IEEE Trans. Pattern Anal. Mach. Intell. (2018)
37. Afouras, T., Chung, J.S., Zisserman, A.: LRS3-TED: a large-scale dataset for visual speech recognition. arXiv preprint arXiv:1809.00496 (2018)
38. Schwiebert, G., Weber, C., Qu, L., Siqueira, H., Wermter, S.: A multimodal German dataset for automatic lip reading systems and transfer learning. arXiv:2202.13403
39. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: Advances in Neural Information Processing Systems, pp. 6626–6637 (2017)
40. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
41. Narvekar, N.D., Karam, L.J.: A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection. In: 2009 International Workshop on Quality of Multimedia Experience, pp. 87–91. IEEE (2009)
42. Wiles O, Sophia Koepke A, Zisserman A (2018) X2face: network for controlling face generation using images, audio, and pose codes. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11217, pp. 670–686. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01261-8_41

43. Wang, T.C., Liu, M.Y., Tao, A., Liu, G., Kautz, J., Catanzaro, B.: Few-shot video-to-video synthesis. arXiv preprint arXiv:191012713 (2019)
44. Zakharov, E., Shysheya, A., Burkov, E., Lempitsky, V.: Few-shot adversarial learning of realistic neural talking head models. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 9459–946 (2019)
45. Cassell, J., et al.: Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, pp. 413–420 (1994)
46. Kopp, S., et al.: Towards a common framework for multimodal generation: the behavior markup language. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 205–217. Springer, Heidelberg (2006). https://doi.org/10.1007/11821830_17
47. Wagner, P., Malisz, Z., Kopp, S.: Gesture and speech in interaction: an overview. Speech Commun. **57**, 209–232 (2014)
48. Ginosar, S., Bar, A., Kohavi, G., Chan, C., Owens, A., Malik, J.: Learning individual styles of conversational gesture. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3497–3506 (2019)
49. Ahuja, C., Lee, D.W., Nakano, Y.I., Morency, L.-P.: Style transfer for co-speech gesture animation: a multi-speaker conditional-mixture approach. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds,) ECCV 2020. LNCS, vol. 12363, pp. 248–265. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58523-5_15
50. Yoon, Y., et al.: Speech gesture generation from the trimodal context of text, audio, and speaker identity. ACM Trans. Graph. (TOG) **39**(6), 1–16 (2020)
51. Yoon, Y., Ko, W.-R., Jang, M., Lee, J., Kim, J., Lee, G.: Robots learn social skills: end-to-end learning of co-speech gesture generation for humanoid robots. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 4303–4309. IEEE (2019)
52. Liao, M., Zhang, S., Wang, P., Zhu, H., Zuo, X., Yang, R.: Speech2Video synthesis with 3D skeleton regularization and expressive body poses. In: Proceedings of the Asian Conference on Computer Vision (2020)
53. Qian, S., Tu, Z., Zhi, Y., Liu, W., Gao, S.: Speech drives templates: co-speech gesture synthesis with learned templates. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 11077–11086 (2021)
54. Yi, R., Ye, Z., Zhang, J., Bao, H., Liu, Y.-J.: Audio-driven talking face video generation with learning-based personalized head pose. arXiv preprint arXiv:2002.10137 (2020)
55. Alexanderson, S., Henter, G.E., Kucherenko, T., Beskow, J.: Style-controllable speech-driven gesture synthesis using normalising flows. In: Computer Graphics Forum, vol. 39, pp. 487–496. Wiley (2020)
56. Xu, J., Zhang, W., Bai, Y., Sun, Q., Mei, T.: Freedom body motion generation from speech. arXiv preprint arXiv:2203.02291 (2022)
57. Aliakbarian, S., Saleh, F.S., Salzmann, M., Petersson, L., Gould, S.: A stochastic conditioning scheme for diverse human motion prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5223–5232 (2020)
58. Shlizerman, E., Dery, L., Schoen, H., Kemelmacher-Shlizerman, I.: Audio to body dynamics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7574–7583 (2018)

# Research on the Mode and Framework of College Entrance Examination and Undergraduate Smart Teaching in the Environment of Smart Media Integration

Teng Zhang and Liang Wang(✉)

School of Economics and Management, Beijing Institute of Graphic Printing, Beijing, China
wangliang@bigc.edu.cn

**Abstract.** Starting from the questionnaire survey and statistical analysis of the current situation of postgraduate entrance examination and professional learning in colleges and universities, combined with the policy of training talents in higher education in my country and the background of the integration and development of intelligent media, this paper comprehensively analyzes the current situation between postgraduate entrance examination and professional learning in colleges and universities. It is difficult to take into account the problems of professional learning, ignoring the opportunities for undergraduate teaching practice, the dilemma of choice after failing the postgraduate entrance examination, and the serious mental health problems of students. It also proposes a smart teaching model for postgraduate entrance examination and undergraduate study [1], in order to better serve postgraduate entrance examination students and colleges and universities, improve the quality of undergraduate teaching classrooms, and provide decision-making basis for the improvement of school-running quality.

**Keywords:** Smart Media · University Postgraduate Examination · Smart Teaching

## 1 Introduction

With the rapid development of the national economy and scientific level, the smart media environment has created conditions for promoting smart teaching in the education industry. Higher education has gradually entered the stage of normalization, and the importance of academic qualifications to employment has become more and more prominent. An important choice for undergraduates after graduation. However, colleges and universities have not seriously dealt with the contradiction between postgraduate entrance examination and undergraduate teaching, resulting in the increasingly severe impact of postgraduate entrance examination on undergraduate teaching. Therefore, it

This work is partially supported by Key Teaching Reform Research Project of Beijing Institute of Graphic Communication.


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 97–108, 2023.
https://doi.org/10.1007/978-981-99-3618-2_10

is necessary to build a complete scientific and intelligent teaching model framework for postgraduate entrance examination and undergraduate study. The construction of smart teaching classrooms in colleges and universities provides new ideas and new methods, and provides a reference for the implementation of smart teaching in the whole process of pre-class, in-class and after-class. It is of great significance for the cultivation of intelligent talents to make colleges and universities better integrate the development of postgraduate entrance examination and undergraduate teaching.

## 2 The Current Situation of Postgraduate Entrance Examination and Undergraduate Teaching in Colleges and Universities

### 2.1 There is a Large Demand for Postgraduate Entrance Examinations in Colleges and Universities

With the country's emphasis on higher education talents, the cultivation of comprehensive talents and the increase in the demand for high-level talents, students' desire to improve their academic qualifications and abilities has gradually increased. It can be seen from Fig. 1 that the biggest motivation for college entrance examinations is personal life planning, accounting for 70.94%; followed by expanding the scope of employment options, accounting for 66.07%; in addition, family expectations, yearning for famous schools, and conforming to the crowd for entrance examinations are all driven to join the postgraduate entrance examination team. Colleges and universities are the main force for postgraduate entrance examinations. At the same time, the volatility of the new crown epidemic in recent years has increased the demand for postgraduate entrance examinations in colleges and universities, and the long-term continuous growth of the postgraduate entrance examination rate in colleges and universities has affected the quality of undergraduate teaching to a certain extent.

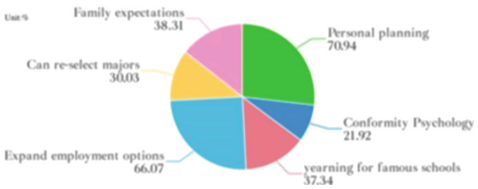


**Fig. 1.** Reasons for college students to choose postgraduate entrance exams.

### 2.2 The Preparation Time of Colleges and Universities is Longer

College students generally start preparing for postgraduate exams in their junior year. The preparation period is one and a half years, or even as long as two or three years. The preparation front is long. As shown in Fig. 2, from the time of preparing for the postgraduate entrance examination, students spend 3–6 h or 6–10 h on average every day for studying knowledge related to postgraduate entrance examination, accounting

for 35.06%, and 1–3 h for studying 20.78%. As can be seen from Fig. 3, when there is a conflict between undergraduate professional study and preparation for entrance exams, students spend 1–3 h studying professional courses accounting for 54.06%, and 3–6 h studying majors account for 25.32%. It can be seen that students not only have a long preparation time for the exam and daily review time, but also spend as much time as possible on reviewing for the postgraduate entrance examination without affecting the achievement of graduation requirements, while courses or courses that are not related to postgraduate entrance examination will be used as much as possible. More or less will become their burden, which will affect their enthusiasm and initiative in learning.


**Fig. 2.** The average daily time spent on studying for postgraduate entrance exams during college exam preparation.

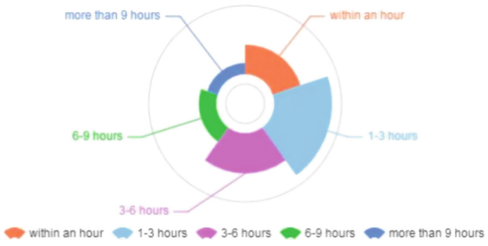


**Fig. 3.** The daily time spent studying for postgraduate entrance examination when there is a conflict between university postgraduate entrance examination and professional course learning.

### 2.3 The Employment Rate of Postgraduate Students is Low

According to a report by the Ministry of Education on December 28, 2021, the number of college graduates in the 2022 class is estimated to reach 10.76 million, a year-on-year increase of 1.67 million[①]. The scale of college graduates increases employment pressure. In addition, after the outbreak of the new crown pneumonia epidemic, the economy has suffered a severe downturn, layoffs in major industries have been large, and the employment situation has become more complicated and severe, and most of the graduate students will miss the opportunity to recruit in the autumn. You are less likely to find your

ideal job accurately. However, due to the harsh employment environment and the fact that factors such as settlement in big cities, starting salary for academic qualifications or talent subsidies are directly linked to higher education, many postgraduate students will choose to join the team to prepare for the test after they can't find the ideal job. At the same time, due to the epidemic, some students who originally wanted to study abroad will eventually choose to join the postgraduate entrance examination army.

### 2.4 The Success Rate of College Entrance Examination is not High

According to the statistics of the Ministry of Education, in 2022, 4.57 million people will apply for the postgraduate examination, 1.11 million people will be admitted, and the registration ratio will be 24.3%[②]. It can be seen from the above that although the Ministry of Education has proposed measures to expand the scale of postgraduate enrollment in order to promote employment, every year postgraduate students The number of registrations and the increase in the number of registrations have reached new heights, resulting in a continuous decline in the number of registrations every year. The proportion of graduates in postgraduate entrance examinations is increasing year by year. Factors such as the high competitiveness of postgraduate entrance examinations, the high psychological pressure on students to postgraduate entrance examinations, and the fact that postgraduate entrance examination goals do not meet their own conditions have resulted in a low success rate of postgraduate entrance examinations; it is difficult for more students to reach the first place. Most of the students entered the transfer team due to the voluntary application scores, which led to the full number of transfer students in the second district colleges and universities, and the admission colleges put forward new requirements for students' undergraduate colleges. Such repetitions make it even more difficult for students to succeed in postgraduate entrance examinations.

### 2.5 Postgraduate Entrance Examination Students Are Under Great Psychological Pressure

In terms of sleep quality of students during the test preparation period, 57.63% of the students have occasional insomnia, and 14.61% of the students have frequent insomnia. Generally approaching the postgraduate entrance examination, students' insomnia will be more serious, indicating that during the preparation period, students' mental health is worthy of the attention and guidance of the school, family and friends, as shown in Fig. 4. In addition, in terms of students giving up hobbies or habits during exam preparation, 24.51% of the students temporarily gave up the habit of reading leisurely books and accounted for the largest proportion, followed by 23.7% of students temporarily giving up sports activities, 15.42% of students would choose to temporarily give up literary and artistic activities, etc. Interests and hobbies, as shown in Fig. 5. In order to have more time to prepare for the final battle, the postgraduate students give up more entertainment, hobbies or social activities for the postgraduate entrance examination, which shows that the postgraduate entrance examination is under a lot of psychological pressure, and they think that the postgraduate entrance examination is equivalent to another college entrance examination, which is mainly based on personal initiative. It is a composite test of multiple factors such as personal learning attitude, learning ability, self-discipline,

management ability and psychological quality. Under the high-intensity test, students are prone to anxiety, insomnia, depression and dryness and other realistic manifestations. Therefore, schools, teachers and parents need to pay more attention to students' mental health problems.

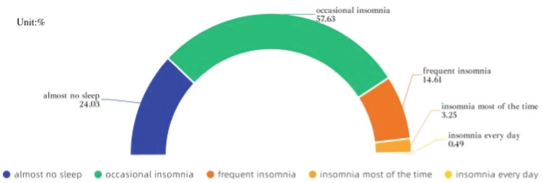


**Fig. 4.** Sleep quality of students during the postgraduate entrance examination in colleges and universities.

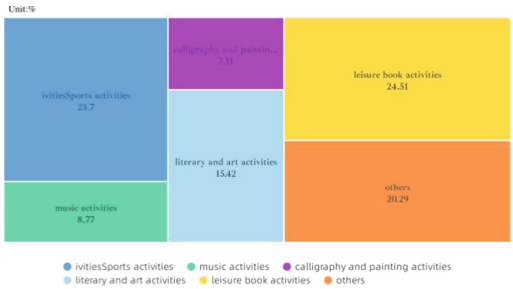


**Fig. 5.** The hobbies or habits that college students give up for the postgraduate entrance examination.

## 3 The Problems of Postgraduate Entrance Examination and Undergraduate Teaching in Colleges and Universities

### 3.1 It is Difficult to Balance Postgraduate Entrance Examination and Undergraduate Teaching

The fact that students apply for training courses during the preparation period greatly affects the learning of this professional course and takes into account the difficulties. The survey data shows that 44.97% of students choose to apply for training courses during the preparation period, 25.78% of students are still hesitant, and 29.25% of students will not apply, indicating that postgraduate entrance examination training courses to a certain extent distract students from studying other courses. Time, more difficult postgraduate and undergraduate teaching time. In addition, the analysis of the factors of not applying for the training class shows that the top factor is due to personal financial pressure, accounting for 43.83%, and 26.62% of the students believe that the role and effectiveness of applying for the remedial class is not sufficient. Large, 13.31% of the students have limited personal time and energy. It can be seen from the above that although

economic problems can affect whether students apply for remedial classes, the more intuitive factors are the role of remedial classes, the limited personal time and energy and other factors that determine whether students apply for remedial classes. Therefore, for students who apply for remedial classes, it is prone to conflict between undergraduate professional course learning and remedial class training courses. As a result, in the case of conflicts between professional courses and postgraduate examination subjects, many students will skip classes to attend remedial class training courses; Students who do not apply for the postgraduate entrance examination are prone to review the subjects related to the postgraduate entrance examination courses in the study of professional courses due to factors such as tight time for the postgraduate entrance examination and high pressure. Learning, which seriously affects the quality of undergraduate classroom teaching (Fig. 6).


**Fig. 6.** Reasons why college students do not apply for postgraduate remedial classes.

### 3.2 Ignoring Opportunities for Undergraduate Teaching Practice

The vast majority of students ignore the opportunities for undergraduate teaching practice due to preparation for exams, and ignoring undergraduate study should be a great opportunity to cultivate personal comprehensive qualities and compound talents. According to the survey data of students' dissatisfaction with the four-year undergraduate study status due to postgraduate entrance examination, it can be seen in Fig. 7 that 46.75% of the students gave up internship and employment opportunities, followed by 43.18% of the students who gave up the cultivation of hobbies and hobbies. 41.88% and 28.41% of students earnestly study professional courses and give up participating in clubs and other student work organizations, respectively. It can be seen from this that candidates who are preparing for postgraduate entrance exams can accumulate social experience


**Fig. 7.** The content of college students' dissatisfaction with the four-year undergraduate study status due to postgraduate entrance examination

and cultivate practical ability by ignoring undergraduate internships; ignoring hobbies or specialties can contribute to self-development; ignoring undergraduate professional courses can increase knowledge and skills and improve professional quality; Student work can improve personal communication ability and practical operation ability, etc., and less attention is paid to undergraduate teaching practice opportunities.

### 3.3 The Dilemma of Choosing After Failing the Postgraduate Entrance Examination

Postgraduate students spend more time studying courses related to the postgraduate entrance examination during the preparation period, which will inevitably have a certain impact on employment after failing the postgraduate entrance examination. Employers pay more attention to the fact that the interviewee has certain practical work ability and potential, has a basic way of life, can quickly adapt to the working environment and work rhythm, integrate into the group as soon as possible, and has certain professional quality and comprehensive ability[2].Some students prepare for the entrance examination and postgraduate entrance exams, which will lead to the shallow learning of undergraduate professional courses, ignoring the opportunities for undergraduate practice inside and outside the school, and giving up more opportunities to cultivate professional skills and comprehensive abilities. This phenomenon is inevitably contrary to the talents required by employers. As a result, some students are caught in a dilemma after failing the postgraduate entrance examination. Students not only lose the opportunity to continue their studies, but also cannot choose a more desirable job. They are caught in a dilemma, and they are not sure whether to continue studying for exams or whether to work, which will inevitably affect their physical and mental health.

### 3.4 Students with Serious Mental Health Problems

Under the learning pressure of fierce competition, students preparing for exams are more prone to stress, anxiety and depression, and students have serious mental health problems. Some students choose the postgraduate entrance examination due to herd mentality; they do not have a clear self-goal orientation, do not have a clear future planning direction, and blindly choose the postgraduate entrance examination; some students have insufficient preparation for the postgraduate entrance examination. The test results are unsatisfactory due to the unsystematic and comprehensive scope of the test; some students fall into self-doubt, anxiety, sensitivity to interpersonal relationships, paranoia, and depression due to factors such as high academic pressure, social difficulties, poor sleep quality, and social pressure during test preparation etc. status. Due to some unhealthy psychological phenomena of college students in the postgraduate entrance examination, students are skeptical about self-choice and difficult to evaluate themselves correctly and objectively, resulting in more frustration and regret in the final choice, and students' mental health problems are more serious. Therefore, we must pay attention to their mental health, guide them correctly, and educate them carefully, so that students can correctly understand the value of society, life and learning from their own perspective, and guide students to develop healthily and upwardly.

# 4 The Smart Teaching Mode of Postgraduate Entrance Examination and Undergraduate Teaching in Colleges and Universities

## 4.1 The Basis for the Construction of the Model

The construction of a smart teaching model in the smart media environment should conform to the actual situation of postgraduate entrance examination and undergraduate teaching in colleges and universities and the development of artificial intelligence technology. The first is to promote the two-way development of postgraduate entrance examinations and undergraduate teaching in colleges and universities. The main purpose of smart teaching is to not affect the quality of undergraduate teaching due to postgraduate entrance examinations, improve the quality of intelligent talent training in colleges and universities, and promote college students to truly become society. Useful pillars. The second is guided by the basic concept of smart education, which emphasizes the design of educational situations, takes the scholars themselves as the center, makes full use of various high-quality information resources, promotes collaboration and communication between teachers and students, and evaluates the learning process, so as to provide correct feedback on learning. Quality and efficiency [3]. The third is based on the subject teaching model, based on the virtuous cycle design before class, during class and after class, referring to the foreign 7e model: import engage, explore explore, explain explain, elaborate elaborate, expand extend, exchange, evaluate, and provide theoretical basis for smart teaching model [4].

The frame structure design of this paper mainly adopts the virtuous cycle design before class, during class and after class from the perspectives of students and teachers, constructs a basic smart teaching model framework, and designs new framework learning ideas from the perspectives of postgraduate entrance examination and undergraduate study, not only taking into account students It also meets the basic requirements of undergraduate teaching and promotes the cultivation of diversified and comprehensive talents in colleges and universities (Fig. 8).

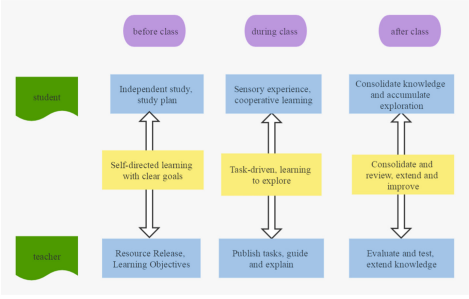


**Fig. 8.** The basic model framework of smart teaching for students.

## 4.2 Interpretation of model results

### 4.2.1 The Basic Model of Students' Smart Teaching

a) *Self-directed learning with clear goals*

Before class, teachers carefully design preview materials based on the learning situation analysis results, and push them to students through electronic whiteboards or online learning apps in a timely manner; Make preview materials and complete the independent preview within the specified time[5].Teachers can check students' preview status, preview progress, and answering status in a timely manner on the terminal, which are visually presented in the form of data or tables, providing preparation conditions for the orderly development of the classroom.

b) *Task-driven, learning and exploration*

In class, teachers can enhance students' sensory experience and stimulate students' interest in learning about classroom problems through introduction of situations in the form of film and television animation scenes, pre-class preview of difficult and difficult problems, and interesting activities. Then teachers guide students to explore problems independently in a task-driven way., in the form of classroom guidance and in-class testing, so that students can master classroom knowledge [6]. Sometimes, the exchange of roles between teachers and students and cooperative learning can deepen students' understanding and mastery of knowledge, and it is more conducive to the cultivation of students' innovative ability.

c) *Consolidate and review, extend and improve*

After class, teachers can not only consolidate and extend knowledge, but also increase students' accumulation and exploration of knowledge by issuing comprehensive, practical and diverse testing tasks. The practical questions can be uploaded to the terminal and shared with the students in time to learn from each other and make progress together. After teachers grasp the differences in individual learning needs, they can push relatively personalized review materials according to different students, teach students according to their aptitude, and guide learning in a personalized way (Fig. 9).

