Interpretable Machine Learning Based on Integration of NLP and Psychology in Peer-to-Peer Lending Risk Evaluation

Lei Li[1(✉)], Tianyuan Zhao[1(✉)], Yang Xie[1], and Yanjie Feng[2]

[1] Beijing University of Posts and Telecommunications, Beijing 100876, China
leili@bupt.edu.cn, zhaotianyuan13@163.com
[2] Shanghai University of International Business and Economics, Shanghai 201620, China

Abstract. With the rapid development of Peer-to-Peer (P2P) lending in the financial field, abundant data of lending agencies have appeared. P2P agencies also have problems such as absconded with ill-gotten gains and out of business. Therefore, it is urgent to use the interpretable AI in Fintech to evaluate the lending risk effectively. In this paper we use the machine learning and deep learning method to model and analyze the unstructured natural language text of P2P agencies, and we propose an interpretable machine learning method to evaluate the fraud risk of P2P agencies, which enhances the credibility of the AI model. First, this paper explains model behavior based on the psychological interpersonal fraud theory in the field of social science. At the same time, the NLP and influence function in the field of natural science are used to verify that the machine learning model really learns the information of part-of-speech details in the fraud theory, which provides the psychological interpretable support for the model of P2P risk evaluation. In addition, we propose "style vectors" to describe the overall differences between text styles of P2P agencies and understand model behavior. Experiments show that using style vectors and influence functions to describe text style differences is the same as human intuitive perception. This proves that the machine learning model indeed learn the text style difference and use it for risk evaluation, which further shows that the model has a certain machine learning interpretability.

Keywords: Interpretable machine learning · Natural Language Processing (NLP) · Fraud theory in psychology · AI in Fintech · Peer-to-Peer (P2P) lending risk evaluation

1 Introduction

P2P lending is a kind of private lending model that gathers small amounts of money to lend to people in need of funds. The main process is to use the Internet credit company as an intermediary platform to provide information release and transactions through the Internet. China's P2P lending company have developed rapidly due to its advantages of convenience, high interest rate. However, there are also many problems, such as absconded with ill-gotten gains and difficult withdrawing. At present, the risk assessment

X. Zhu et al. (Eds.): NLPCC 2020, LNAI 12431, pp. 429–441, 2020.
https://doi.org/10.1007/978-3-030-60457-8_35

of P2P lending agencies is still very scarce. In particular, P2P network lending has generated a lot of data, especially unstructured natural language text, which contain more plentiful information than structured one. These data can be used to effectively evaluate the risk fraud in the lending process, and then analyze its interpretability according to the effect of different machine learning models, which is significant to verify whether the behavior of the model conforms to human cognition, and can enhance the user's understanding and trust of the system. These are significant to reduce the risk of online lending, strengthen market supervision, assist in making policies and decisions, and establish a good financial investment environment.

In the current information age, data processing and analysis is very important. Admittedly, machine learning and deep learning can automatically process and analyze a large number of data, but the interpretability of their models does not have sufficient theoretical support. Only interpretable models can be applied to the market more safely, which is necessary in financial industry that requires a high degree of accuracy and stability. The current research of machine learning interpretability is mostly based on the methods in the field of natural science, which explain the model behavior through the analysis of model structure and data, but rarely analyze whether the model conforms to human cognitive behavior in the field of social science.

In summary, this paper uses machine learning method to model and analyze unstructured natural language text information in P2P lending companies, and to evaluate the potential fraud risks of various companies from the perspective of machine learning interpretability. For the first time, this paper proposes a psychological fraud theory based on social sciences to explain the results of P2P lending model risk assessment. At the same time, we use the influence function and computational linguistics technology in the field of natural sciences to verify that the machine learning model really learns the important information in the fraud theory, which provides the interpretable support for the machine learning model of risk fraud. The main contributions include two points:

(1) For the explicit details features such as part-of-speech distribution, we combine psychology and computational linguistics to propose and verify an interpretable machine learning in P2P lending risk evaluation.
(2) For the implicit abstract features such as doc2vec, we propose a machine learning interpretability research based on text style. First, we define a style vector, and then combine the influence function and least square method to describe the overall difference of P2P company text, and use it in risk evaluation.