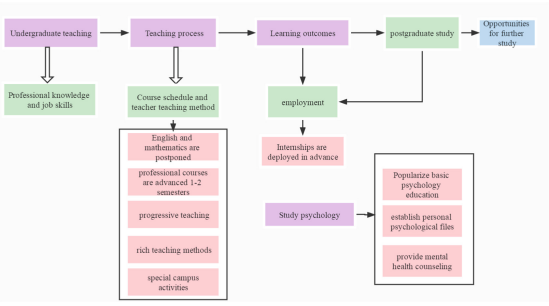


**Fig. 9.** Special model framework for postgraduate entrance examination and undergraduate smart teaching.

### 4.2.2 Special *Model* for Postgraduate Entrance Examination and Undergraduate Smart Teaching

d) *Undergraduate teaching objective*

Institutions of higher learning are the foundation for students to achieve higher education and an important source of talents for training vocational education. Undergraduate teaching should not only cultivate first-class compound talents with professional knowledge and professional skills, but also enable students to apply their professional expertise in work or life. Therefore, the goal of undergraduate teaching is an extremely important point for colleges and universities under the background of intelligent media integration. At the same time, it is necessary to pay attention to the individualization of student training to ensure that it does not deviate from the basic requirements of undergraduate teaching.

e) *Flexibility in the teaching process*

When colleges and universities face the conflict between postgraduate entrance examination and undergraduate teaching, it is easy to find it difficult to balance postgraduate entrance examination and professional course learning, and the teaching process of schools should be flexible. In terms of teaching curriculum arrangement, the first is to flexibly adjust the English and mathematics courses in the freshman year to the junior year, to synchronize with the review time for the postgraduate entrance examination; the second is to advance the study time of professional courses by 1–2 semesters[7].In terms of teaching methods, the first is to adopt different training methods for students with different specialties such as postgraduate entrance examinations, employment, and going abroad; the second is to use progressive teaching methods to guide students with pre-class previews and relevant cases and cutting-edge research. Independent self-study, and targeted design of classroom discussion and report writing to consolidate learning [8]; The third is to expand the platform for students to practice and exchange, and to launch on-campus activities based on course topics, not only to master and consolidate professional knowledge, but also to cultivate students' unity and cooperation, etc. ability.

f) *Significant learning outcomes*

Students can achieve the expected teaching goals through higher education, apply what they have learned, and find the next crossing that they like, which is the performance of significant learning outcomes. The school not only cultivates students' professional knowledge, innovative skills development and skill training, but also achieves high evaluation of students' self-perception satisfaction. Obtaining further study opportunities or a bright employment future is the greatest manifestation of significant learning outcomes.

g) *Learning to be mentally sound*

Learning psychology runs through the entire teaching process of students. Colleges and universities should pay attention to students' mental health and improve psychological education. First, it is necessary to repeat the educational role of classroom teaching in the cultivation of college students' mental health; second, the school should establish personal psychological files for students with departments as a unit, and pay attention to students' mental health problems in a timely manner[9]; Third, the school should make

use of associations, new media, school broadcasts, newspapers and campus networks and other channels to actively carry out mental health counseling activities; Fourth, schools should play the role of psychological counseling centers.

## 5 Epilogue

Colleges and universities should pay attention to the phenomenon of students taking postgraduate entrance examinations, but students who are keen on postgraduate entrance examinations cannot ignore the essential responsibilities of professional learning. Postgraduate entrance examination should become a mechanism to stimulate students to study hard, and should not hinder the cultivation of students' professional skills. Therefore, under the environment of intelligent media integration, schools should correctly guide students and develop intelligent teaching models in order to improve the quality of undergraduate classroom learning, give full play to the essence of undergraduate teaching, and make contributions to cultivating higher professional talents with a sense of social responsibility, innovative spirit and practical ability., to promote socialist modernization.

**Acknowledgment.** This thesis was completed under the careful guidance of Professor Wang Liang. Professor Wang Liang has had a profound influence on me with his rigorous attitude towards knowledge, high professionalism and rich social experience. The teacher not only taught me to explore and learn with humility, but also taught me the principles of life and doing things, expanding my thinking and broadening my horizons. Once again, I would like to express my heartfelt thanks to him.

I would also like to thank Beijing Jiaotong University for giving me the opportunity to submit this manuscript. It made me feel the internationalization of knowledge learning methods, and I am deeply honored to have a better platform to listen to the opinions and research results of experts in various fields.

Finally, I would like to thank my school, classmates, family and friends for their great support and encouragement, as well as the senior teachers who participated in reviewing and commenting on this paper during their busy schedules. I would like to express my heartfelt thanks to them. I will study hard with my family, study hard, apply what I have learned, and better contribute my meager strength to the development of the motherland.

## References

1. Ruitian, H.: The construction of smart teaching design model supported by artificial intelligence technology. Hainan Normal University (2021)
2. Ronghui, S.: The impact of college students' postgraduate entrance examination fever on undergraduate teaching and the countermeasures of colleges and universities. China Science and Education Innovation Guide (2008)
3. Yin, G., Hui, Z.: Research and application of smart teaching space model based on cloud-edge collaboration. Heilongjiang Higher Education Research (2020)
4. Weijun, W.: Research on the design and evaluation of remote smart teaching mode under the big data education ecology. J. Liaoning Radio Telev. Univ. (2019)

5. Karagöza, Ö., Zeki Sakab, A.: Development of teacher guidance materials based on 7e learning method in virtual laboratory environment. Procedia-Soc. Behav. Sci. (191), 810–827 (2015)
6. Xiangdong, X.: The development path and exploration breakthroughs of intelligent media. People's Forum (2021)
7. Jiandi, F., Zhenzhen, Z.: Research on the analysis methods and countermeasures of the impact of postgraduate entrance examinations on undergraduate teaching (2021)
8. Yan, G., Yunbo, Z., Wen, Z.: Research and practice on the teaching mode of applied undergraduate postgraduate entrance examination courses. Contemp. Educ. Pract. Teach. Res. (2018)
9. Xiaofeng, L., Jianjun, L., Gang, L., Xiaobing, T.: The influence and countermeasures of "postgraduate entrance examination fever" on undergraduate teaching. China Adult Educ. (2007)


# Research and Application Status of Text Generation Tasks Based on Generative Adversarial Network

Weiqi Wang, Dan Jiang(✉), and Shaozhong Cao

College of Information Engineering, Beijing Institute of Graphic Communication, Beijing, China
{jiangdan,caoshaozhong}@bigc.edu.cn

**Abstract.** In recent years, in the field of natural language processing, significant progress has been made in text generation. Text generation has gained widespread popularity in many fields such as abstract extraction, poetry creation, and response to social network comments. Given the excellent generative capabilities of Generative Adversarial Networks (GAN), it is often used as main model for text generation with remarkable results. This review aims to provide the core tasks of generative adversarial network text generation and the architecture used to deal with these tasks, and draw attention to the challenges in text generation with generative adversarial network. Firstly, we outline the mainstream text generation models, and then introduce datasets, advanced models and challenges of text generation tasks in detail. Finally, we discuss the prospects and challenges of the fusion of generative adversarial networks and text generation tasks in the future.

**Keywords:** Deep learning · text generation · natural language processing · generative adversarial networks

## 1 Introduction

As a research branch of Natural Language Processing (NLP), text generation has received more and more attention in recent years, and it has been widely used in machine translation [1–6], automatic question answering systems [7–11], text summarization [12–16] and text style transfer [17, 18]. All text generation tasks require the output texts to be fluent, coherent and faithful to the inputs. The development of text generation technology has also continuously promoted the development of human society. Many text generation technology achievements have been widely used in the industry [19, 20], such as Xiao Ai, Microsoft Xiaobing, etc., all of which are based on the application results of text generation technology. While the commercial value of text generation technology continues to increase, users' demands for its related products are also increasing. Today, models are not only required to generate coherent and coherent text, but also to generate diverse and high-quality text based on the information provided by the user.

Previous research on text generation methods mainly focused on template rule-based methods, which generated text by artificially creating language rules, such as lexical, part-of-speech, and grammar annotations. This method is only suitable for the case


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 109–122, 2023.
https://doi.org/10.1007/978-981-99-3618-2_11

where the amount of data is not large and the syntax and semantics are relatively simple, and its effect is relatively good compared to the method based on statistics. Nowadays, with the increasing amount of data and the continuous enrichment of various data rules, the rule-based method is not only time-consuming and labor-intensive, but also far less effective than the method used today. Therefore, the method based on template rules has now withdrawn from the historical stage.

At present, with the development of deep learning, more and more people apply deep learning to text generation. Including feedforward neural network and recurrent neural network, these models are based on Maximum Likelihood Estimation (MLE). This method completes the parameter update through backpropagation during the training process, so that the training results match the training data set , so as to obtain better results. Because of this, this method is prone to problems such as overwriting and exposure bias, that is, the model generates subsequent characters based on the real sample prefix during training, and predicts the next character based on the character prefix generated by the model during inference. This difference accumulates with the increase of sequence length, so in the long text generation task, it will lead to the disadvantages of poor text quality and poor diversity.

For problems such as overfitting and exposure bias, researchers have proposed the following solutions:

1. A neural language model [21] is proposed, through which the process of solving the optimal value of the model is converted into the process of solving the word vector, and the semantically smooth text is obtained by simulating the real distribution of the training text dataset.
2. The classic approach is to train a recurrent neural network (RNN) [22]. Since the language model adds Markov features, the nodes between its hidden layers are connected to each other, and each hidden layer receives the output of the input layer and the output of the hidden layer at the previous moment. Due to this property, the text generated by the RNN model has contextual connections.
3. The original RNN model cannot solve the problem of long-term dependence on sequences, so subsequent researchers continue to introduce a variant of RNN: long short-term memory (LSTM) network [23]. The key idea of LSTM is to add three extra gates to control the hidden vector passed to the next unit of the original RNN unit, which can better handle context-sensitive data such as text. The Gated Recurrent Unit (GRU) [4] also directly propagates backward by adding intermediate state information to solve the vanishing gradient problem. Make LSTM and GRU the standard model of RNN.
4. Employing Convolutional Neural Networks (CNN) [24], using temporal convolutional networks (CN) to apply deep learning to text understanding, from character-level input to abstract textual concepts. Compared with RNN, CNN has efficient parallel computing capability.
5. Apply Transformer [25] to text generation tasks. Based on each pre-trained model of Transformer, a self-attention mechanism and joint position encoding are introduced to

resolve the dependencies between words in the generative model. Using a pretrained model can make the natural language processing model independent of the recurrent neural network structure, which is beneficial to the parallel computing of large-scale data. For example, Bert [26], XLNet [27], GPT-2 [28], etc. all use large amounts of data for pre-training and then fine-tune for specific tasks and specific datasets. These pre-trainings can give the model a better initial state, leading to better performance.

6. Using Generative Adversarial Network (GAN) [29], GAN can achieve the purpose of synthesizing data by simulating the data distribution of the original data set and making judgments. This is also the focus of this study.

To sum up, text generation is an important research direction in the field of natural language processing. Researchers hope that computers can understand human language and automatically generate texts that are logical and fluent, so as to write articles like humans. These applications will continue to benefit human society and further liberate productivity. This paper takes the text generation task based on generative adversarial network as the core, and summarizes it. First, the background knowledge of generative adversarial networks and the modeling of text generation tasks are introduced. Then, the application of generative adversarial network in text generation is reviewed and the advantages and disadvantages of each algorithm are analyzed. Finally, the full text is summarized, and the research trends and application prospects of the combination of generative adversarial network technology and text generation tasks are analyzed.

## 2 Generative Adversarial Network Background Knowledge

Generative adversarial network is not a complete model, but a framework whose basic structure consists of a generator G responsible for generating data and a discriminator D for distinguishing generated data from real data. The specific network structure of the generator and the discriminator needs to be formulated by ourselves. The relevant vector of the image or text generated by G is input into D to connect the generator and the discriminator. During the training process, the generator and the discriminator play against each other, and the generator hopes to deceive the discriminator by continuously generating fake data similar to the real samples. The discriminator needs to continuously distinguish the generated data from the real data, and the output data is the probability value that the sample is the real sample. In this process, the generator and the discriminator are continuously improved. In the end, not only the discriminator is proficient in data recognition, but the data generation capability of the generator is also greatly improved. Therefore, usually the end of the discriminator network is a binary classifier, and the feedback from the discriminator in turn motivates the generator. The overall model is shown in Fig. 1 below.

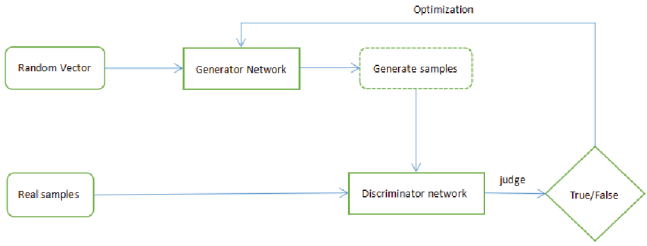


**Fig. 1.** The general structure of the generative adversarial network framework

Suppose the input of the generator is a random variable z, the prior distribution of the variable is $P_z$, the generator is $G(z,\theta_g)$, where $\theta_g$ is the parameter of the generator, and the discriminator is $D(x,\theta_d)$, where $\theta_d$ is the parameter of the discriminator.Assuming that the final distribution of the generator is $P_g$, and the distribution of the real data is $P_{data}$, the goal of the generator is to make $P_g$ fit $P_{data}$ as much as possible. The overall training objective of GAN is as follows:

错误!嵌入对象无效。 (1)

Among them, V(G, D) is the minimization game problem between the discriminator and the generator, as shown in Eq. (2). The first term is the discriminator's discriminative expectation of real samples, and the second term is the discriminator's discriminant expectation of generated samples.

错误!嵌入对象无效。 (2)

The goal of the discriminator is to make the output of the real sample as close to 1 as possible, and the output of the generated sample as close to 0 as possible, while the goal of the generator is the opposite. When the distribution of the data is continuous distribution, formula (2) can be rewritten as:

错误!嵌入对象无效。 (3)

In the training process of GAN, the optimization of G(.) and D(.) is alternately performed, and the discriminator is trained by fixing the G(.) parameter. When the discriminator can correctly distinguish the generated samples from the real samples, Eq. (3) obtains the extreme value, at this time:

错误!嵌入对象无效。 (4)

This is because for real numbers a and b that are not all 0, a function of the form $a\log(y) + b\log(1-y)$ will always have a maximum value on $a/(a+b)$. Substitute Eq. (4) into Eq. (3) to get:

错误!嵌入对象无效。 (5)

错误!嵌入对象无效。 (6)

错误!嵌入对象无效。 (7)

If $P_g(x) = P_{data}(x)$, then $D_G^*(x)$ is 0.5. At this time, formula (7) is transformed into:

错误!嵌入对象无效。 (8)

Compared with other mainstream deep generative models, GAN has the following advantages:

(1) The GAN network has a small amount of computation, simple construction, efficient training, and requires fewer iterations, and avoids the tedious Markov learning mechanism [30]. It can directly perform efficient sampling and inference, so it has a wide range of application scenarios.
(2) It can better extract the implicit association of the data distribution and generate better quality samples.
(3) GAN is flexible, and the loss function can be designed according to different tasks and integrated into the GAN model for learning and optimization.
(4) GAN can also establish a connection with reinforcement learning and apply it to reinforcement learning[31].

However, the original structure of generative adversarial network has problems when dealing with discrete data generation such as text. Since the "sampling" step exists in the middle layer of the network, the discriminator network cannot back-propagate the gradients to the generator network. Specifically, the discriminator network can only receive the vector after sampling, and cannot obtain all the information of the generator network. As a result, the feedback signal is lost, and the generator network cannot find the direction of parameter optimization.

In addition to modeling reasons, primitive structures have mathematical problems with discrete data. As shown in Eq. (8), the measure of the gap between the generated samples and the real samples is the JS divergence. JS divergence cannot measure the distance between two non-overlapping distributions. For two distributions P and Q that do not have any intersection, the JS divergence between them is a constant value:

错误!嵌入对象无效。 (9)

Since the relevant parameters of the generator network are all randomly initialized, they cannot be guaranteed to overlap. Therefore, no matter which direction the generator network performs gradient-based parameter fine-tuning in the parameter space, the JS divergence between it and the real sample distribution is always log2, and the generator network cannot adjust the parameters in this case.

Based on these problems, subsequent researchers have improved the GAN model for discrete data.

## 3 Generative Adversarial Network-Based Text Generation Method

Generative Adversarial Networks have achieved great success in the task of generating images. However, there are two obstacles to applying generative adversarial networks to text generation tasks. The first is that the Generative Adversarial Network was originally

designed to generate real-valued continuous data, and it is difficult to directly generate discrete sequences. The reason is that the generator starts with random sampling first, and the gradient of the loss of the discriminative model D guides the update of the generative model G with slight changes. If the generated data is based on discrete sequences, then this "slight change" of the discriminative network does not make much sense for the update of the generative network. Second, the generative adversarial network can only give the score or loss of the entire sequence when generating the sequence. For the partially generated sequence, it cannot balance the current good or bad with the future score of the entire sequence.

When GAN is applied to the field of text generation, the text data is not differentiable during model optimization due to the discreteness of language, which is prone to gradient explosion and gradient disappearance. In 2016, Arjovsky et al. [32] proposed Wasserstein GAN. This variant of generative adversarial network uses Wasserstein distance to calculate the distance of two distributions, replacing the previous JS divergence. However, its essential problem as a generative adversarial network has not been solved, that is, the problem of non-steerability of the sampling process has not been solved. Therefore, Kusner et al. [33] proposed a new model, GSGAN, to solve the above problems through the Gumbel-softmax distribution. This model can use the GS distribution to be infinitely close to the original distribution, which ensures that both the generator and the discriminator can complete gradient-based parameter updates, enabling text sequence generation. However, since GSGAN still uses JS divergence for the discrimination between generated text and real text, the problems of difficult model training and mode collapse remain unsolved. In the same year, Zhang et al. [34] used the idea of smooth approximation to solve the problem of non-derivable discrete gradients. Therefore, the discrete features of text can be solved, so that GAN can be applied to text generation tasks. In 2018, Fedus et al. [35] pointed out that the traditional evaluation of language perplexity cannot directly measure the quality of language models. Therefore, they use the Actor-Critic algorithm of reinforcement learning to optimize the GAN generator. In order to solve the problem of model collapse and instability during GAN training, they use the context-based method to fill in the missing words. In 2019, Xu et al. [36] proposed a DP-GAN model for the problems of poor text diversity, easy repetition, and low information content in existing text generation models. The model adopts a reinforcement learning policy gradient optimization algorithm, uses cross-entropy as a reward, and guides the generator to produce novel and smooth text through high reward value. Nie et al. [37] proposed the RelGAN model, and they pointed out that the model collapse of the adversarial network is caused by the lack of the generator's ability to express diversity or the lack of feedback information from the discriminator. Therefore, they use multiple word embedding representations in the discriminator to provide richer feedback information to guide the generator optimization, and use relational memory networks instead of traditional LSTM network-based generators to enhance the generator's ability to express diversity.

Yu et al. [38] proposed a sequence generative adversarial network model (SeqGAN) by combining reinforcement learning and generative adversarial networks. The core concept of the algorithm model of the sequence generative adversarial network is to regard a complete generative adversarial network as a reinforcement learning system,

and the update of the parameters in the model is realized by means of the policy gradient algorithm. At the same time, it also draws on the concept of Monte Carlo Tree Search (MCTS) to achieve real-time evaluation of incomplete sequences at any point in time. The algorithm of the sequence generative adversarial network itself still maintains the architecture of the generator and the discriminator in the generative adversarial network algorithm model. The algorithm framework is shown in Fig. 2.

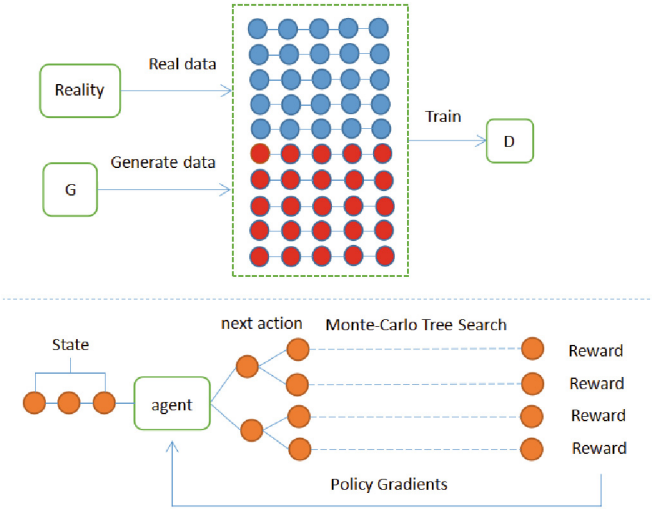


**Fig. 2.** Sequence Generative Adversarial Network Structure Diagram

This model treats text generation models using recurrent neural networks as agents in reinforcement learning tasks. The current state is the word that has been generated. When the next action is determined, the current state moves to the next state. The discriminator model takes the real text data and the generated text data as input, and the output data is the probability of real data.

In order to solve the problem that the gradient mentioned above cannot be returned, SeqGAN uses the reinforcement learning policy gradient method [39, 40] to update the generator. In order to maximize the cumulative reward of the sequence, its objective function is defined as:

错误!嵌入对象无效。 (10)

where $R_T$ is the cumulative reward of a complete sequence. $Q_{D_\phi}^{G_\theta}(s_t, a_{t+1})$ is the state-action value function of the sequence, indicating how good or bad the selected action $a_{t+1}$ is in the current state $s_t$. The SeqGAN model uses the output probability of the discriminator D as the reward function in reinforcement learning as.

错误!嵌入对象无效。 (11)

where $Y_{1:T}$ is the complete sequence of length T. It can be seen that the discriminator can only evaluate the complete sequence. To solve this problem, the Monte Carlo method can

be used to complete the partial sequence into the complete sequence, and approximate the action value function of the intermediate state. The discriminator D is updated in an iterative training manner, providing dynamic guidance to the generator G, and its objective function can be expressed as:

错误!嵌入对象无效。 (12)

Among them, $Y \sim P_{\text{data}}$ is the real data, $Y \sim G_\theta$ is the generated data.

In the SeqGAN algorithm, due to the use of the policy gradient algorithm, each time the generator generates a new parameter, a feedback can be obtained to update the parameters of the generator. Compared with the original GAN, which needs to use the back-propagation of the discriminant network to achieve parameter update, SeqGAN gets rid of the dependence of the discriminator. The Monte Carlo search also allows the algorithm to immediately evaluate the quality of the current word without waiting for all sequences to be generated. Therefore, SeqGAN solves the problems encountered when applying GANs to text generation mentioned above.

However, the SeqGAN model suffers from two major problems of gradient vanishing and mode collapse during training. Subsequent researchers made further improvements to the SeqGAN model on how to address these two problems.

Goodfellow et al. [29] replaced the discriminator with a ranker, and the model structure consisted of a sequence generator G and a ranker R, and designed RankGAN. Among them, the ranker R can rank the real data and the generated data relatively when a reference is given, and use the relative ranking result as a reward to guide the generator. The ranker R expresses the relevance score of the input sequence given the parameters by calculating the cosine similarity, and then uses the softmax formula to calculate the ranking score. RankGAN actually replaces the binary classifier with a ranking score based on multiple sentences, which can effectively solve the problem of gradient disappearance. Because the reward value provided by the binary classifier is sparse, the resulting generator cannot be updated effectively and cannot generate diverse and logical texts. However, since it requires additional sampling of the reference set, its computational cost is higher than other models.

In addition to this approach, Che et al.[41] propose that the reward function can be re-scored, and they design a maximum likelihood augmented discrete generative adversarial network (MaliGAN). MaliGAN optimizes the generator to utilize importance sampling, combined with the output of the discriminator, and recalculates the obtained score as a reward using Eq. (13). MaliGAN brings the training process closer to maximum likelihood training of autoregressive models, resulting in more stable gradients.

错误!嵌入对象无效。 (13)

It is also mentioned in the text that MaliGAN uses Monte Carlo number search, and gradually transitions from MLE to MaliGAN to reduce variance.

Based on the SeqGAN model, Guo et al. [42] combined the hierarchical reinforcement learning framework [43] to construct LeakGAN, which made GAN obtain excellent results in long texts. The discriminator in the LeakGAN model uses the CNN classifier [44] to score the generated sentences. The generator adopts the LSTM model [45] to automatically generate some sentences. It can be seen from the above that when the

generator generates a complete sentence, it can be sent to the discriminator model for feedback. When the text is long, this mechanism cannot effectively guide the generator parameter update. Therefore, on the feedback mechanism, the LeakGAN model leaks its own extracted features through the discriminator to further guide the generator, and uses the policy gradient algorithm of reinforcement learning to compare the cosine distance between the real text and the generated text to evaluate the generator. At the same time, use the Monte Carlo tree search [46] to complete the currently generated part, and evaluate the quality of the current part, without waiting until the entire sentence is generated to evaluate the quality. At the same time, the generator is modeled as a hierarchical reinforcement learning problem [47], including a high-order Manager module and a low-order Worker module, both of which are constructed using long-short-term memory networks. At each time step, the Manager module takes as input the high-dimensional feature representation received from the discriminator, and outputs a guiding target vector. The Worker module encodes the currently generated words through a long-short-term memory network, and combines its output with the target vector by matrix product to ensure that a suitable new word can be generated according to the Manager's guidance and the current state.

The biggest improvement of the LeakGAN model is that it trains the generator by leaking internal features for the first time, and combines hierarchical reinforcement learning to solve the problems existing in the previous generation model in generating long texts.

For the mode crash problem. That is, in the training process, the generator deceives the discriminator to obtain high scores by fitting only a specific part of the target distribution, and often can only generate simple and short repetitive sentences, which greatly reduces the diversity of generated texts. Xu et al. [36] argue that existing classifier-based discriminators suffer from saturation problems. That is, it can only distinguish between true and false sentences, but cannot judge the novelty of new sentences, which leads to the tendency of text generation models to generate some repetitive and meaningless texts. Therefore, the DP-GAN (diversity-promoting generative adversarial net-work, DP-GAN) model is proposed. The key idea is to build a discriminator responsible for providing a reward to the generator based on the novelty of the generated text. Treat texts that are frequently generated by the generator as low-novelty texts, and texts that are infrequent in the generated data as high-novelty texts. The discriminator is trained with real-world text as positive examples and generated text as negative examples. This training mechanism encourages the discriminator to give higher rewards to text that looks like real-world data. The reward is fed back to the generator, which generates diverse and fluent text through policy gradients. To generate multiple sentences, it builds a standard hierarchical LSTM decoder [48]. The two layers of LSTM are hierarchical structure. The bottom layer decodes the sentence representation, and the top layer decodes each word based on the output of the bottom layer. The attention mechanism is used for word decoding [5, 49]. In addition, DP-GAN uses two reward methods, local word-level reward and global sentence-level reward. For a sentence $y_t$ containing K words, the sentence-level reward is the average reward per word:

错误!嵌入对象无效。 (14)

Considering that the reward $y_t$ should be different for different words in a sentence, use the reward at the word level as follows:

错误!嵌入对象无效。 (15)

Experiments show that the DP-GAN model significantly outperforms the baseline methods in both automatic and human evaluation. This shows that DP-GAN is able to generate more diverse and informative texts. Second, compared to traditional classifier-based discriminators, this discriminator can better discriminate between new and repeated texts with saturation problems. Finally, with the increased diversity, the data distribution generated by DP-GAN is closer to the real-world data distribution than MLE.

Zhou et al. [50] propose a new self-adversarial learning (SAL) model for the problem of reward sparsity and mode collapse. SAL uses a comparison discriminator, which is a pairwise classifier that compares the quality of text between a pair of samples. Assuming that its input is two text sequences A and B, the output labels contain three categories, corresponding to the quality of sequence A is higher ($>$), lower ($<$) and indistinguishable ($\approx$) than B. During training, SAL rewards the generator if it is found that the sentences currently generated by the generator are better than previously generated samples. In the early training stage, when the quality of the generated samples is much lower than the real data, this self-improving reward mechanism makes it easier for the generator to receive non-sparse rewards with informative learning signals, effectively alleviating the reward sparsity problem. And in later training phases, SAL prevents the samples from continuing to have high returns, as self-improvement on popular patterns will become increasingly difficult, thus helping the generator to avoid collapsing towards the limited patterns of real data. This self-improving reward mechanism makes it easier for the model to obtain rewards and avoid collapse on a limited number of real samples, which not only helps alleviate the reward sparsity problem, but also reduces the risk of model collapse. Experiments show that the text generated by the SAL model has a good performance in terms of relevance, diversity, and fluency.

## 4 Experiment and Analysis

In order to unify the data set and make the final generated results more comparable, this paper uses the EMNLP2017 WMT News (48M) data set, in which the data has been preprocessed by word segmentation and divided according to the standards of the literature [51]. The EMNLP2017 WMT News dataset contains a training set of 270,000 sentences and a test set of 10,000 sentences. The number of samples is set to 10,000, the maximum length of a single sentence is 51, and the vocabulary size is 5,256.

The evaluation metric used in this paper is BLEU. BLEU was proposed by IBM scientist Kishore [52] as an indicator for evaluating the difference between generated sentences and real sentences, and is often used in machine translation tasks. The BLEU score is calculated by calculating the N-grams model of the generated sentence and the real sentence, and then counting the number of matches, which is an evaluation method independent of word order. The BLEU algorithm is related to the matching rules of n-grams, where n represents n words that are close to each other. Common evaluation

**Table 1.** Performance comparison of different models in EMNLP2017 WMT news generation task. The metrics from left to right represent generation quality and generation diversity, respectively. For all BLEU metrics, higher is better.

| Method | BLEU-2 | BLEU-3 | BLEU-4 | BLEU-5 | P-value |
|---|---|---|---|---|---|
| MLE | 0.768 | 0.473 | 0.240 | 0.126 | $< 10^{-6}$ |
| SeqGAN | 0.777 | 0.491 | 0.261 | 0.138 | $< 10^{-6}$ |
| RankGAN | 0.736 | 0.441 | 0.204 | 0.097 | $< 10^{-6}$ |
| MaliGAN | 0.764 | 0.468 | 0.231 | 0.113 | $< 10^{-6}$ |
| LeakGAN | 0.826 | 0.645 | 0.437 | 0.272 | $< 10^{-6}$ |
| SAL | 0.876 | 0.654 | 0.404 | 0.233 | $< 10^{-6}$ |

indicators include BLEU-1, BLEU-2, BLEU-3 and BLEU-4. BLEU-1 can measure the accuracy of words. The higher-order BLEU can measure the fluency of sentences.

A comparative experiment is performed on the EMNLP2017 WMT NEWS dataset, and the experimental results are shown in Table 1. The experimental results show that the evaluation indicators of the SAL model are most significantly improved on BLEU-2 and BLEU-3. As the sequence becomes longer and the data complexity becomes larger, the performance of the SAL model decreases slightly compared to the LeakGAN model, but it is still better than other models.

## 5 Conclusion and future work

This paper conducts related research on text generation in natural language processing. The generative adversarial network does not rely on any prior assumptions, and learns the statistical laws of real samples through the generative model and the discriminant model, and has a strong generative ability. If it is applied to discrete data such as text, it will not be possible to update parameters directly through backpropagation, and the original GAN framework also brings problems such as gradient disappearance, difficulty in training, and mode collapse. Therefore, this paper also summarizes the previous solutions to these problems, and briefly summarizes the background, basic concepts, model structure, advantages and disadvantages. Although text generation based on generative adversarial networks has achieved certain results, there are still some problems and challenges in research, and new text generation models are still being proposed. How to apply better generative adversarial networks to text generation tasks is also a research direction worthy of researchers' efforts. In addition, there is room for improvement in research containing keywords and emotion generation, so that computers can generate more accurate and emotional texts. It will surely help mankind usher in a better intelligent world.

## References


1. Kalchbrenner, N., Blunsom, P.: Recurrent continuous translation models. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp.1700–1709 (2013)
2. Gehring, J., Auli, M., Grangier, D., et al.: A convolutional encoder model for neural machine translation (2016)
3. Zhen, Y., Wei, C., Feng, W., et al.: Improving neural machine translation with conditional sequence generative adversarial nets (2018)
4. Cho, K., Merrienboer, B.V., Gulcehre, C., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. Computer Science (2014)
5. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. Computer Science (2014)
6. Cho, K., Merrienboer, B.V., Bahdanau, D., et al.: On the properties of neural machine translation: encoder-decoder approaches. Computer Science (2014)
7. Li, J., Monroe, W., Shi, T., et al.: Adversarial learning for neural dialogue generation. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (2017)
8. Weizenbaum, J.: ELIZA: a computer program for the study of natural language communication between men and machines (1974)
9. Colby, K.M.: Artificial paranoia: a computer simulation of paranoid processes. Elsevier (2013)
10. Shieber, S.M.: Lessons from a restricted Turing test. Commun. ACM **37**, 70–78 (1994)
11. Wallace, R.S.: The Anatomy of A.L.I.C.E. In: Epstein, R., Roberts, G., Beber, G. (eds.) Parsing the Turing Test. Springer, Dordrecht (2009). https://doi.org/10.1007/978-1-4020-6710-5_13
12. Rush, A.M., Chopra, S., Weston, J. : A neural attention model for abstractive sentence summarization. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 379–389 (2015)
13. Zeng, W., Luo, W., Fidler, S., et al.: Efficient summarization with read-again and copy mechanism. arXiv preprint arXiv:1611.03382 (2016)
14. Nallapati, R., et al.: Abstractive text summarization using sequence-to-sequence RNNs and beyond. In: Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (2016)
15. Rush, A.M., Chopra, S., Weston, J.: A Neural attention model for abstractive sentence summarization. Computer Science (2015)
16. Alifimoff, A.: Abstractive sentence summarization with attentive deep recurrent neural networks (2016)
17. Prabhumoye, S., Tsvetkov, Y., Salakhutdinov, R., et al.: Style transfer through back-translation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (2018)
18. Zhu, J.-Y., Park, T., Isola, P., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
19. Zadeh, L.A.: From search engines to question-answering systems - the need for new tools. In: The 12th IEEE International Conference on Fuzzy Systems, 2003, pp. 1107–1109. FUZZ 2003. IEEE (2003)
20. Zhao, T., Ran, Z., Eskenazi, M.: Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (2017)
21. Bengio, Y., Ducharme, R., Vincent, P., et al.: A neural probabilistic language model. J. Mach. Learn. Res. **3**, 1137–1155 (2003)

22. Karpathy, A.: The unreasonable effectiveness of recurrent neural networks. Andrej Karpathy Blog **21**, 23 (2015)
23. Schmidhuber, H.S., et al.: Long short-term memory. In: Neural Computation, vol. 9, no. 8, pp. 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
24. Zhang, X., Lecun, Y.: Text understanding from scratch. arXiv:1502.01710 (2015)
25. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. arXiv:1706.03762 (2017)
26. Devlin, J., Chang, M.W., Lee, K., et al.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
27. Yang, Z., Dai, Z., Yang, Y., et al.: XLNet: generalized autoregressive pretraining for language understanding. In: Proceedings of the International Conference on Neural Information Processing Systems, pp. 5753–5763 (2019)
28. Radford, A., Wu, J., Child, R., et al.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8), 9 (2019)
29. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In: Proceedings of the International Conference on Neural Information Processing Systems, pp. 2672–2680 (2014)
30. Neal, R.M.: Markov chain sampling methods for Dirichlet process mixture models. J. Comput. Graph. Stat. **9**(2), 249–265 (2000)
31. Guo, H.: Generating text with deep reinforcement learning. Comput. Sci. **40**(4), 1–5 (2015)
32. Martin, A.S.C., Bottou, L.: Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, pp. 214–223 (2017)
33. Kusner, M.J., Hernández-Lobato, J.M.: GANs for sequences of discrete elements with the gumbel-softmax distribution. arXiv preprint arXiv:1611.04051 (2016)
34. Zhang, Y., Gan, Z., Carin, L.: Generating text via adversarial training. In: NIPS workshop on Adversarial Training (2016)
35. Fedus, W., Goodfellow, I., Dai, A.M.: MaskGAN: better Text Generation via Filling in the_____ (2018)
36. Xu, J., Ren, X., Lin, J., et al.: DP-GAN: diversity-promoting generative adversarial network for generating informative and diversified text (2018)
37. Nie, W., Narodytska, N., Patel, A.: RelGAN: relational generative adversarial networks for text generation. In: ICLR 2019 Conference Paper 1462 (2019)
38. Yu, L., Zhang, W., Wang, J., et al. SeqGAN: sequence generative adversarial nets with policy gradient (2016)
39. Sehnke, F., Osendorfer, C., Rückstieß, T., et al.: Parameter-exploring policy gradients. Neural Netw. **23**(4), 551–559 (2010)
40. Williams, R.J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach. Learn. **8**(3–4), 229–256 (1992)
41. Che, T., Li, Y., Zhang, R., et al.: Maximum-likelihood augmented discrete generative adversarial networks. arXiv preprint arXiv:1702.07983 (2017)
42. Guo, J., Lu, S., Han, C., et al.: Long text generation via adversarial training with leaked information (2017)
43. Vezhnevets, A.S., Osindero, S., Schaul, T., et al.: FeUdal networks for hierarchical reinforcement learning (2017)
44. Kim, Y.: Convolutional neural networks for sentence classification. Eprint Arxiv (2014)
45. Mou, L., Yan, R., Li, G., et al.: Backward and forward language modeling for constrained sentence generation. Comput. Sci. **4**(6), 473–482 (2015)
46. Sutton, R.S., Mcallester, D., Singh, S., et al.: Policy gradient methods for reinforcement learning with function approximation. MIT Press (1999)

47. Dietterich, T.G.: An overview of MAXQ hierarchical reinforcement learning. In: Choueiry, B.Y., Walsh, T. (eds.) Abstraction, Reformulation, and Approximation. SARA 2000. Lecture Notes in Computer Science, vol. 1864. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44914-0_2
48. Li, J., Luong, M.T., Dan, J.: A hierarchical neural autoencoder for paragraphs and documents. Computer Science (2015)
49. Luong, T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. EMNLP **2015**, 1412–1421 (2015)
50. Zhou, W., Ge, T., Xu, K., et al.: Self-adversarial learning with comparative discrimination for text generation (2020)
51. Zhu, Y., Lu, S., Zheng, L., et al.: Texygen: A benchmarking platform for text generation models. In: Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1097–1100 (2018)
52. Papineni, K., Roukos, S., Ward, T., et al.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311–318 (2002)

# Research on Copyright Protection of We Media Based on Blockchain

Ruidan Yang[1] and Liang Wang[2(✉)]

[1] School of Journalism and Publication, Beijing Institute of Graphic Communication, Beijing, China

[2] School of Economics and Management, Beijing Institute of Graphic Communication, Beijing, China

wangliang@bigc.edu.cn

**Abstract.** We media has gained rapid development on the fast train of digital era, and everyone has a microphone. However, due to the low threshold, large flow and fast realization, we media is also very prone to problems such as infringement, etc. Blockchain, as a kind of distributed ledger, can provide technical support and application in the problem of copyright protection of we media with its characteristics of decentralization, non-tamperability, trustworthiness and privacy protection, etc. This paper discusses and researches the problems existing in the process of we media copyright protection, and provides reasonable ideas and measures for we media copyright protection based on blockchain technology from three perspectives of confirming, using and defending rights, in order to better use technical means to combat infringement in we media field, safeguard the legitimate rights and interests of we media, and build a positive and clear we media communication environment.

**Keywords:** blockchain · we media · copyright protection

## 1 Introduction

The Internet era has given everyone the opportunity to have a voice, and we media has seen a booming development. While creating a large amount of rich content, the complex and diversified we media has also brought new problems and challenges to copyright protection. Blockchain is a decentralized management technology based on data encryption, consensus mechanism and timestamp, which can establish peer-to-peer trusted value transactions between uncontacted nodes [1], which also provides technical solutions and ideas for we media copyright protection.

This work is partially supported by National Social Science Fund of China under contract with No. 21BXW037.


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 123–131, 2023.
https://doi.org/10.1007/978-981-99-3618-2_12

## 2 The Current Situation of the Development of the We Media Industry and the Difficulties of Copyright

### 2.1 Development Status

In 1991, American scholar Russell Newman predicted the trend of "4Cs", i.e., increasing participation and control of members of society in the communication process (control), increasing media convergence (convergence), more convenient use of communication media (convenience) and The trend is also becoming a reality in the evolving online environment, which has led to the rapid growth of the we meida industry.

"We media" is a universal, private, autonomous and popular communicator. It transmits normative and non-normative information to an unspecified majority or a specific individual in an electronic and modern way. General term for new media that provide and share their own facts and news [2]. The lower entry barrier, the traffic of large cash, fast profit characteristics, etc., prompted the we media to become an increasing volume of industry. Data from a Chinese company called Qi Cha Cha shows that the number of we media related enterprises in China has developed from 313 in 2014, more than 1,000 in 2017, and 2,747 in 2019 to 12,089 in 2021 [3] (see Fig. 1), published by China Research Institute of Industrial《Report on market survey and development prospect analysis of China we-media industry (2020–2025)》 It is also pointed out that the number of practitioners in China's we media industry has ushered in significant growth since 2014, and has exceeded 2 million in 2015, and by 2021, the number of people engaged in we media full-time in China has reached 3.7 million, and from a global perspective, the we media industry is also basically in the wind, and it can be seen that whether at present or in the predictable future,we can see that the we meida industry will continue to prosper, with huge development opportunities and space.