2 Related Work

The existing P2P risk assessment is mainly based on the theory of economics and personal credit risk, using the method of combining theoretical research with case analysis. For example, the credit risk assessment of P2P lending investment decision based on examples [1] and the enhancement of P2P lending investment decision [2]. However, there is still less work to identify the operational risk of P2P lending from the perspective of fraud theory. In previous work, we proposed a data-driven risk assessment framework [3] for 4554 unstructured natural language text data of P2P companies, and NLP technologies such as keywords [4], LDA [5], word2vec [6] and doc2vec [7] were used to

extract the features of the text for each P2P company, and then meta-learning method were used to integrate multiple machine learning and deep learning models. Experiments showed that the precision of risk identification using text-based features such as company profile and executive profile was higher than that of numerical features (volume, yield, etc.). Textual features include not only explicit features such as part-of-speech distributions, but also implicit features such as doc2vec. In this paper, in order to further analyze the interpretability of the evaluation results, we build on that work and propose an interpretable machine learning method based on the combination of psychology and computational linguistics for the explicit features, and propose text style-based machine learning interpretability research for implicit features.

Although researchers are eager to explore the interpretable truth from the performance of machine learning model, there is little consensus on the specific definition and evaluation method of machine learning interpretability [8], and even less research on the interpretable machine learning of P2P lending risk fraud. At present, there are three kinds of interpretable evaluation methods in general. The first type is ante-hoc interpretability: the model itself is interpretable due to its simple structure and easy to understand, such as decision tree [9], generalized linear model [10], etc. The second type is post-hoc interpretability: for the trained model, the relationship between the input and output of the sample is analyzed by using the interpretable method to explain the working mechanism and operation principle of the model, such as the influence function [11] and LIME [12]. The third type is based on the multi-disciplinary point of view: through philosophy, psychology and other theories to explain the model of human cognition. What makes psychology stand out is that many theories of psychology have been proved and verified in a large number of psychological experiments (such as cognitive psychology [13] and experimental psychology [14]).

Koh PW et al. [11] use the influence function to track the prediction results of the model and trace them to the training samples through the learning algorithm, so as to obtain the training points with the greatest influence on the prediction results. Compared with influence function, other interpretable algorithms (such as decision tree) simply analyze the relationship between model input and output from the perspective of feature itself or the principle of easily interpretable model, while influence function is strictly defined and proved by reasoning in the paradigm of machine learning. The whole process is very consistent with the research process of machine learning. This is why the influence function is chosen as the focus of this paper to study the interpretability of machine learning.

In psychology, Criteria-Based Content Analysis (CBCA) [15] is usually used to identify cases and adult lies. The CBCA theory is able to distinguish between lies and the truth because the person who is actually experiencing the event gives a more detailed description, and therefore meets more CBCA criteria. Interpersonal Deception Theory (IDT) is also used to explain, predict and identify lying behaviors in interpersonal situations. According to the theory, liars use the following strategies to control the information in a conversation in order to avoid getting caught: (a) Quality Manipulations: liars will deviate completely or partially from the facts and will use fewer adjectives and adverbs to make the meaning of the sentence ambiguous; (b) Quantity Manipulations: liars use fewer words and sentences and cannot provide rich details.

In summary, the previous researches mainly focused on the analysis of numerical and textual information, involving a variety of machine learning and deep learning models. However, the current research rarely analyzes the risk fraud evaluation results of P2P companies from the perspective of machine learning interpretability. Also, the research on interpretability is only the analysis of the model mechanism in natural science, and rarely can be given from the psychological theory of social science. There is also rare research of the efficacy of implicit features used in machine learning models from a textual perspective. These have caused great obstacles to the application of artificial intelligence in the financial field.

3 Model

3.1 Machine Learning Based on Integration of Psychology and NLP

(1) Text Details of CBCA and IDT in Psychology.