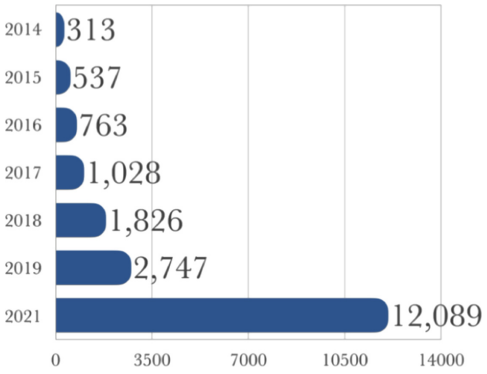


**Fig. 1.** We media related business registration in China, 2014–2021

### 2.2 Copyright Difficulties

a. Confirmation of copyrights: "Text poachers" abound, copyright attribution is difficult to determine

The concept of "text poachers" was first introduced by communication scholar Henry Jenkins in his 1988 paper "Star Trek Returns, Reread, Rewrite: Fan Writing as Text Poachers," and further elaborated in "Text Poachers - Television Fans and Participatory Culture. In Jenkins' view, text poachers, are "fans" of reinterpreting and reproducing texts [4], In the current "anonymous" online environment, the low cost of infringement and the huge amount of we media works have given rise to a large number of "text poachers" who "recreate" and "shuffle" we media works in pursuit of profits. We media works such as articles, short videos, pictures, etc. are subject to different degrees of infringement.

However, as a kind of creative content, the creation process of we media works is also extremely complicated, involving multiple rights holders, and the infringed we media works that have been disseminated several times in the network have changed owners many times and become "completely different" in the process of continuous dissemination. In addition, since copyright has the property of automatic protection, many we media do not spend a lot of time and energy to register their copyrights, and there are "white space" and "absence" in the confirmation of rights of we media.

b. Use of copyrights: redundant authorization process, increasing the difficulty of copyright use

In practice, it is very difficult to find the real owner of the copyrighted works in the vast network space, the vast amount of we media works, and the anonymous online works, and after overcoming the difficulty of finding the information of the owner, it is also necessary to contact the owner and negotiate the authorization, which takes a lot of energy and a long period of time. It is enough to jam a large number of we media, not to mention that they have to apply to the copyright management department to use such works [5]. Meanwhile, "the traditional offline "one-to-one" negotiation and transaction model is obviously unable to adapt to the demand of copyright transactions in the online environment [6]".

c. Defending copyrights: the cost of infringement and the cost of defending rights are imbalanced, and the we media is deep in the quagmire of defending rights

The original we media works are not technically protected by the original we media authors, which gives the lawless elements a chance to take advantage of the situation. "Some infringers will even "transfer" the original we media works directly to other platforms through watermark masking and other means to gain economic benefits. In contrast, in order to defend the copyright of their own works, we media need to go through complicated steps such as analysis of works and proof of infringement, each step of which costs we media a lot of time, money and human resources, and the cost of infringement and the cost of defending rights are in a dichotomous state. For "Big V" we media, if they have enough power to fight against infringers and actively defend their works, but for small we media, the high cost of defending rights often makes them give up the pursuit of infringers, which also causes infringing subjects who almost do not have to bear the consequences of infringement to be more reckless and continue to infringe, forming a vicious circle.

## 3 Blockchain in the Application of We Media Copyright Protection

Blockchain technology is a space-biased technology that utilizes a blockchain data structure to verify and store data, a distributed node-based consensus algorithm to generate and update data, a new distributed infrastructure and computing paradigm to secure data transmission and access with the help of cryptography, and a composed smart contract to program and manipulate data with the help of automated scripting code [7]. Countries increase the development and application of blockchain, and the role of blockchain in deepening information technology applications is becoming more and more important, becoming one of the hottest technologies today, "The distributed architecture of blockchain has the characteristics of transparency and openness, consistent state, high proof strength, timestamp, and strong dependence on cryptography [8]." It is very suitable for the current practical needs of confirming, using and maintaining the rights of we media works, and various digital copyright platforms based on blockchain technology in the field of intellectual property rights are emerging, such as Mediachain and Ethereum in the United States, ENT in South Korea, Ujo Music in the United Kingdom, Voise in Spain, Colu in Israel, [9], and originally, PaperGuy in China Technology, Yishu Platform, Huiju.com, Ant Chain, Copyright Home, Xiao Rhino Copyright Chain, etc. Specifically, blockchain's "decentralization", "smart contract" and "non-tamperability " can help the application and practice of copyright protection for we media.

### 3.1 Confirmation of Copyrights: "Decentralization" to Achieve Inexpensive and Efficient Copyright Registration for We Media

Blockchain is a specific data structure in a chain by arranging data blocks in chronological order, which is essentially a distributed ledger with decentralized characteristics and free from the reliance on traditional third-party trusted institutions. It can realize decentralized and trustworthy storage [10].

The existence of "text poachers" makes it difficult for we media to confirm copyright ownership, but the costly and long-period offline copyright registration has also become an intrinsic factor preventing many we media from not doing copyright confirmation, and the "decentralization" feature of blockchain can provide powerful technical support to solve this problem. The blockchain copyright registration system can complete the registration of copyright instantly after the applicant uploads the we media work, which is convenient and instantaneous. In addition, the operating cost of the blockchain platform system is very low compared with that of traditional centralized data centers, only 4% of the operating cost of traditional centralized data centers and 1% of the infrastructure cost, which further increases the possibility of low-cost and efficient rights verification. As a we media platform for photography enthusiasts, "Tuchong" has already carried out instant free online blockchain deposit confirmation of photographic works based on ant blockchain technology, and the applicant can complete copyright registration and obtain a blockchain deposit certificate of digital works within a moment after uploading photographic works. The new paradigm of high-efficiency and low-cost copyright verification is formed.

### 3.2 Use of Copyrights: "Smart Contract", Perfect the We Media Authorization Transaction Model

The concept of "smart contracts" was first introduced in 1995 by Nick Szabo, a cross-disciplinary legal scholar, as "a set of digitally represented promises, including agreements on which the contracting parties can execute them" [11]. Smart contracts are also a prominent technological advantage of blockchain. It is a computer protocol designed to propagate, validate, or enforce contracts in an informationized manner, allowing trusted transactions to take place without a third party.

Usually, the authorization and licensing model of most copyright objects is "author-registration platform or intermediary-consumer", which also requires a series of tedious authorization process such as finding the right subject and authorization negotiation, and the longer the time spent, the more likely it is to hinder the copyright transaction. In the blockchain system, you can use "smart contracts" to set the trading conditions or authorize others to use them independently. Through the automated execution of smart contracts, we media can place the content of their works on a unified "copyright trading platform", and mark the price so that those who need the copyright can buy it directly on the platform, forming a "peer-to-peer" automated trading model. In addition, according to the transaction information block and its unique timestamp and hash value, it can form a unique digital identification of all transaction information, and realize the tracking and monitoring of copyrighted works after authorization through cross-chain technology and API interface, etc. Once the situation is found to be contrary to the use in the smart contract, the data can be fed back to the copyright owner at [12], which further helps the we media to quickly and sensitively react to the possible infringement risks. This will help we media react quickly to possible infringement risks, solve the problem of uncontrollable authorization in the traditional authorization management model, and further improve the authorization transaction model of we media.

### 3.3 Defending Copyrights: "Untamperable", Optimizing the Remedy Path for the Maintenance of Copyrights of We Media

At present, the difficulties suffered by we media to defend their works mainly include the difficulty of proving ownership and the difficulty of depositing evidence, because of the unstable and easily tampered characteristics of traditional electronic data, rights holders often need to use third-party notary institutions, using professional technical means to achieve the forensic evidence of infringement, such a high cost of judicial remedies, usually become a factor that prevents we media to actively defend their rights, while the block blockchain has the technical characteristics of "real-time record" and "difficult to tamper", which can facilitate the proof of ownership and evidence of infringement for copyright owners of we media [6].

From the block data structure in Fig. 2, it can be seen that the blocks of transaction information in the blockchain are linked in order from far to near, and each block points to the previous block, and in general, a block has only one corresponding parent block, which contains the corresponding field of "parent block hash value" in the block header. If the hash value of the parent block is changed, the child blocks will also be changed accordingly, and so on. When the block chain has sufficient length, the blocks will not be

changed unless all the blocks after the changed block are forced to be recalculated, and this hash algorithm adopted by the block chain and its unique chain data storage structure can keep the data consistent on the basis of multiple consensus and prevent the data from being tampered with. If the wrongdoers want to tamper with the copyright information, at least 51% of the nodes in the system must give permission and modification, which is extremely difficult to operate [13]. It can "dissuade" most of them to a large extent, so that the objectivity and authenticity of the we media works stored in the blockchain can be more safely and effectively guaranteed. At the same time, since the uploaded we media works have a unique electronic identity, it is possible to monitor and trace the work authentication, authorization, circulation path and the creation process of the works [14]. When a we media infringement occurs, the infringed person can directly collect and verify the evidence from the node, and reduce the cost of time, money and man power for proof and right defense by means of technical self-evidence. It is also possible to trace and query the suspected infringement information of high-frequency communication nodes in the Internet, realize early warning of infringement, change the current situation of passive rights defense of copyright owners of we media, and promote judicial remedies more effectively.

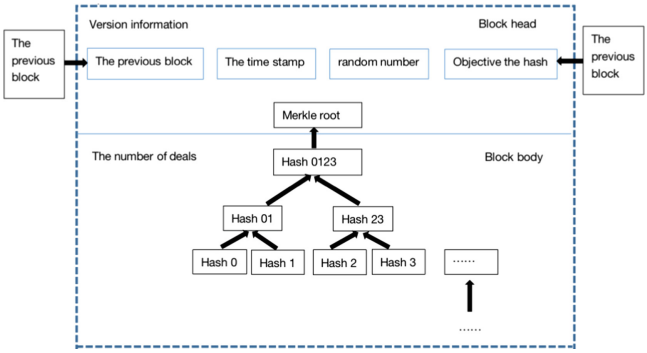


**Fig. 2.** Block data structure

In terms of specific cases, on November 26, 2019, Xinhua Zhiyun Technology Co., Ltd. Released the "Media Brain 3.0 Intelligent Solution for Financial Media Center", in which, with the support of blockchain technology, through Xinhua Zhiyun's copyright protection platform, with the help of its network-wide monitoring system, users can easily and efficiently understand the spread situation of their own work. Once the infringement is found, the creator of the we media content can fix the evidence with the help of electronic forensic tools, and the data of key links can be chained in real time to form effective evidence that cannot be tampered with. When a dispute occurs, through blockchain cross-chain technology, the court can also directly retrieve the relevant evidence for a quick trial, which greatly improves the judicial efficiency [15].

## 4 Cold Thinking Under the Hot Technology

Scholar Neil Bozeman once said, "Any technology is both a gift and a burden". While we see that blockchain provides certain technical support for copyright protection of we media, we should also let blockchain "step down" and face some of its challenges and problems.

In terms of confirmation of rights, the blockchain-based copyright registration system for we media is not yet fully mature and perfect, and the unique way of generating hash value also hinders the determination of originality of the content of we media works and the exercise of the author's right to modify to a certain extent.

In terms of the use of rights, the anonymity of blockchain also makes it difficult to identify the subjects of copyright transactions in we media, and the overly professional code text, if not supervised, can easily become a tool for illegal crimes [16]. In addition, there are many contradictions between the smart contract and the contract part of the Civil Code and the existing laws [17].

In terms of rights protection, the blockchain judicial deposit system is not yet complete, and the existing blockchain deposit platform is uneven, which may lead to the limitation of the utility of blockchain technology in the fixed evidence deposit in copyright disputes of we media works, and the inconsistency of blockchain evidence judicial certification rules may also lead to the problem of "different judgments in the same case".

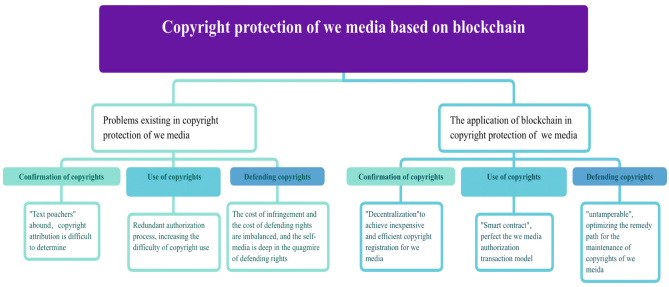


**Fig. 3.** The framework diagram of copyright protection of we media based on blockchain

## 5 Conclusion

As an emerging technology, the "decentralization" feature of blockchain can greatly improve the efficiency of we media rights, and meet the demand for rights verification of massive we media works. The "smart contract" can innovatively build a more simple and efficient trading platform for we media copyright, provide convenience for authorization and use of rights, and the "non-tamperable" feature of blockchain also enables we media works to be stored and queried in a safer and more effective way, and trace the whole process of application of works, so that once infringement occurs, it can help we media to effectively prove their rights. "decentralized", "smart contract", "non-tamperable"

and other technical features have shown unique advantages in the management and protection application of we media copyright (see Fig. 3), but at the same time, we should also see the shortcomings and loopholes of blockchain, a technology that is not fully mature and perfect, in the areas of confirmation, use and maintenance of rights. We should accelerate the pace of improving blockchain-related applications, guide the standardized application and development of blockchain technology in the field of we media copyright protection, further combat infringement in the field of we media, and build a harmonious and healthy network communication environment for we media copyright protection.

## References


1. Didi, C.: Research and implementation of trusted depository system based on blockchain smart contract. Nanjing University of Posts and Telecommunications (2019). https://doi.org/10.27251/d.cnki.gnjdc.2019.000069
2. Propaganda Department of the General Political Department, ed. Selected New Words on the Internet 2012 Revision. Beijing: PLA Publishing House, 2013.02
3. The network GuanYan. China We Media Marketing Market Analysis Report 2022 - Industry Supply and Demand Status quo and Development Trend Research. [EB/OL]. https://baogao.chinabaogao.com/guanggaochuanmei/550849550849.html
4. Jenkins, H.: Fans, Bloggers, and Gamers: Exploring Participatory Culture. New York University Press, New York and London (2006)
5. Ouyang, A.F., Tan, Z.L.: Study on the copyright protection model of orphan works utilization in libraries in the era of big data. Libr. Constr. **09**, 4–9 (2016)
6. Yifei, Z.: Research and Application of Digital Content Distribution Platform based on Blockchain Smart Contract. Beijing Printing Institute, Beijing (2021). https://doi.org/10.26968/d.cnki.gbjyc.2021.000084
7. Department of Information Technology and Software Service Industry, Ministry of Industry and Information Technology, and China Blockchain Technology and Industry Development Forum: White Paper on China Blockchain Technology and Application Development, p. 5, http://www.cbdforum.cn/bcweb/index/dd/7.do. 18 October 2016
8. Tschorsch, F., Scheuermann, B.: Bitcoin and beyond: a technical survey on decentralized digital currencies. IEEE Commun. Surv. Tutorials **18**(3), 2084–2123 (2016)
9. Qixun, M., Jing, L.: Research on the change of music copyright protection path from the perspective of blockchain. Technol. Law **06**, 17–24 (2018). https://doi.org/10.19685/j.cnki.cn11-2922/n.2018.06.003
10. Yuan, Y., Feiyue, W.: Current status and outlook of blockchain technology development. J. Autom. **42**(4), 481–494 (2016)
11. Szabo, N.: Smart contracts - Building Blocks for Digital Markets. https://kameir.com/smart-contracts/.2019
12. Jiaqian, Z.: Research on the Construction and Application of Blockchain-based Digital Copyright Industry Ecosystem. Beijing Printing Institute, Beijing (2021). https://doi.org/10.26968/d.cnki.gbjyc.2021.000022
13. José, D.M., Laura, S.-R.M.: Metadata-based measurements transmission verified by a Merkle Tree. Knowl.-Based Syst. **219**, 106871 (2021)
14. Feng, Z.: Around exploration of digital copyright protection based on blockchain technology. Technol. Law **01**, 59–70 (2017)

15. Danwen, Z.: Xinhua News Agency's exploration of copyright protection using blockchain technology. China Media Technol. **10**, 138–141 (2021). https://doi.org/10.19483/j.cnki.11-4653/n.2021.10.042
16. Jie, Y.: Research on Legal Issues in the Application of Blockchain Technology to Digital Copyright Management. Nanjing Normal University, Nanjing (2020). https://doi.org/10.27245/d.cnki.gnjsu.2020.000568
17. Xudong, L., Songyuan, M.: Research on blockchain smart contracts in the perspective of contracts of the civil code. J. Shanghai Normal Univ. (Philos. Soc. Sci. Ed.) **49**(05), 58–69 (2020). https://doi.org/10.13852/J.CNKI.JSHNU.2020.05.007

# Load Prediction of Electric Vehicle Charging Station Based on Residual Network

Renjie Wang(✉)

Beijingjiaotong University, Beijing, China
20120611@bjtu.edu.cn

**Abstract.** In the context of the rapid development of electric vehicles, the uneven space-time distribution of charging station load has caused the loss of efficiency and user experience. Therefore, the space-time prediction of charging station load has become an important research problem. In this paper, based on the St-ResNet model, which has achieved excellent results in space-time flow prediction in the field of traffic flow, we establish a space-time prediction model for a load of electric vehicle charging stations. In the model, we convert the spatial features of multiple charging stations with different geographical locations into 16*16 charging areas. And then, we fuse the three temporal features of the regional spatial distribution of the charging station load, and then use ResPlus to capture the long-distance spatial dependence of the charging load. Finally, we improved 3% to 20% compared with the baseline model.

**Keywords:** charging load · prediction · residual network

## 1 Introduction

With more and more attention paid to environmental protection, electric vehicles (EVs) have become the focus of society. Electric vehicles are pollution-free, low-noise, and have high performance beyond fuel vehicles, which has become a new trend in the development of the automotive industry in the future [1]. At present, the market scale of electric vehicles is developing rapidly. The Table 1 below shows the sales growth trend of electric vehicles in the Chinese market in the past 6 years.

**Table 1.** Production and sales of electric vehicles in China

| Year | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|
| Production | 51.7 | 79.4 | 127 | 124.2 | 120.5 | 294.2 |
| Sales | 50.7 | 77.7 | 125.6 | 120.6 | 120.5 | 291.6 |

With the rapid growth of the electric vehicle market, the number of charging stations supporting charging services has increased rapidly in recent years to meet the rapidly


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 132–143, 2023.
https://doi.org/10.1007/978-981-99-3618-2_13

expanding charging demand. However, for the charging station, the charging load fluctuates violently and is unevenly distributed in time and space, which leads to excessive congestion during the peak period of charging demand, and the queuing behavior of the charging station will increase the total cost by 1%–26% [2]. Therefore, it is a problem being studied by scholars to find some methods to predict the charging load of electric vehicle charging stations. Besides, space-time charging load forecasting of electric vehicles is the basis of charging station planning and vehicle network interactive decision-making.

## 2 Related Work

### 2.1 Load Predication of Charging Station

In the field of load prediction of electric vehicle charging stations, there have been many research results. Luo, Xu, and Qian used mathematical statistics to predict and modeled the charging load in the charging station by simulating the charging state of electric vehicles [3–5]. Paper [3] and paper [4] combine the dynamic road network information to predict the load distribution of each charging station by simulating the dynamic driving process of electric vehicles. Paper [5] uses queuing theory to predict the waiting time after vehicles arrive at the charging station. Paper [6] considers the influence of road conditions and temperature on EV power consumption and uses the Markov decision process to simulate the charging load of electric vehicles. Paper [7] uses a support vector machine (SVM) to predict the short-term load. Paper [8] is based on grey correlation theory, Based on the data samples of similar days, the support vector regression prediction model of daily load of electric bus stations is established. Paper [9] based on the historical charging load data of electric vehicles, a charging load prediction method based on a Bayesian regularized BP neural network is proposed. Paper [10] proposed a multi-scenario generation charging load prediction method based on the study of the time correlation between the charging load to be predicted and the historical charging load. Paper [11] carries out correlation analysis on the influencing factors of the operation state of the charging station under the multi-information coupling, and carries out a simulation based on the actual historical operation data of a city charging station. Paper [12] considering the parking laws and user behavior habits of large-scale users in different time and spaces, the Monte Carlo method is used to predict the charging demand. However, in the face of charging station load data with relatively discrete distribution in time and space, the prediction accuracy and efficiency of the above methods still have the potential to improve.

### 2.2 Spatio-Temporal Prediction

Subsequent studies have discovered that electric vehicle charging behavior data exhibit substantial spatial-temporal patterns [13]. In recent years, the field of deep learning models related to spatial-temporal prediction has experienced rapid advancements, achieving outstanding data fitting performance in traffic flow prediction tasks. Liu employs the ConvLSTM module to process short-term traffic flow information from nearby areas,

extracting spatial-temporal features, and using bidirectional LSTM to handle prediction points' historical traffic data, thereby revealing traffic flow data's cyclical characteristics [14]. Li introduced the Diffusion Convolutional Recurrent Neural Network (DCRNN) for traffic flow prediction, modeling traffic flow as a diffusive process on a directed graph. DCRNN employs a bidirectional random walk on the graph to capture spatial and temporal dependencies using an encoder-decoder structure with scheduled sampling [15]. Sun proposed a deep learning-based multi-branch prediction model for short-term urban traffic flow forecasting. The model accounts for external factors such as weather, unexpected traffic incidents, holidays, and other variables that impact traffic flow. The model utilizes the spatial-temporal traffic flow matrix and external factors as inputs to deduce and output the entire road network's future short-term traffic status (flow) [16]. Yu introduced a novel deep learning architecture, the spatiotemporal graph convolution network, comprising multiple spatiotemporal convolution blocks, each consisting of a graph convolution layer and a convolution sequence learning layer, to model spatial-temporal dependencies [17]. Zhao's Temporal Graph Convolutional Network (T-GCN) model combines Graph Convolutional Network (GCN) and Gated Recurrent Unit (GRU) [18]. GCN is employed to learn intricate topologies to capture spatial dependencies, while GRU is used to understand traffic data's dynamic changes to capture temporal dependencies. Guo's traffic flow prediction is based on Attention-based Spatial-Temporal Graph Convolutional Networks (ASTGCN) [19]. The ASTGCN model primarily consists of three independent components, each modeling one of the three temporal characteristics of traffic flow: proximity, daily, and weekly dependencies. In Zhang's research, Spatial-Temporal Residual Networks (St-ResNet) are used to model the closeness, periodicity, and trends of traffic congestion flow [20]. A branch of the residual convolution unit is designed for each attribute, with each unit modeling traffic congestion flow's spatial features. St-ResNet dynamically combines the outputs of the three residual neural network branches, assigning varying weights to different branches and regions. The integrated results are then combined with external factors, such as weather and the day of the week, to predict each region's final flow. More specifically, this study employs local CNN and LSTM to process spatial-temporal information, respectively, using a gated local CNN to model spatial dependencies based on dynamic similarities between regions, and employing a periodic shift attention mechanism to learn long-term periodic dependency models. Overall, this constitutes a deep learning framework for modeling long-term periodic information and time translation through the attention mechanism.

### 2.3 Innovative Content

Inspired by the St-ResNet model on the traffic flow prediction method, we establish a promoted residual neural network model to capture the temporal and spatial distribution characteristics of the charging station load and predict the temporal and spatial distribution of the charging station load. In addition, we added the "ResPlus" part based on the St-ResNet model to capture the long-distance spatial dependence of the load change of charging stations in the city.

# 3 Prliminaries

In this chapter, we will clarify the specific problems predicted by the model in this study, and introduce the original data set and data preprocessing process.

## 3.1 Datase and Data Preprocessing

We have obtained a data set of charging piles in Shanghai, which contains 201737 charging order records of 1521 charging stations in Shanghai from September 1, 2020, to February 28, 2021. The data attributes include charging station ID, charging station name, charging start time point, charging end time point, etc.

Based on the existing data sets, we obtained the real physical coordinates (longitude and latitude) of each charging station by calling the API of the Gaode map and Baidu map. We generated the following scatter diagram for these coordinates, as shown in the Fig. 1.

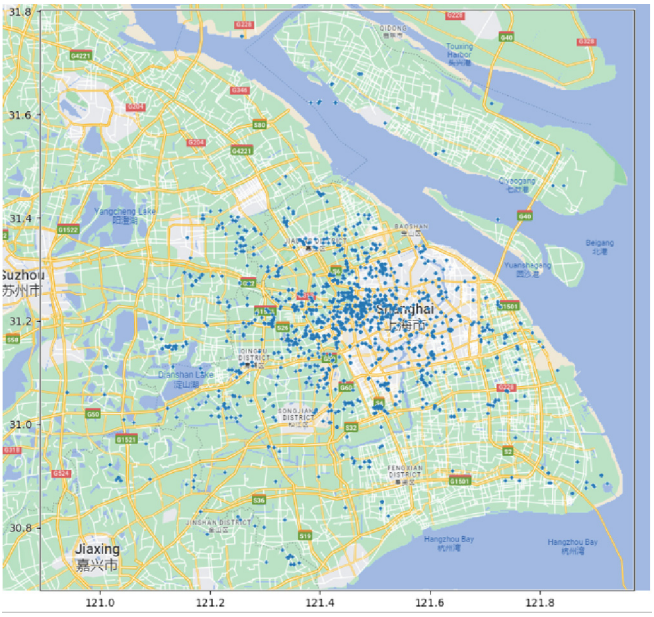


**Fig. 1.** Distribution diagram of charging station

Through observation in the diagram, we found that the charging stations at the edge of Shanghai city are scattered, and the number of charging orders per day is small, and the amount of data is not enough, which is unfavorable to the prediction. Therefore, we decided to select a relatively dense location of charging stations in Shanghai for grid processing.

### 3.2 Problem Formulation

- Definition 1 (Charging Zone):

In this study, we divided the dense area of charging stations in Shanghai into 16 * 16 grids as 256 Charging Zones, and filled all charging stations into the grids of their longitude and latitude positions, as shown in the Fig. 2 below.

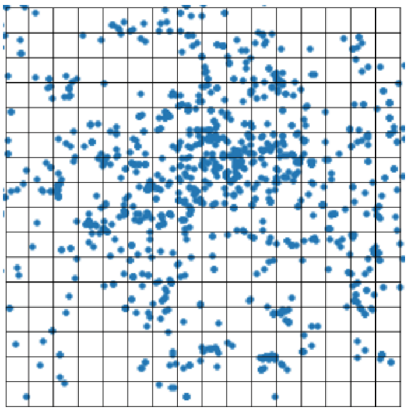

**Fig. 2.** Charging zone filled with charging stations

On this basis, we will aggregate the spatial features of scattered charging stations into more regular spatial features of the 156 Charging Zones, which is more conducive for us to input the spatial features into the model.

- Definition 2 (Charging Period):

In this study, we divided the 6-month time period from '2020/9/1 00:00:00' to '2021/2/28 23:00:00' into 4343 time intervals, each of which is 1 hour long. For each charging period, we will allocate the load of the charging station to each charging period of each charging zone (16 * 16) according to the charging start and end times of all charging orders, and finally generate a three-dimensional matrix of 16 * 16 * 4343, representing the change process information of the number of charging zones with a charging station load of 16 * 16 in the 4343 charging periods, as shown in the Fig. 3.

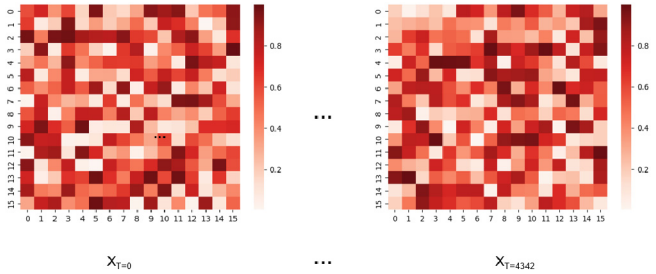


**Fig. 3.** Spatial region distribution of charging station

### 3.3 Charging Load Prediction Tas

Given the historical observations,

$$\{\mathbf{X}t \mid t = 0, 1, \cdots, n-1\} \text{ predict } \mathbf{X}t.$$

## 4 Improved Model

As shown in the Fig. 4, the overall framework of improved model model is mainly composed of two parts, one is information input, and the other is ResPlus. The information input of $\mathbf{X}_t$ includes three parts: trend of week, trend of day, and near hour. The ResPlus unit is a designed by us to directly capture long-distance dependence. We make statistics on the usage of charging stations in each area within each hour to form a series of change diagrams of the usage of charging stations. These change diagrams will normalize min

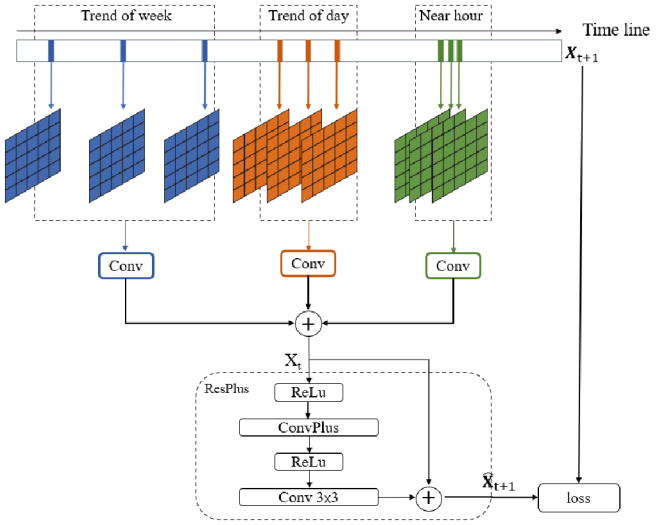


**Fig. 4.** Model architecture

max to [−1, 1], and then the three parts of the information input will be fused and sent to ResPlus unit for processing. After that, the result will be mapped to [−1, 1] through a tanh activation function. Finally, we get the predicted $\widehat{X}_{t+1}$ to evaluate the loss by comparing with original $\mathbf{X}_{t+1}$.

### 4.1 Temporal Dependence

As the parts of information input: The trend of week part $\mathbf{X}_{i-l_w}$ means X of the same hour period, of the same day of week and of the first $l_w$ weeks of the forecast time period; The trend of day $\mathbf{X}_{i-l_d}$ means X of the same hour period of in the first $l_d$ days of the forecast time period; The near hour means $\mathbf{X}$ of the first $l_h$ hours of the forecast period.

### 4.2 Spatial Dependence

There is a spatial dependence between the charging load in a region and the charging load in its adjacent region. In a period, the dependence includes 1. The area with a saturated charging load will transmit the excess charging demand of this area to the adjacent area. 2. There is also a competitive relationship between adjacent charging areas. 3. In terms of user behavior, users' charging needs will migrate to further areas with their personal driving behavior. Therefore, long-range spatial dependence should also be taken into account.

In the data set, although the coverage area we selected does not cover the whole city, the coverage area is still very large, which will involve the spatial dependence of distance and distance. In the context of this problem, convolution neural network has shown a very superior ability to extract spatial features. Although the size of the convolution kernel is limited, we can extract long and short spatial features by gradually expanding the receptive field by increasing the depth of the convolution network. However, when we increase the depth of the convolution network, we may encounter the problem of gradient disappearance to solve this problem, we choose St-ResNet as the infrastructure to solve the problem of spatial-temporal prediction of charging load.

### 4.3 ResPlus

The optimization of the St-ResNet model is reflected in the settings of the ResPlus part. At present, many models based on deep learning are proposed for spatiotemporal prediction, mainly including two basic structures: RNN based structures such as ConvLSTM [21], periodic-CRN [22], and St-ResNet [20]. However, the training based on RNN structure always consumes a lot of time. Therefore, we choose St-ResNet based on CNN structure as the basic model. Convolutional neural network (CNN) has powerful ability to capture image region features. However, with the development of battery technology of electric vehicles, the range of activities of electric vehicles is becoming wider and wider, and it can cause loads on charging stations that are farther away in space, resulting in the long-distance space becoming more and more important. The long-distance space dependence varies from region to region, which makes it difficult for a stack of scrolls to effectively capture this relationship. Therefore, we designed ConvPlus to capture the long-distance

spatial dependence of the load change of the charging station. As shown in Fig. 3, the ResPlus unit adopts a ConvPlus and a common convolution. We separate some channels of the common convolution to capture the long-distance spatial dependence of each region.

## 5 Training

The training procedure of model as shown in the Table 2.

**Table 2.** Model training proceduce

| Model Training Procedure |
|---|
| **Input:** historical observations: $\{ \mathbf{X}_1, \ldots, \mathbf{X}_{i-1} \}$<br>trend of week: $l_w$;<br>trend of day: $l_d$;<br>near hour: $l_h$ |
| **Output:** learned model |
| / construct the training data $\mathbb{D}$ |
| 1 $\mathbb{D} \Leftarrow \emptyset$ |
| 2 **for** all available time interval |
| 3 $\mathbf{X}_i^h = \left[\mathbf{X}_{i-l_h}, \mathbf{X}_{i-(l_h-1)}, \cdots, \mathbf{X}_{i-1}\right]$ |
| 4 $\mathbf{X}_i^d = \left[\mathbf{X}_{i-l_d}, \mathbf{X}_{i-(l_d-1)}, \cdots, \mathbf{X}_{i-d}\right]$ |
| 5 $\mathbf{X}_i^w = \left[\mathbf{X}_{i-l_w}, \mathbf{X}_{i-(l_w-1)}, \cdots, \mathbf{X}_{i-w}\right]$ |
| 6 put training instance$(\{\mathbf{X}_i^h, \mathbf{X}_i^d, \mathbf{X}_i^w, \}, \mathbf{X}_i)$ into $\mathbb{D}$ |
| 7 end // $\mathbf{X}_i$ is the target at time i |
| //train the model |
| 8 initialize all learnable parameters θ in the model |
| 9 **Repeat** |
| 10 randomly select a batch of instances D from $\mathbb{D}$ |
| 11 optimize θ using Adam and D |
| **12 Until model overfitting** |

## 6 Performance Evaluation

### 6.1 Baslines

We compare the improved St-ResNet model with the four baselines listed below:

- HA: It forecasts crowd inflow and outflow using the average of prior inflow and outflow values for the respective periods;

- VAR [23]: Due to its numerous parameters and high computing cost, Vector Auto-Regressive can capture pairwise associations among all flows.
- ARIMA [24]: Auto-Regressive Integrated Moving Average combines moving average (MA) with a difference process using auto-regression (AR)
- ConvLSTM [21]: This neural network (convolution and LSTM combined) captures both spatial and temporal data but requires a lengthy training process due to its recurrent structure.

### 6.2 Metrics and Parameters

We employ Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) as metrics:

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}\left\|\mathbf{X}_i - \widehat{\mathbf{X}}_i\right\|_2^2}$$

$$MAE = \frac{1}{T}\sum_{i=1}^{T}\left|\mathbf{X}_i - \widehat{\mathbf{X}}_i\right|$$

where and $\mathbf{X}_i$ and $\widehat{\mathbf{X}}_i$ denote the ground-truth and the prediction at the $i^{th}$ time interval. $T$ is the total number of samples in the testing data. RMSE is also used as the loss function of improved St-ResNet model.

### 6.3 Performance Comparison

See Table 3

**Table 3.** Performance comparsion

| Model | RMSE | MAE |
|---|---|---|
| HA | 1.1225 | 0.5964 |
| VAR | 1.4378 | 1.1589 |
| ARIMA | 1.5512 | 0.6857 |
| ConvLSTM | 0.913 | 0.5372 |
| Improved St-ResNet | 0.8847 | 0.5045 |

### 6.4 Effects of Hyper Parameters

The impact of the number of separated channels for long-range spatial dependence is depicted in the Fig. 5. The model benefits immediately as the number of separated channels increases from 0 to 1, indicating that the city's long-range spatial dependence

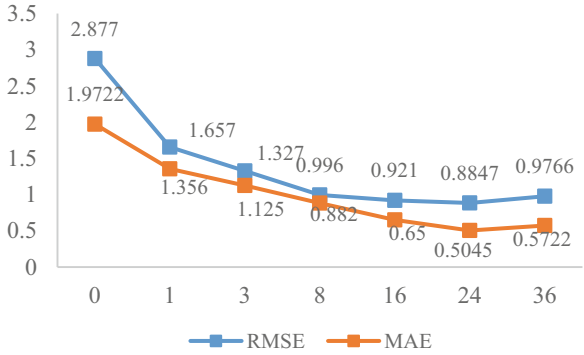


**Fig. 5.** Influence of Amount of ConvPlus Channels

is strong and is considered by the ConvPlus structure. The model performs better with 24 separated channels and worse with more separated channels, which shows that both short-range spatial dependency and long-range spatial reliance are significant.

The effect of the total number of channels in the convolution and ConvPlus is depicted in Fig. 6. We choose a total channel count between 16 and 128. The results demonstrate that, despite greater computational costs, our approach will continue to outperform as the number of channels increases.

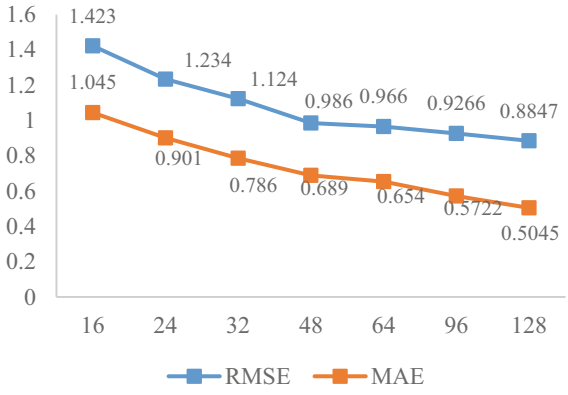


**Fig. 6.** Influence of Amount of All Channels

## 7 Conclusion

In order to predict the change of the load of electric vehicle charging station in time and space, we establish a model based on St-ResNet (space-time residual network) to capture the time characteristics (including "trend of week", "trend of day" and "near hour") and spatial regional distribution characteristics of the charging load, and finally achieve a better prediction effect (3%–20%) than the existing baseline model.

The main contributions of the paper include:

- For the date process method, we transform the spatial feature of discrete longitude and latitude coordinates of the charging station into a more structured charging zone, and the time feature of charging time point into a time slice with an hour slice with and hour as the length. Such a data processing method can improve the prediction effect.
- We proposed an improved St-ResNet model for space-time prediction of charging load of charging stations. And the prediction effect is improved 3%–20%, compared with the baseline model in this field;
- We build a ResPlus part in the model, which can capture the long-distance spatial dependence of each region. And it helps the model is more suitable for field of charging load precidtion. This is because users' charging behavior of selecting a charging stations is related to the user's long-distance moving trajectory, which leads that the charging load has a long-distance dependency.

**Acknowledgment.** This research was supported by the Beijing Logistics Informatics Research Base. I would like to sincerely thank Professor Zhang Zhenji and Professor Gong Daqing for their careful guidance of the paper. In addition, many thanks go to Beijingjiaotong university and ICIR for their support.

## References

1. Chen, L., Zhang, Y., Figueiredo, A.: Overview of research on charge and discharge load prediction of electric vehicles. Power Syst. Autom. **43**(10), 177–191 (2019)
2. Keskin, M., Laporte, G., Çatay, B.: Electric vehicle routing problem with time-dependent waiting times at recharging stations. Comput. Oper. Res. **107**, 77–94 (2019)
3. Luo, Z., Hu, Z., Song, Y.: Study on plug-in electric vehicles charging load calculating. Autom. Electric Power Syst. **35**(14), 36–42 (2011)
4. Xu, G., Yang, G., Xu, Z.: Impacts of electric vehicle charging on distribution networks in Denmark. Autom. Electric Power Syst. **35**(14), 18–23 (2011)
5. Qian, K., Zhou, C., Allen, M.: Modeling of load demand due to EV battery charging in distribution systems. IEEE Trans. Power Syst. **26**(2), 802–810 (2011)
6. Xing, Q., Chen, Z., Leng, Z.: Electric vehicle path planning and charging navigation strategy based on real-time traffic information. Chin. J. Electr. Eng. **40**(2), 534–549 (2020)
7. Kong, X., Zheng, F., Zhijun, E.: Short-term load forecasting based on deep belief network. Autom. Electric Power Syst. **42**(5), 133–139 (2018)
8. Xu, X.B., Liu, W.X., Zhou, X.: Short-term load fore-casting for the electric bus station based on GRA-DE-SVR. In: 2014 IEEE Innovative Smart Grid Technologies-Asia (ISGTASIA), Kuala Lumpur, Malaysia, pp. 388–393. IEEE (2014)
9. Xiao, X., Pan, M., Duan, D.: Electric vehicle charging load prediction method based on grid division. J. Electric Power Sci. Technol. **36**(3), 19–26 (2021)
10. Huang, N., Liu, D., Cai, G.: Electric vehicle charging load interval prediction based on multi correlation daily scenario generation. Chin. J. Electr. Eng. 1–11. https://doi.org/10.13334/j.0258-8013.pcsee.201806
11. Zhao, Y., Li, Y., Su, S.: Research on information prediction method of charging station under multi information coupling. Electric Measur. Instrum. 1–10 (2022). http://kns.cnki.net/kcms/detail/23.1202.TH.20210519.1311.004.html

12. Hocevar, M., Strok, B., Blagojevic, B.: Prediction of cavitation vortex dynamics in the draft tube of aFrancis turbine using radial basis neural networks. Neural Comput. Appl. **14**(3), 229–234 (2005)
13. Ge, S., Zhou, H., Liu, H.: Collaborative planning of Expressway charging station and service area considering the impact of vehicle accidents. J. Power Syst. Autom. **31**(3), 1–8 (2019)
14. Liu, Y., Zheng, H., Feng, X.: Short-term traffic flow prediction with Conv-LSTM. In: 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6. IEEE (2017)
15. Li, Y., Yu, R., Shahabi, C.: Diffusion convolutional recurrent neural network: data-driven traffic forecasting. arXiv preprint arXiv:1707.01926 (2017)
16. Sun, S., Wu, H., Xiang, L.: City-wide traffic flow forecasting using a deep convolutional neural network. Sensors **20**(2), 421 (2020)
17. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875 (2017)
18. Zhao, L., Song, Y., Zhang, C.: T-GCN: a temporal graph convolutional network for traffic prediction. IEEE Trans. Intell. Transp. Syst. **21**(9), 3848–3858 (2019)
19. Guo, S., Lin, Y., Feng, N.: Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, pp. 922–929 (2019)
20. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
21. Shi, X., Chen, Z., Wang, H.: Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems, vol. 28 (2015)
22. Zonoozi, A., Kim, J., Li, X.L.: Periodic-CRN: a convolutional recurrent model for crowd density prediction with recurring periodic patterns. In: IJCAI, pp. 3732–3738 (2018)
23. Hansen, B.E.: Time series analysisjames d. hamilton princeton university press, 1994. Econometric Theory **11**(3), 625–630 (1995)
24. Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: Time Series Analysis: Forecasting and Control. Wiley, Hoboken (2015)

# Study on Copyright Infringements of We Media Based on Media Integration Blockchain

Tongtong Yu and Liang Wang(✉)

School of Economics and Management, Beijing Institute of Graphic Communication, Beijing, China
wangliang@bigc.edu.cn

**Abstract.** With the maturity and improvement of blockchain theory and technology, many media have created platforms based on blockchains such as Civil, Decentralized News Network, Steemit and so on. All these successful cases showed us the potential in protecting media industry especially in digital copyright area. In 2021, the Xinhua News Agency's search engine ChinaSo created Media Integration Blockchain to protect copyright of news media. This paper introduces the design of the Media Integration Blockchain and try to give some advice to upgrade it to increase the range of it. The paper focuses on not only news media but also on We Media. The purpose of this study is to avoid the copyright infringements and contribute to the protection of digital copyright based on the application of blockchains in China's digital copyright protection.

**Keywords:** Blockchain · We Media · Media Integration Blockchain · Copyright Infringements

## 1 Introduction

We media is a form of participatory user-generated content, it allows anyone to post and upload files, information, and news without a formal editorial moderation or filter process [1], these contents including picture, music, video and article. 2013 is the first year of an era of China's we media. According to the 49th Statistical Report on China's Internet Development based on China Internet Network Information Center in February 2022, as of December 2021, China had 1,032 million netizens, up by 414 million over December 2013, and its Internet penetration had reached 73.0%, which implies us that the users of we media has increased quickly.

Due to the characteristics of virtualization and concealment of we media, copyright infringements come into people's sight. This illegal activity persist despite repeated efforts to counter it, meanwhile, it undermine user's trust in online content.

At present, many media have used digital fingerprint [2] and digital watermarking [3] to avoid copyright infringement. Along with the wave of information technology revolution, the digital protection based on blockchain is growing at a break-neck speed.

This work is partially supported by the Key Program of Beijing Social Science Fund under contract with No. 19JDXCA001.


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 144–151, 2023.
https://doi.org/10.1007/978-981-99-3618-2_14

Blockchain is a chain of blocks that contains information. It was originally described in 1991 by a group of researchers and was originally intended to timestamp digital documents so that it's not possible to backdate them or to tamper with them.

In 2008, Satoshi Nakamoto came up with blockchain in a technical paper. In that paper, he explained conception of blockchain and created the digital cryptocurrency Bitcoin [4].

After that, decentralized and distributed ledger technologies have received widespread attention in all over the world. Security is the most dominant feature of blockchain. That is to say once some data has been recorded inside a blockchain, it becomes very difficult to change it because of the hash. Each block has a different hash, changing something inside the block will cause the hash to change. What's more, the security of a blockchain also depend on the proof-of-work mechanism to calculate hash [5]. Decentralization, openness, transparency, immutability [6], tamper-proof and whole process trace are all blockchain's characteristics [7]. This new technology are widely used in various fields. For instance, it can be used in finance to supply-chain-related industries [8], it also can be used in social service, risk management, healthcare facilities [9], medical field, education credit platform [10], and so on.

This article will discuss the application of blockchain in copyright protection and provide advice for improvement it.

In a nutshell, the main contributions of this essay are as follows.

1) We briefly conclude the application of the basis blockchains' circumstance in news media area and contrast the different using situations between China and foreign countries.
2) We give four suggestions to improve the efficiency of the blockchains' use in we media which also can be used in other platforms.

## 2 The Application of Blockchain in Copyright

With the increasing use of blockchain, digital publications and copyright protection are becoming far more integrated to spark new development opportunities which drive we media's copyright protection to achieve a new stage of leapfrog development.

Annabel Tresise, Jake Goldenfein and Dan Hunter concluded five main areas which blockchain can do for avoiding copyright infringements including registration processes, the creation of open digital rights management systems, the development of automated licensing, the reduction of need for collective management organizations, and the identification of orphan works [11]. Some specific models are proposed in copyright protection area like the framework of music copyright protection system [12] and the novel public participation consortium blockchain system [13].

At present, many media tried to use this technology to solve the problem of copyright infringement. For example, users who join the Civil Media Company's platorm, should submit an application to show ability and identity in news media, after the committee's censorship, they can post on it. The blockchain helps Civil to create a trusted web. Steemit used the Digital Specie Reward Incentive Consensus to inspire users create works so that they can collect works as much as possible. Decentralized News Network Encourage all users to read and write works, and it used reputation mechanism distinguish users which

is an innovative method. Table 1 briefly shows the blockchain application in news media which helps us understand the use of blockchain in western countries.

**Table 1.** Blockchain application in media

| media | Circumstance of media | | |
|---|---|---|---|
| | blockchain | characteristic | flaw |
| Civil | Ethereum blockchain | To create the Trusted Web | It can not guarantee news authenticity |
| Decentralized News Network | Ethereum blockchain | Encourage all users to read and write works | News people on platform may out of time |
| Steemit | Steem blockchain | Adopt reputation mechanism to govern users | It can not limit or punish the user who infringe the law |

What should be noted is Ethereum blockchain and Steem blockchain are Public Blockchains, which is very different from Consortium Blockchains and private blockchains.

Public blockchains allow read access and ability to create transactions to all people who use blockchain. However, private blockchains bound access to the predefined list of users that means private blockchains are more safer and resilience to attacks in a way [14].

Consortium blockchains are something like a combination of public blockchains and private blockchains. In the consortium blockchain, access and upload files are only grant for members who engage in consortium [15].

All these blockchains use consensus mechanism to guarantee security and reliability. The remarkable thing is Media Integration Blockchain is based on the consortium blockchain.

## 3 The Application of Media Integration Blockchain.

There is a abroad consensus that blockchain technology may be an efficacious way to solve the current copyright infringements in China [16], and that is the foundation of why Media Integration Blockchain were created. We want to extend the use of blockchain to a broader application. That is to say blockchain can be used in not only news media, but also in we media.

The basic principle of Media Integration Blockchain is to generate codes for each users who upload files, and then the blockchain needs to store the information of copyright for the original code on the nodes. The blockchain will trace all the user who use these codes and every code can be verify through the construction of blocks. Finally, users can use these data as evidences to sue people who infringe.

### 3.1 Design of Media Intergration Blockchain

Media Integration Blockchain is a platform invented by the Media Copyright Protection Association of China which including seven mainstream media in China, such as Xinhua News Agency, People's Daily, China Media Group, China News Service and so on. It has two user interface, one is closed platform that only services for associations and another is open platform which can be use by we media.

The platform can be used to inquiry works, trace content, monitor users and store certificate and so on. All the records are based on time of Bei Dou Navigation Satellite System.

To use the functions of Media Integration Blockchain, people need to sign up a new account with binding phone number and passing the real-name certification system. After that, users can upload their works to get their proof of existence. One thing that's important to note is the number of upload works are limited.

The platform can track almost works posted on the Internet. When some works are suspected infringing, the system will contrast the work and inform the user to take action for copyright protection,

If users find copyright infringement happens, they can take depositions through screenshot and screencast by the Application Programming Interface (API) which must be purchased. All these evidences will be synchronized on Beijing Internet Court and its Tianping Blockchain and also can be used in court as direct evidences. Up to June 2022, Media Integration Blockchain has got 45 nodes, 122,684 blocks and 428,304 ledger data.

To address these steps, we draw a flow chart to depict the general operational procedure of Media Integration Blockchain in Fig. 1.

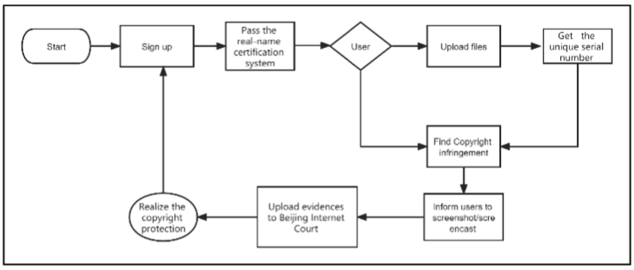


**Fig. 1.** How the platform work.

According to Fig. 1, the key functions of this platform are uploading files and finding copyright infringements automatically. Furthermore, all the copyright information are stored permanently in the blockchain, making the records of copyright more authority.

### 3.2 Strengths and Shortcomings

Media Integration Blockchain has helped many users to figure out copyright infringements. First of all, it helps users to define the ownership of copyright. Next, the connection

of Beijing Internet Court can help judges to confirm responsibility identification. Last but not least, it can save time and decrease cost of transaction.

However, there are still many limitations in this platform. Firstly, different media use different blockchains which means China don't have an official blockchain admitted by government and law, information cannot be shared by different blockchains. Secondly, this platform only appeals users who have well aware of copyright protection. People who do not have health citizen consciousness and sense of social responsibility will not take part in it [17]. Thirdly, there are still some legal issues existing in platform which have been solved in copyright protection like "the identification of real creator" and "the cognizance of responsible subject of transaction" [18].

In summary, most of the current blockchain plan used in copyright protection are not perfect which more or less exist some flaws.

## 4 Adcice of Media Integration Blockchain

As we know, the most serious problem on copyright infringement of we media are changing the works and pirating the works. In fact, not only works of journalism face the problem with copyright infringements, but also works of we media. Both of them have to solve the problems like the difficulties in burden of proof, the high cost of safeguard and the low compensation of verdict. To counter these questions, we propose some suggestions for Media Integration Blockchain to protect copyright better.

### 4.1 Use Rewards to Appeal Users and Expand the Platform

The platform needs to encourage users to record their works on the blockchain so that the works on blockchain can keep growth continuously. Besides, if someone who deals the copyright infringement successfully or invites new users for the platform, he or she should be given some discounts when they use the API.

To some extent, the more works are on blockchain, the less copyright infringements will happen. In addition, Media Integration Blockchain should try to invite key opinion leader to share their successful cases of copyright protection, that can help people to arouse the awareness of copyright protection.

### 4.2 Cooperate with Other Blockchains

To strengthen copyright protection effectively and keep sustainable development of the platform healthily, consortium needs more media and more verification works. Although there's no public blockchain used in China in copyright area, Media Integration Blockchain needs to avoid to be an information silo and try hard to break down barriers of diverse platforms.

### 4.3 Establish the Copyright Trading System

The western developed countries' experience have proved that build an official copyright trading system can reduce the copyright infringement. For instance, England built

Copyright Hub and America established Copyright Clearance Center, both of the two platforms helped the reuse of scattered copyright works.

Although China have built some platforms for trading copyright, all these platforms are created by enterprisers, the number of the works and the range of impact are limited.

### 4.4 Create Credit System on the Platform

The platform needs to create a credit system for users. Artificial intelligence can be used to assess the users' credit point on a regular basis to make sure all users are treated fairly.

In addition, platform can set levels for different point of users. For example, divide users in 1–5 level, addition and subtraction items are adopted for user's evaluation. The higher level users get, the lower point users have. Different level means different punishments and limitations. If the user was in level 1, he or she can enjoy all the functions of the platform by a discount price which is at half of the original price. If the user was in level 2, he or she can get a 20% discount for all the services on the platform. If the user was in level 3, he or she can enjoy all the serves in a normal price. If the user was in level 4, he or she would be limited to use the upload function, the number of the upload file would be limited no more than 5. If the user was in level 5, he or she may have a higher price to use the API, that's about twice as they usually cost. What's more only could use a few functions of the platform, the number of the automatic detection of the copyright infringements would be limited.

As illustrated in Fig. 2, our suggestions are concentrating on different processes of Media Integration Blockchain. We think in a process of develop user, the platform needs to create reward mechanism to mobilize users' enthusiasm. In the process of key functions, collaborate with other blockchains can realize win-win cooperation. When copyright infringements happen, except using legal measures, people can use copyright trading system to resolve disputes. After the whole process, the platform can use credit system to evaluate users.

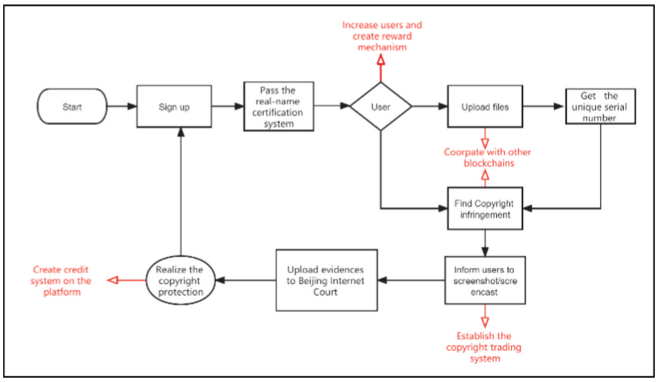


**Fig. 2.** Suggestion in different processes.

These advice optimize the user experience and improve the platform design process. All these methods can help users standardize behavior and guide them to focus on copyright protection.

## 5 Conclusion

Cyberspace concerns the future of humanity, and the future of we media should be jointly built by all users. Blockchain technology has offered new opportunities to rethink the way of copyright protection. In this essay, we introduce basic background and application of blockchain in media and we try to deconstructing Media Integration Blockchain. In order to explore the more possibility of it, we give four suggestions to make it better. It is important to build an online environment that is fairer and more equitable, more open and inclusive, safer and more stable, and more vibrant. We hope that this platform would contribute to all the copyright protection of we media and global internet development.

## 6 Limitation and Future Work

Our research only study on the original basis of Media Integration Blockchain. In fact, these advice also can be used on other platforms too. Our goal is aiming to build a platform which can provide the gamut of copyright services in China. Nonetheless, Media Integration Blockchain has already been a platform which has relatively mature conditions and realistic significance until now.

A promising direction for further study of copyright infringement on we media is to find an effective way to identify original works to avoid copyright preemptive registration and try to distinguish the level of infringement through blockchain.

If possible, we need to research and collect more cases and data from copyright protection platforms base on blockchain. Turning the platform into a multi-function platform which could also provide copyright trade and user evaluation for copyright.

**Acknowledgment.** This work has been partially supported by the Key Program of Beijing Social Science Fund under contract with No. 19JDXCA001.

## References

1. Deuze, M.: Participation, remediation, bricolage: considering principal components of a digital culture. Inf. Soc. **22**, 63–75 (2006)
2. Gold, S.: Understanding the digital fingerprint. Netw. Secur. **12**, 15–18 (2013)
3. Katzenbeisser, S., Petitcolas, F.A.P.: Digital watermarking. Artech House Lond. **2**, 2 (2000)
4. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash. System [DB/OL] (2008). https://bitcoin.org/bitcoin.pdf
5. Finck, M., Moscon, V.: Copyright law on blockchains: between new forms of rights administration and digital rights management 2.0. IIC – Int. Rev. Intellect. Property Compet. Law **50**(1), 77–108 (2018). https://doi.org/10.1007/s40319-018-00776-8
6. Kshetri, N.: Blockchain's roles in meeting key supply chain management objectives. Int. J. Inf. Manag. **39**, 80–89 (2018)

7. Zhang, J.: Application of blockchain technology in financial media industry. Video Eng. **7**, 6–8 (2021)
8. Denter, N.M., Seeger, F., Moehrle, M.G.: How can Blockchain technology support patent management? A systematic literature review. Int. J. Inf. Manag. **68**, 102506 (2022)
9. Liu, L., Zhang, W., Han, C.: A survey for the application of blockchain technology in the media. Peer-to-Peer Network. Appl. **14**, 3143–3165 (2021)
10. Monrat, A.A., Schelén, O., Andersson, K.: A survey of blockchain from the perspectives of applications, challenges, and opportunities. IEEE Access **7**, 117134–117151 (2019)
11. Jing, N., Liu, Q., Sugumaran, V.: A blockchain-based code copyright management system. Inf. Process. Manag. **58**, 102518 (2021)
12. Bai, Y., Hu, Q., Seo, S.H., Kang, K., Lee, J.J.: Public participation consortium blockchain for smart city governance. IEEE Internet Things J. **9**, 2094–2108 (2021)
13. Nana, W., Hui, X., Feng, X., Lei, C.: The algorithmic composition for music copyright protection under deep learning and blockchain. Appl. Soft Comput. **112**, 107763 (2021)
14. Tresise, A., Goldenfein, J., Hunter, D.: What blockchain can and can't do for copyright. Australian Intell. Property J. **28**, 144–157 (2018)
15. Liu, C., Zhang, X., Medda, F.: Plastic credit: a consortium blockchain-based plastic recyclability system. Waste Manag. **121**, 42–51 (2021)
16. Savelyev, A.: Copyright in the blockchain era: promises and challenges. Comput. Law Secur. Rev. Int. J. Technol. Law Pract. **34**, 550–561 (2018)
17. Feng, W., Weiyuan, W.: Identification and remedy of self-media video infringement—taking the Bilibili for example. J. Henan Judicial Police Vocat. Coll. **19**, 65–69 (2021)
18. Qi, W.: Research on copyright protection of short video in the era of we media. Sci. J. Humanit. Soc. Sci. **3**, 76–82 (2021)

# The Determinants of Carbon Emissions in Macau: Based on the Analysis of STIRPAT, EKC, and LMDI Models

Wenlan Liang, Ruizhen Pan(✉), Yilin Cao, Wentong Wang, Shixuan Lin, and Yujia Zhao

School of Business, Macau University of Science and Technology, Macau SAR, China
1832966292@qq.com

**Abstract.** The global spotlight has been increasingly drawn to the topic of the greenhouse gas effect. This research endeavors to explore the factors that influence Macau's carbon emissions by analyzing yearly figures from 1991 to 2020 and then proffering strategic proposals for the Macau government. The empirical results of the study demonstrate a long-term cointegration relationship, which is based on the natural logarithm of carbon emissions as a dependent variable, and population, real GDP per worker, financial development, industrial structure upgrading, and fixed capital formation as independent variables. The LMDI model shows that, in Macau, the environmental Kuznets curve is present, with energy consumption and carbon emissions being the primary contributors to carbon emissions. Moreover, real GDP per worker, industrial structure improvement, and fixed capital formation have a beneficial effect on carbon emissions, while financial development has a detrimental effect.

**Keywords:** Carbon Emissions · Environmental Kuznets Curve · Greenhous Gas Effect

## 1 Introduction

Recently, the Greenhouse gas effect has substantially caused environmental degradation, and carbon emissions are a serious issue in the global climate-change problem. The world's rapid economic expansion has caused a steady rise in carbon emissions, which has had a detrimental effect on global economic activities and the quality of life for humans [1]. According to the Paris Climate Agreement of 2016, governments have dedicated themselves to dealing with pollution problems.

Relative environmental theory suggests that population is a major determinant of environmental quality [2]. Moreover, prior studies have shown that economic growth can also have a detrimental effect on environmental quality [3], though it may decrease carbon emissions [4]. Moreover, environmental degradation has had a detrimental effect on sustainable economic growth [5].

Over the past two decades, Macau has experienced rapid economic growth due to the liberalization policy of casino gaming in 2002. This results that large amounts of


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 152–161, 2023.
https://doi.org/10.1007/978-981-99-3618-2_15

tourists have flown into Macau, which causes pollution of transportation and lodging. The issue of economic growth and environmental quality is of great importance to the Macau government, thus requiring more attention.

The STIRPAT model is employed in this study to explore the factors influencing carbon emissions in Macau. Moreover, it is examined whether the environmental Kuznets curve phenomenon in Macau is eliminated and the LMDI model is used to break down the factors impacting carbon emissions.

The study has five sections. In the initial part, the introduction is provided. Subsequently, the literature review is outlined, and hypotheses are put forward. Data and methodology are then discussed in the third section, followed by empirical results in the fourth. Lastly, the ultimate conclusion is presented.

## 2 Review of Literature

### 2.1 STIRPAT Model

The STIPAT model, proposed by Ehrlich and Holdren [2], was created to illustrate the influence of human activity on environmental quality. Dietz and Rosa [6] then adapted this model to create the TIRPAT model, which evaluates the effect of population, property, and technology on environmental degradation through stochastic regression.

Dietz and Rosa's STIRPAT model is frequently employed to investigate the demographic, economic, and technological components that are responsible for environmental deterioration.

$$I_i = \alpha P_i^b A_i^c T_i^d \varepsilon_i \tag{3.1}$$

Logarithms are then employed for all variables in both sides of the equation.

$$logI_i = a + blogP_i + clogA_i + dlogT_i + \varepsilon_i \tag{3.2}$$

where I denote pollution emissions, P denotes population size, A denotes real GDP per capita, and T denotes technology. ε is the error item.

The STIRPAT model has been frequently employed in prior studies to investigate the elements that influence carbon emissions. Khan, Saleem, & Fatima [7] applied the model to Bangladesh, India, and Pakistan, and determined that financial growth has a detrimental effect on the correlation with emission, apart from India. In Pakistan and India, income disparity lessens emission, whereas Bangladesh's outcome is the reverse. Utilizing the STIRPAT model, Li et al. [8]. Investigated the effect of manufacturing structural transmission and natural resources reliance on carbon emissions by taking population as a representation of population, GDP per capita as an indicator of economic factor, and the ratio of value-added of the second industry to value-added of the third industry as a representation of technology, with energy usage having a beneficial effect on emission.

### 2.2 Environmental Kuznets Curve

Grossman and Krueger [9] proposed an inverted U-shaped relationship between economic growth and environmental degradation, which they termed the Environmental Kuznets Curve (EKC). This curve indicates that, initially, environmental degradation rises in tandem with economic growth, but then begins to decline once per capita income has attained the benchmark point.

In South Africa, Shahbaz, Tiwari, and Nasir [10] uncovered a U-shaped correlation between real GDP per capita and carbon emissions per capita. In Bangladesh, Murshed, Alam, and Ansarin [11] also uncovered an inverted U-shaped relationship between GDP per capita and energy consumption of natural gas, liquefied petroleum, and hydropower, thereby confirming the presence of an inverse U-shaped relationship between economic growth and environmental degradation.

### 2.3 LMDI (Logarithm Mean Divisia Index) Model

The LMDI model, a decomposition analysis, is employed to differentiate the primary drivers of carbon emissions: population, GDP per capita, energy intensity, and carbon emission per unit of energy consumption [12]. Equation 3.3 illustrates the LMDI model's decomposition of these factors.

$$\mathrm{F} = \mathrm{P} * \frac{G}{P} * \frac{E}{G} * \frac{F}{E} = \mathrm{Pgef} \tag{3.3}$$

As for the LMDI decomposition technique, prior studies struggle to explore the factors affecting carbon emissions. Raupach et al. [13]. A regionalized analysis of trends in emissions, based on the Kaya identity of the LMDI technique, was conducted to explore the contributing factors of fossil-fuel combustion and industrial processes, as well as their demographic, economic, and technological drivers. Olanrewaju [14] employed the LMDI model to assess the energy consumption in South Africa's manufacturing sector, which could be more effective in enhancing energy efficiency. The results indicated that activity was the primary factor in energy consumption, thus requiring the government to focus more heavily on activity when formulating policies.

## 3 Data and Methodology

Data are collected from Statistics Bureau of Macau over the period from 1991 to 2020. Subject to the limited energy sources, the study selects petroleum gas, liquid fuel, and electricity as energy consumption commonly used in Macau.

Exploring the determinants of carbon emissions, the Kuznets curve phenomenon is examined, and decomposition analysis of factors impacting them is conducted through STIRPAT, EKC, and LMDI analysis in this study.

The estimation regression, as presented by the STIRPAT model, is as follows:

$$CO_{2,t} = \alpha_0 + \alpha_1 POP_t + \alpha_2 RGDPPW_t + \alpha_3 FD_t + \alpha_4 ISUP_t + \alpha_5 FCF_t + \varepsilon_t \tag{3.4}$$

The natural logarithm of carbon emission, denoted by CO2, is represented by POP, the natural logarithm of populations is represented by POP, RGDPPW stands for real GDP per employee, FD is credit to the private sector, which is scaled by GDP, ISUP is the ratio of value-added of the tertiary industry to value-added of the second industry, and FCF is fixed capital formation, which is scaled by GDP. In addition, ε represents error items and α1, α2, α3, α4, and α5 as proxies of elasticity coefficient.

The equation constructed by the study is to investigate the hypothesis of an environmental Kuznets curve.

$$CO_{2,t} = \alpha_0 + \alpha_1 RGDPPW_t + \alpha_2 RGDPPW_t^2 + \alpha_3 FD_t + \alpha_4 ISUP_t + \alpha_5 FCF_t + \varepsilon_t \tag{3.5}$$

Referring to LMDI model of Kaya, the study rewrite Eqs. 3.3 to Eq. 3.6–3.10, which decompose carbon emissions.:

$$\Delta C = C_t - C_{t-1} = \Delta C_p + \Delta C_g + \Delta C_e + \Delta C_f \tag{3.6}$$

$$\Delta P = \sum\nolimits_{t=1}^{3} \frac{C_{it} - C_{it-1}}{lnC_{it} - lnC_{it-1}} \ln[\left(\frac{P_t}{P_{t-1}}\right)] \tag{3.7}$$

$$\Delta g = \sum_{t=1}^{3} \frac{C_{it} - C_{it-1}}{lnC_{it} - lnC_{it-1}} \ln[\left(\frac{g_t}{g_{t-1}}\right)] \tag{3.8}$$

$$\Delta f = \sum_{t=1}^{3} \frac{C_{it} - C_{it-1}}{lnC_{it} - lnC_{it-1}} \ln[\left(\frac{f_t}{f_{t-1}}\right)] \tag{3.9}$$

$$\Delta e = \sum_{t=1}^{3} \frac{C_{it} - C_{it-1}}{lnC_{it} - lnC_{it-1}} \ln\left[\left(\frac{e_t}{e_{t-1}}\right)\right] \tag{3.10}$$

The number of workers, denoted by P, the amount of GDP, G, and energy consumption, E, are all represented by F, which is carbon emission. Additionally, G/P stands for GDP per worker, E/G is energy intensity, and F/E is carbon emission intensity.

## 4 Empirical Results

### 4.1 Description Analysis

To understand the characteristics of all variables, the maximum, minimum, mean, and standard deviation of all variables are shown in Table 1.

**Table 1.** Description Analysis

| Variables | Obser | Mean | St. Dev | Min | Max |
|---|---|---|---|---|---|
| **$CO_2$** | 30 | 7.061 | 1.543 | 4.240 | 8.704 |
| POP | 30 | 6.223 | 0.195 | 5.897 | 6.527 |
| RGDPPW | 30 | 6.151 | 0.568 | 5.371 | 7.025 |
| FD | 30 | 0.547 | 0.164 | 0.295 | 0.834 |
| ISUP | 30 | 2.416 | 0.121 | 2.236 | 2.606 |
| FCF | 30 | 0.208 | 0.078 | 0.092 | 0.366 |

### 4.2 Correlation Analysis

Table 2 shows the correlation among all variables. Any two variables have a lower correlation except that POP and RGDPPW have collinearity because they have a correlation coefficient of 0.93. Thus, the study will delete the variable of POP from estimation regression.

**Table 2.** Correlation Analysis

| Variables | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| (1) $CO_2$ | 1.000 | | | | | |
| (2) POP | −0.937*** | 1.00 | | | | |
| (3) RGDPPW | −0.952*** | 0.953*** | 1.00 | | | |
| (4) FD | −0.632 | 0.577* | 0.547 | 1.00 | | |
| (5) ISUP | 0.570** | −0.603*** | −0.702*** | 0.121** | 1.00 | |
| (6) FCF | 0.183 | −0.134 | −0.215 | 0.023 | 0.288 | 1.00 |

Note: *** indicates it is significant if $p < 0.01$

### 4.3 Unit Root Test

To avoid spurious regression, the time series of each variable should be stationary. The study employs the Unit Roots Test of Phillips-Perron to examine whether the stationary of every variable or not. Table 3 indicates that each variable is non-stationary at I (0) and stationary at I (1).

**Table 3.** Unit Root Test of Phillips-Perron

| Variable | level | | |
|---|---|---|---|
| | Adj t-Stat | 5% | results |
| | −1.876 | −1.953 | Non-stationary |
| RGDPPW | 0.561 | −1.953 | Non-stationary |
| CREDIT | 0.565 | −1.953 | Non-stationary |
| ISUP | −0.135 | −1.953 | Non-stationary |
| FCF | −0.834 | −1.953 | Non-stationary |
| Variable | The first difference | | |
| | Adj t-Stat | 5% | results |
| $CO_2$ | −4.425 | −1.953 | Stationary |
| RGDPPW | −2.027 | −1.953 | Stationary |
| CREDIT | −3.596 | −1.953 | Stationary |
| ISUP | −3.598 | −1.953 | Stationary |
| FCF | −2.667 | −1.953 | Stationary |

### 4.4 Optimal Lag Structure

Table 4 displays the different criteria for optimal lag. According to the criteria of AIC or SC, the study selects one lag as the optimal lag.

**Table 4.** Criteria of Optimal Lag

| Lag | Log L | LR | FPE | AIC | SC | HQ |
|---|---|---|---|---|---|---|
| 0 | 59.94 | NA | 1.36E−08 | −3.92 | −3.686 | −3.852 |
| 1 | 172.43 | 176.77* | 2.72e−11* | −10.17* | −8.747* | −9.738* |
| 2 | 196.89 | 29.70 | 3.41E−11 | −10.13 | −7.519 | −9.336 |

### 4.5 Johansen Cointegration

Based on the results of the Johansen cointegration test shown in Table 5, it has five cointegration relationships. According to the trace test in Table 5, the equation of the long-term Johansen cointegration relationship is as follows:

$$CO_2 = -1.61RGDPPW_t + 5.59FD_t - 2.34ISUP_t - 25.765FCF_t$$
$$(-1.459) \quad (-3.745) \quad (-5.358) \quad (-3.726).$$

where the values of parathesis are t-values.

By adding error correction item (CointEq1) in equation of vector error correction model (VECM), the study eventually has equation as follows:

$$\begin{aligned}\text{CointEq1} = &-1.612RGDPPW(-1) - 5.592FD(-1) - 2.347ISUP(-1) \\ &- 25.765FCF(-1) + 10.685.\end{aligned}$$

**Table 5.** Johansen Cointegration Test (Trace Test)

| Hypothesized | | Trace | 0.05 | |
|---|---|---|---|---|
| No. of CE (s) | Eigenvalue | Statistic | Critical Value | Prob |
| None * | 0.810 | 111.738 | 69.819 | 0.000 |
| At most 1 * | 0.585 | 66.917 | 47.856 | 0.000 |
| At most 2 * | 0.556 | 43.175 | 29.797 | 0.001 |
| At most 3 * | 0.416 | 21.286 | 15.495 | 0.006 |
| At most 4 * | 0.221 | 6.756 | 3.841 | 0.009 |

### 4.6 Analysis of Environmental Kuznets Curve

The empirical results for the environmental Kuznets curve are shown in Table 6. It indicates that it exists an inverted U-shape relationship between RGDPPW and because the β coefficient of $\text{RGDPPW}^2$ is −1.130. As RGDPPW increases, $CO_2$ will increase. After reaching the turning point of −2.260, $CO_2$ will decrease with an increase in PGDGPW.

**Table 6.** Empirical Results of Environmental Kuznets Curve

| Variable | Coefficient | St. Error | p-value |
|---|---|---|---|
| FGDPPW | 11.893 | 6.867 | 0.097 |
| $\text{RGDPPW}^2$ | −1.130 | 0.547 | 0.051 |
| FD | −0.609 | 0.184 | 0.003 |
| ISUP | 2.249 | 1.979 | 0.268 |
| FCF | −0.838 | 1.878 | 0.659 |
| Constant | −26.930 | 23.448 | 0.263 |
| AR(1) | 0.525 | 0.186 | 0.010 |
| Adj-$\text{R}^2$ | 0.954 | | |

(*continued*)

**Table 6.** (*continued*)

| Variable | Coefficient | St. Error | p-value |
|---|---|---|---|
| F-Stat | 87.558*** | | |
| D-W Stat | 1.899 | | |

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

### 4.7 LMDI Model

Referring to LMDI model developed by Kaya [12], the study constructs the equation to decompose carbon emissions as its determinants. Table 7 indicates the decomposition analysis of carbon emissions over the period 2011 to 2020.

The results show that major drivers of changes in carbon emissions by carbon consumption and carbon emission intensity affect the change in carbon emission by 75 million tons in 2015. In 2016, an increase in carbon emissions to 65 million tons resulted from carbon emissions by energy consumption. Change in carbon emission shows a negative figure during the period 2017 to 2020, which resulted from a decrease in carbon emissions by energy consumption.

**Table 7.** Decomposition Analysis in LMDI

| Year | $\Delta$ C | $\Delta$ $C_p$ | $\Delta$ $C_g$ | $\Delta$ $C_e$ | $\Delta$ $C_f$ |
|---|---|---|---|---|---|
| 2011 | 26.998 | 2.213 | 19.379 | −2.843 | 33.842 |
| 2012 | 15.839 | 3.449 | 12.406 | 7.931 | −7.947 |
| 2013 | −35.528 | 2.658 | 10.947 | −15.717 | −33.417 |
| 2014 | −5.010 | 2.316 | 3.161 | −10.609 | 0.121 |
| 2015 | 75.209 | 1.379 | −16.535 | 42.485 | 47.879 |
| 2016 | 65.867 | −0.390 | 0.234 | 61.480 | 4.542 |
| 2017 | −30.199 | 1.520 | 14.014 | −11.713 | −34.021 |
| 2018 | −153.860 | 1.445 | 6.502 | −39.369 | −122.438 |
| 2019 | −38.402 | 0.629 | −0.058 | −22.662 | −15.950 |
| 2020 | −29.475 | 0.154 | −23.461 | −4.325 | −1.843 |

## 5 Conclusion

Based on the empirical results, the study concludes that real GDP per worker, financial development, industrial structure upgrading, and fixed capital formation have long-term Johansen cointegration with carbon emissions. Among them, real GDP per worker, industrial structure upgrading, and fixed capital formation has a positive impact on

carbon emissions, while financial development negative impact on carbon emissions. In addition, it exists a phenomenon environment Kuznets curve in Macau. According to LMDI decomposition approach, energy consumption by GDP and carbon emissions by energy consumption are major drivers of carbon emissions.

Some suggestions that the Macau government should be adopted to take out:

1. Macao government should be able to introduce new energy vehicle industry, so that it can replace traditional fuel vehicles with electric vehicles or gasoline-electric hybrid vehicles.
2. Macao government should actively adopt green financial policies, such as issuing green bonds, green credit, green funds, and green insurance and so on, to guide the development of green economy.
3. Macao government can guide residents to use renewable energy products, which can reduce carbon dioxide emissions during the period of recycling (caused by burning garbage).

## References

1. Zhang, J., Dai, Y., Su, C.W., Kirikkaleli, D., Umar, M.: International change in the effect of economic growth on carbon emission in China. Energy Environ. **32**(7), 1207–1225 (2021)
2. Ehrlich, P.R., Holdren, J.P.: Impact of population growth: complacency concerning this component of main's predicament is unjustified and counterproductive. Science **171**, 1212–1217 (1971)
3. Bakirtas, I., Bayrak, S., Cetin, A.: Economic growth and carbon emission: a dynamic panel data analysis. Eur. J. Sustain. Dev. **3**(4), 91–102 (2014)
4. Narayan, P.K., Saboori, B., Soleymani, A.: Economic growth and carbon emissions. Econ. Model. **53**, 388–397 (2016)
5. Alvarado, R., Toledo, E.: Environmental degradation and economic growth: Evidence for a developing country. Environ. Dev. Sustain. **19**, 1205–1218 (2017)
6. Dietz, T., Rosa, E.A.: Rethinking the environmental impacts of population, affluence, and technology. Human Ecol. Rev. **1**, 277–300 (1994)
7. Khan, A.Q., Saleem, N., Fatima, S.T.: Financial development, income inequality, and emissions in Asian countries using STIRPAT model. Environ. Sci. Pollut. Res. **25**, 6308–6319 (2018)
8. Li, X., Du, J., Long, H.: Dynamic analysis of international green behavior from the perspective of the mapping knowledge domain. Environ. Sci. Pollut. Res. **26**(6), 6087–6098 (2019). https://doi.org/10.1007/s11356-018-04090-1
9. Grossman, G.M., Krueger, A.B.: Economic growth and the environment. Q. J. Econ. **110**(2), 353–377 (1995)
10. Shahbaz, M., Tiwari, A.K., Nasir, M.: The effects of financial development, economic growth, coal consumption, and trade openness on CO2 emissions in South Africa. Energy Policy **61**, 1452–1459 (2013)
11. Murshed, M., Alam, R., Ansarin, A.: The environmental Kuznets curve hypothesis for Bangladesh: the importance of natural gas, liquefied petroleum gas, and hydropower consumption. Environ. Sci. Pollut. Res. **28**(14), 17208–17227 (2021)

12. Kaya, Y.: Impact of Carbon Dioxide Emission Control on GNP Growth: Interpretation of Proposed Scenarios. IPCC Energy and Industry Subgroup, Response Strategies Working Group, Paris (1990)
13. Raupach, M.R., et al.: Global and regional drivers of accelerating CO2 emissions. Proc. Natl. Acad. Sci. U.S.A. **104**(24), 10288–10293 (2007)
14. Olanrewaju, O.: Multiplicative LMDI approach to South Africa's industrial energy consumption. S. Afr. J. Ind. Eng. **30**(1), 69–77 (2019)

# Literature Review of Research on Common Methods of Grapheme-To-Phoneme

Yating Zhang, Han Zhang(✉), and Shaozhong Cao

Beijing Institute of Graphic Communication, Beijing, China
{zhanghan,caoshaozhong}@bigc.edu.cn

**Abstract.** Grapheme-to-phoneme (G2P) conversion techniques have been used in many fields, most notably speech synthesis (text-to-speech, TTS). Nowadays, the development of speech synthesis is facilitated with the continuous improvement of G2P conversion techniques. The purpose of the paper is to provide an review of grapheme-to-phoneme conversion methods. First, the grapheme-to-phoneme conversion methods in recent years are sorted out; then the relevant data sets and evaluation metrics are listed, and finally the problems and development trends faced by grapheme-to-phoneme conversion are described.

**Keywords:** Grapheme-to-phoneme conversion · Speech synthesis · Machine learning · Artificial intelligence

## 1 Introduction

G2P conversion is the task of converting words from orthography (sequences of letters/characters/graphemes) to their pronunciation (sequences of phonemes) and is the front-end task of TTS, which is currently the main part of speech synthesis technology studied. A good G2P conversion method generates all possible pronunciations for each word, and inaccurate G2P conversion can lead to unnatural pronunciations or even unintelligible synthetic speech. Therefore the quality of grapheme-to-phoneme conversion plays an important role in the development of high quality speech synthesis systems.

G2P conversion algorithms developed so far can be divided into rule-based approaches, data-driven approaches and deep learning approaches. Rule-based approaches require linguistic expertise to artificially formulate language rule sets, and several rule-based approaches were proposed in [1], while [2] also showed that this approach is not well suited for languages where the connection between grapheme and phoneme is not clear, since the number of rules to convert from grapheme to phoneme in these languages grows almost linearly with the size of the lexicon. In general, although rule-based systems can produce relatively correct results, it is difficult to go for covering almost all possible cases.

This subject was supported by the Research Project of Education Commission of Beijing (KM202110015003), Initial funding for the Doctoral Program of BIGC, and Innovation team project of BIGC (Eb202103).


M. Li et al. (Eds.): ICIEIS 2022, LNOR, pp. 162–170, 2023.
https://doi.org/10.1007/978-981-99-3618-2_16

The data-driven approach is to build a G2P conversion model using probabilistic statistics and machine learning algorithms with sufficient training data, and one important step in this approach is alignment. Reference [3] proposed to achieve one-to-one alignment of graphemes and phonemes using the expectation maximization (EM) algorithm and to build a G2P conversion model using N-gram. In fact, G2P conversion can be regarded as a machine translation problem: [4] proposed the joint sequence approach when applied to G2P conversion. Reference [5] first applied the weighted finite state sensor WFST [6, 7] to the G2P conversion task when WFST is a good choice for indexing, it is effective in reducing storage space and also speeds up decoding. Reference [8] proposed EM-driven one-to-many and many-to-one alignment based on WFST, and also proposed N-best decoding algorithms for language models based on RNN [9] and minimal Bayesian risk word graph decoding algorithms. It can also be considered as a classification problem: when [10] proposed a conditional distribution model for phonemes using the maximum entropy criterion. It can be treated as a sequence labeling problem: [11, 12] introduced the Hidden Conditional Random Field (HCRF) model to model the alignment of grapheme and phoneme sequences using hidden variables. Reference [13] applied Hidden Markov Model(HMM) to G2P transformations, and [14] proposed an extension of [13]. When the case of many-to-one or one-to-many occurrences of grapheme and phoneme strings arises, [15] proposes a many-to-many alignment and proposes to find the most likely phoneme sequence for a given word using a supervised HMM method embedded in local classification. Hybrid models can produce relatively better transformation results than single models, and [16] combined a joint n-gram model with a decision tree model as well as [17] combined a joint n-gram model with a CRF model, which can improve the generalization ability of the model. The data-driven approach is a great improvement in many aspects compared to the previous methods, but requires additional alignment. In recent years, with the continuous development of artificial intelligence, people prefer to apply deep learning to G2P conversion, and the deep learning method is used to obtain better accuracy compared with previous methods.

In recent years, with the continuous development of artificial intelligence, there is a preference for applying deep learning to G2P conversion, and the deep learning approach is used to obtain better accuracy compared to previous methods.

In this paper, we focus on the technical principles and development of grapheme-to-phoneme conversion, focusing on recent grapheme-to-phoneme conversion methods; then we list relevant data sets and evaluation metrics, and finally, we summarize the current status and problems faced by grapheme-to-phoneme conversion, and make an outlook on future development.

## 2 Deep Learning Based Approach

In recent years, with the proposal of neural networks, more and more researchers have applied deep learning models to the G2P conversion problem. Deep learning trains enough sample data to capture more fine-grained and effective features. As a result, the deep learning approach is not only more robust but also has better generalization performance compared to previous methods. Nowadays most deep learning based g2p

systems are based on monolingual, but such systems are difficult to extend to low resource languages because these languages do not have data and manually developed rules, but some scholars have exploited the inherent similarity between different writing systems for G2P conversion of low resource languages by jointly learning multiple languages. Therefore, deep learning-based G2P conversion is described in two parts: monolingual and multilingual.

### 2.1 Monolingual

Attention mechanism [17] a technique that allows the model to focus its limited attention on the focused information, thus saving resources and obtaining the most effective information quickly, adding it to the deep learning model can significantly improve the model performance, so the monolingual part is divided into two parts according to whether the Attention mechanism is added or not.

#### 2.1.1 No Attention Mechanism Added

Rao et al. [19] first applied LSTM [20] to G2P transformation, which trained one-way models with different output delays since one-way long short-term memory (ULSTM) only utilizes past contexts.

Yao et al. [21] proposed to utilize an alignment-based one-layer ULSTM model using the alignment package of [15] to generate the alignment needed for training the model. The LSTM was trained using back-propagation through time (BPTT), and the hypothetical sequence with the highest posterior probability was selected as the decoding result using a beam search solver to generate phoneme sequences in the decoding stage. The one-way LSTM with default window size has a higher WER due to the failure to observe the entire input sequence.

Since the one-way LSTM requires the construction of a context window, Rao et al. [19] used a deep bidirectional LSTM (DBLSTM) to better understand contextual information using a bidirectional LSTM, increased the number of hidden layers to better linearly partition different types of data, and represented the output of the G2P of the LSTM as a finite state sensor (FST), and intersected the output with the connectionist temporal classification ( CTC) [15] after combining it with the output of the joint n-gram to obtain better model generalization. Yao et al. [21] proposed an alignment-based Bi-LSTM model in which the alignment package of [15] was used to generate the alignment needed for this model, and the alignment was explicitly modeled by the context of the grapheme during the G2P transformation to achieve better model results.

Mousa et al. [23] proposed the use of complex many-to-many alignment to improve the performance of BLSTM RNN-based G2P models. The method considers various alignment schemes through hyperparameter optimization (including the number of hidden layers, optional linear projection layers, and optional input splicing windows).

Inspired by the translation model of [24], Yao et al. [21] proposed to use an encoder-decoder structure for G2P conversion, using an LSTM as the basic recursive network unit. In this model, since the intermediate vector length is fixed regardless of the input and output sequence lengths, display alignment between grapheme and phoneme sequences

is not required. The BPTT [25, 26] was also used to train the encoder-decoder, and the decoding was performed using a beam search solver.

Since LSTMs read inputs sequentially and the output of further inputs depends on the previous inputs, LSTMs cannot parallelize these networks and applying CNNs can reduce the computational load by using large receptive domains. Yolchuyeva et al. [27] first applied convolutional neural networks (CNNs) to G2P transformations, proposing a model that uses CNNs as encoders and Bi- LSTM as a decoder model. The CNN layer in this method performs convolutional operations with grapheme as input and introduces residual connections to reduce the vanishing gradient problem and improve the model performance while increasing the number of network layers.

Chae et al. [28] studied the non-sequential greedy decoding (NSGD) method for G2P and combined it with a fully convolutional encoder-decoder structure. Non-sequential decoding of G2P transformations was proposed using NSGD, which iteratively infers the most probable part of the candidate positions that have not yet been inferred.

### 2.1.2 Add Attention Mechanism

Toshniwal et al. [29] proposed an encoder-decoder model supporting an attention mechanism [30] applied to G2P conversion, in which a global attention mechanism is used to make it consider all encoder states when predicting each decoder output, giving the encoder-decoder network the ability to consider "soft" alignments and learns these alignments together with the sequence prediction task, eliminating the previous dependence on display alignments and thus improving model performance.

Also [29] proposed the use of a local attention mechanism-monotonic alignment (local-m) for application to G2P transformations, which finds an aligned position on each time slot while considering only the encoder states at the aligned position that lie within a fixed-length context window. Both this model and the global attention-based model are good choices when the grapheme sequence length is short.

When deploying G2P models in production systems, how to utilize untagged data is crucial to improve performance. Sun et al. [31] first applied Transformer networks [32] to G2P transformations and proposed a token-level integration distillation method based on Transformer models, which uses knowledge distillation to extract knowledge from additional untagged words as well as from integrated models. Knowledge distillation provides a way to use untagged source data. The teacher model can generate target phoneme sequences based on untagged source grapheme sequences, and the resulting phoneme sequences can be used as labels for the student model, thus improving the accuracy of G2P conversion of untagged words. Sequence-to-sequence based G2P models are trained with different model structures for integration, including Transformer, Bi-LSTM [33] and CNN-based sequence-to-sequence models [34], and then the knowledge from the integrated models is extracted into the Transformer model to reduce the model size for online deployment.

Transformer-based models are able to be trained in parallel and have significantly reduced learning time, which has received more attention from scholars. Yolchuyeva et al. [35] applied Transformer networks to G2P transformations to improve the performance of the models by increasing the number of layers of encoder-decoder in comparison.

Řezáčková et al. [36] proposed to convert input text sentences into phoneme sequences with high accuracy using a text-to-text transfer transformer (T5) neural network model. The T5 model in this approach is trained in a semi-supervised manner from a large text corpus to a full encoder-decoder Transformer.

Dong et al. [37] proposed a pre-trained grapheme model BERT (GBERT) to improve the Transformer-based G2P model. The model was built by self-supervised training on a large, language-specific list of words. The pre-training task of GBERT is a masked grapheme prediction task, as the pre-trained GBERT-derived grapheme representations do not provide all the necessary information for G2P transformation, making it slightly worse than the medium- and low-resource G2P-based models in both medium- and low-resource G2P tasks. The model performs slightly worse than the Transformer-based model in medium and low resource G2P tasks.

GBERT is fine-tuned in [37] by replacing the encoder in Transformer with GBERT, trained in an end-to-end manner, during which the pre-trained GBERT encoder and the randomly initialized Transformer decoder use different learning rates, making it effective for most languages under low-resource conditions.

Inspired by the BERT fusion model [37, 38] fused GBERT into the Transformer-based G2P model. Specifically, an additional GBERT-Enc attention module was added to each encoder layer to adaptively control how that layer interacts with the GBERT representation, with each decoder layer operating similarly. One of the GBERT attention modules employs GBERT as a feature extractor, not only using the GBERT representation but also utilizing the raw input from the transformer-based G2P model. Since this model has a more complex model structure and is more sensitive to the amount of data trained, it can reduce the error rate of G2P under moderate resource conditions.

In general, deep learning eliminates the dependence on display alignment compared to data-driven methods, which can capture more fine-grained and effective features and better model performance. Since the Attention mechanism can solve the problem that RNN cannot be computed in parallel, and the focus can be captured from the middle when the data is long without losing important information, the introduction of the Attention mechanism can not only reduce the model parameters, which can speed up the model computation, but also, more importantly, improve the model performance.

### 2.2 Multilingual

Peters et al. [39] proposed a neural sequence-to-sequence approach to train spelling-pronunciation pairs for hundreds of languages, which shares a single encoder and decoder across all languages, allowing it to exploit the inherent similarities between different writing systems. Sokolov et al. [40] proposed a single end-to-end trained neural G2P model that spans multiple languages sharing the same encoder and decoder, exploiting the large amount of multilingual data to improve prediction accuracy, especially for low-resource languages.

Yu et al. [41] proposed a Transformer-based multilingual G2P model with byte-level input representation, in which each character is represented at the byte level based on UTF-8 encoding, reducing the size of the input vocabulary and further supporting the merging of multiple languages to accommodate different grapheme systems. Zhu et al.

[42] proposed a ByT5-based G2P model to Kim et al. [43] proposed a fast and high-performance bilingual G2P model to solve the large-scale multilingual G2P conversion problem. In this method, for fast and accurate decoding, a non-autoregressive structured Transformer conditional random field based architecture (NART-CRF) is used with byte representation of the input and data augmentation techniques for accurate output length inference.

Overall, learning multiple languages jointly before obtaining low-resource phoneme results may work better than monolingual G2P models using datasets.

## 3 Data Set and Evaluation Indicators

### 3.1 Dataset

The main G2P conversion datasets are CMUdict [44] (1993), Pronlex Pronunciation Dictionary [45] (1995), CELEX Pronunciation Dictionary [46] (1993), and NetTalk Manual Alignment Dataset [47] (1986).

The CMUdict pronunciation dictionary was developed for the ASR and contains over 134,000 word forms and their pronunciations. The latest version of the Pronlex pronunciation dictionary contains pronunciations for 90,988 word forms, covering words used in the Wall Street Journal corpus and the Switchboard corpus over the years. The CELEX pronunciation dictionary includes the Oxford Advanced Learner's English The NetTalk dataset was originally developed to train the NetTalk neural network and contains 19,802 words, some with multiple pronunciations, resulting in a total of 20,008 entries. Total of 20,008.

The three pronunciation dictionaries CMUdict, Pronlex, and CELEX all represent stress as three levels: primary stress, secondary stress, and no stress. The CELEX pronunciation dictionary is for British English pronunciation, and the other three pronunciation dictionaries are for American English pronunciation. Each pronunciation dictionary uses a different phone set, CMU and PRONLEX's subsets are introduced from ARPAbet, and CELEX's subsets are introduced from IPA.

### 3.2 Evaluation Indicators

As performance metrics, we use the phoneme error rate (PER) and word error rate (WER) to evaluate the quality of the G2P model on text data.

PER is used to measure the distance between the predicted phoneme sequence and the reference pronunciation divided by the number of phonemes in the reference pronunciation. The edit distance (also known as Levenshtein distance [48]) is the minimum value of insertion (I), deletion (D) and substitution (S) required to convert one sequence to another and can be calculated by dynamic programming methods [49]. If there are multiple pronunciation variants of a word in the referenced data, the variant with the smallest Levenshtein distance from the candidate word is used. For the calculation of WER, the calculation is performed only if the predicted pronunciation does not match any of the reference pronunciations, and the number of word errors is divided by the total number of words. Lower WER and PER represent better model performance. Table 1 shows the comparison of the results of the G2P transformation algorithm on the CMUDict dataset.

**Table 1.** Comparison of results of G2P transformation algorithm on CMUDict dataset

| Algorithm | Evaluation Indicators | |
|---|---|---|
| | PER(%) | WER(%) |
| Encoder-decoder LSTM [21] | 7.63 | 28.61 |
| Joint sequence model [4] | 5.88 | 24.53 |
| Deep Bi-LSTM with many-to-many alignment [23] | 5.37 | 23.23 |
| Joint maximum entropy (ME) n-gram model [3] | 5.9 | 24.7 |
| Encoder CNN, decoder Bi-LSTM [27] | 4.81 | 25.13 |
| End-to-end CNN [27] | 5.84 | 29.74 |
| Encoder-decoder LSTM with attention [27] | 5.68 | 28.44 |
| Transformer 4 × 4 [35] | 5.23 | 22.1 |

## 4 Summary and Outlook

This paper summarizes the progress of common research methods of G2P conversion, which has evolved from rule-based methods to deep learning methods. G2P conversion technology has made some progress in model simplification, but there are still difficult problems to be improved and solved in terms of reducing training cost, improving training speed, and optimizing generation quality. Meanwhile, most of the currently used public datasets are related to medium-resource languages, and there are not enough datasets available for G2P conversion of low-resource languages. As language synthesis technology is applied to more fields, G2P conversion has a broader research prospect as an important basis for improving the quality of speech synthesis.

## References

1. Kaplan, R.M., Kay, M.: Regular models of phonological rule systems. Comput. Linguist. **20**(3), 331–378 (1994)
2. Kominek, J., Black, A.W.: Learning pronunciation dictionaries: language complexity and word selection strategies. In: Proceedings of the Human Language Technology Conference of the NAACL, Main Conference, pp. 232–239 (2006)
3. Galescu, L., Allen, J.F.: Bi-directional conversion between graphemes and phonemes using a joint n-gram model. In: 4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis (2001)
4. Bisani, M., Ney, H.: Joint-sequence models for grapheme-to-phoneme conversion. Speech Commun. **50**(5), 434–451 (2008)
5. Yang, D., Dixon, P., Sadahi, F.: Rapid development of a grapheme-to-phoneme system based on weighted finite state transducer (WFST) framework. In: The Japanese Acoustical Society 2009 Fall Lecture Proceedings, no. 3, pp. 111–112 (2009)
6. Novak, J.R., Dixon, P.R., Minematsu, N., et al.: Improving WFST-based G2P conversion with alignment constraints and RNNLM N-best rescoring. In: Thirteenth Annual Conference of the International Speech Communication Association (2012)

7. Novak, J.R., Minematsu, N., Hirose, K.: Failure transitions for joint n-gram models and G2p conversion. In: Interspeech, pp. 1821–1825 (2013)
8. Novak, J.R., Minematsu, N., Hirose, K.: WFST-based grapheme-to-phoneme conversion: open source tools for alignment, model-building and decoding. In: Proceedings of the 10th International Workshop on Finite State Methods and Natural Language Processing, pp. 45–49 (2012)
9. Mikolov, T., Karafiát, M., Burget, L., et al.: Recurrent neural network based language model. In: Interspeech, vol. 2, no. 3, pp. 1045–1048 (2010)
10. Chen, S.F.: Conditional and joint models for grapheme-to-phoneme conversion. In: INTERSPEECH (2003)
11. Lehnen, P., Hahn, S., Guta, V.A., et al.: Hidden conditional random fields with m-to-n alignments for grapheme-to-phoneme conversion. In: Thirteenth Annual Conference of the International Speech Communication Association (2012)
12. Lehnen, P., Allauzen, A., Lavergne, T., et al.: Structure learning in hidden conditional random fields for grapheme-to-phoneme conversion. In: Annual Conference of the International Speech Communication Association (2013)
13. Taylor, P.: Hidden Markov models for grapheme to phoneme conversion. In: Ninth European Conference on Speech Communication and Technology (2005)
14. Ogbureke, K.U., Cahill, P., Carson-Berndsen, J.: Hidden Markov models with context-sensitive observations for grapheme-to-phoneme conversion. In: Eleventh Annual Conference of the International Speech Communication Association (2010)
15. Jiampojamarn, S., Kondrak, G., Sherif, T.: Applying many-to-many alignments and hidden Markov models to letter-to-phoneme conversion. In: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, pp. 372–379 (2007)
16. Hahn, S., Vozila, P., Bisani, M.: Comparison of grapheme-to-phoneme methods on large pronunciation dictionaries and LVCSR tasks. In: Thirteenth Annual Conference of the International Speech Communication Association (2012)
17. Wu, K., Allauzen, C., Hall, K., et al.: Encoding linear models as weighted finite-state transducers (2014)
18. Mnih, V., Heess, N., Graves, A.: Recurrent models of visual attention. In: Advances in Neural Information Processing Systems, vol. 27 (2014)
19. Rao, K., Peng, F., Sak, H., et al.: Grapheme-to-phoneme conversion using long short-term memory recurrent neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4225–4229. IEEE (2015)
20. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
21. Yao, K., Zweig, G.: Sequence-to-sequence neural net models for grapheme-to-phoneme conversion. arXiv preprint arXiv:1506.00196 (2015)
22. Graves, A., Fernández, S., Gomez, F., et al.: Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 369–376 (2006)
23. Mousa, A.E.D., Schuller, B.: Deep bidirectional long short-term memory recurrent neural networks for grapheme-to-phoneme conversion utilizing complex many-to-many alignments (2016)
24. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems, vol. 27 (2014)
25. Robinson, A.J., Fallside, F.: The Utility Driven Dynamic Error Propagation Network. University of Cambridge Department of Engineering, Cambridge (1987)
26. Williams, R.J., Peng, J.: An efficient gradient-based algorithm for on-line training of recurrent network trajectories. Neural Comput. **2**(4), 490–501 (1990)

27. Yolchuyeva, S., Németh, G., Gyires-Tóth, B.: Grapheme-to-phoneme conversion with convolutional neural networks. Appl. Sci. **9**(6), 1143 (2019)
28. Chae, M.J., Park, K., Bang, J., et al.: Convolutional sequence to sequence model with non-sequential greedy decoding for grapheme to phoneme conversion. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2486–2490. IEEE (2018)
29. Toshniwal, S., Livescu, K.: Jointly learning to align and convert graphemes to phonemes with neural attention models. In: 2016 IEEE Spoken Language Technology Workshop (SLT), pp. 76–82. IEEE (2016)
30. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
31. Sun, H., Tan, X., Gan, J.W., et al.: Token-level ensemble distillation for grapheme-to-phoneme conversion. arXiv preprint arXiv:1904.03446 (2019)
32. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
33. Wu, Y., Schuster, M., Chen, Z., et al.: Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144 (2016)
34. Gehring, J., Auli, M., Grangier, D., et al.: Convolutional sequence to sequence learning. In: International Conference on Machine Learning, pp. 1243–1252. PMLR (2017)
35. Yolchuyeva, S., Németh, G., Gyires-Tóth, B.: Transformer based grapheme-to-phoneme conversion. arXiv preprint arXiv:2004.06338 (2020)
36. Řezáčková, M., Švec, J., Tihelka, D.: T5g2p: using text-to-text transfer transformer for grapheme-to-phoneme conversion (2021)
37. Dong, L., Guo, Z.Q., Tan, C.H., et al.: Neural grapheme-to-phoneme conversion with pre-trained grapheme models. In: ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6202–6206. IEEE (2022)
38. Zhu, J., Xia, Y., Wu, L., et al.: Incorporating bert into neural machine translation. arXiv preprint arXiv:2002.06823 (2020)
39. Peters, B., Dehdari, J., van Genabith, J.: Massively multilingual neural grapheme-to-phoneme conversion. arXiv preprint arXiv:1708.01464 (2017)
40. Sokolov, A., Rohlin, T., Rastrow, A.: Neural machine translation for multilingual grapheme-to-phoneme conversion. arXiv preprint arXiv:2006.14194 (2020)
41. Yu, M., Nguyen, H.D., Sokolov, A., et al.: Multilingual grapheme-to-phoneme conversion with byte representation. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8234–8238. IEEE (2020)
42. Zhu, J., Zhang, C., Jurgens, D.: ByT5 model for massively multilingual grapheme-to-phoneme conversion. arXiv preprint arXiv:2204.03067 (2022)
43. Kim, H.Y., Kim, J.H., Kim, J.M.: Fast bilingual grapheme-to-phoneme conversion. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 289–296. Industry Track (2022)
44. The CMU Pronouncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict
45. CALLHOME American English Lexicon (PRONLEX). Linguistic Data Consortium
46. CELEX lexical database. http://www.kun.nl/celex/
47. Sejnowski, T.J.: The NetTalk Corpus: Phonetic Transcription of 20008 English Words (1988)
48. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions, and reversals. Soviet physics doklady. **10**(8), 707–710 (1966)
49. Zhu, R., Huang, Y.: Efficient privacy-preserving general edit distance and beyond. Cryptology ePrint Archive (2017)

# Author Index





M. Li et al. (Eds.): ICIEIS 2022, LNOR, p. 171, 2023.
https://doi.org/10.1007/978-981-99-3618-2