CBCA theory points out that fraud can be identified by identifying "general description" and "detailed description". The CBCA theory states that people who actually experience the event will make a more detailed description and therefore will meet more CBCA standards. The IDT theorizes that at least four types of cues are involved in lie detection as shown in Table 1: the number of words, the use of pronouns, the emotional vocabulary, and the cognitive complexity of the presenter.

Table 1. CBCA&IDT detailed description.

Detailed feature	Content
Number of words	Distribution of part-of-speech
Number of details	Specific place, time, person, etc.
Unusual details	Unusual but meaningful details of people, objects and events
Redundant details	Peripheral information with no actual contribution to the statement

(2) Extract Part-of-speech Details Based on NLP.

According to Table 1, for company profile text, we can get part-of-speech sequence and frequency of each part-of-speech at the same time after tagging. The 42 part-of-speech include: adjectives, adverbs, nouns, adjective morphemes, distinguishing words, conjunctions, adverbs, interjections, prefixes, orientation words, idioms, abbreviations, suffixes, idioms, numerals, noun, nominal morpheme, person name, place name, organization, other proper names, onomatopoeia, preposition, quantifier, pronoun, personal pronoun, demonstrative pronoun, place word, time word, time word morpheme, auxiliary word, auxiliary morpheme, auxiliary word, idioms, verbs, adverbs, nominal verbs, intransitive verbs, verbal morphemes, mood words, state words, state morphemes.

(3) Use Influence Functions to Verify the Importance of each Part-of-speech.

The influence function algorithm [11] can observe the change of model parameters by increasing the weight of training samples or disturbing the training samples. The prediction results of the model can be traced back to the training samples, so as to obtain the training data with the greatest influence on the prediction results, and then further analyze the influence degree of each feature on the final results of the model. The influence function (I) of a single training sample (z) on all model parameters (θ) is as follows:

$$I_{up,\ params}(z) \stackrel{\text{def}}{=} \frac{d\hat{\theta}_{\epsilon,z}}{dx}|_{\epsilon=0} = -\mathrm{H}_{\hat{\theta}}^{-1}\nabla_{\theta}L\left(z,\hat{\theta}\right) \tag{1}$$

$$H_{\theta} \stackrel{\text{def}}{=} \frac{1}{n}\sum\nolimits_{i=1}^{n}\nabla_{\theta}^{2}L\left(z_i,\hat{\theta}\right) \tag{2}$$

Where ε is the weight of sample z relative to other training samples, H_θ is the Hessian second-order partial derivative matrix, including the influence of all N training samples on the model parameter θ. The gradient $\nabla_{\theta}L\left(z,\hat{\theta}\right)$ includes the influence of a single training sample z on model parameter θ, where L is loss of training samples.

In the experiment, we use part-of-speech features to train a variety of machine learning models, obtain the importance of each part-of-speech feature through the influence function, verify whether the detailed information extracted by the machine learning model is the same as the detailed information believed by psychological theory, and explain the behavior of the machine learning model from a psychological perspective.

3.2 Interpretable Machine Learning Based on Text Style

For executive profile texts, there are large differences in text styles between normal and abnormal companies. We believe that it is the machine learning model that captures such differences in text styles that has a high accuracy rate. Therefore, this paper proposes to study the interpretability of machine learning based on text styles.

(1) Detailed Description.

There are obvious differences in the text style between the normal and abnormal executive profile texts in the two categories. Table 2 lists the detailed analysis. It can be seen that the distribution of language structural units provides an important basis for the analysis of text style. Through the statistics of language structure features in different texts, we can get the consistency or difference features of text style. The distribution data of language structure, such as part-of-speech, becomes a kind of measurement feature reflecting different types of language style. After the NLP method is used to extract 42 part-of-speech features, we also use the 200 dimensional doc2vec for feature representation. Because the part-of-speech distribution features are not comprehensive. Some important text style information, such as context semantic relation, will be implicitly reflected in the doc2vec in some way. Therefore, it is necessary to analyze the interpretability of doc2vec carefully.

Table 2. Text style of executive profile.

Business status	Text style of executive profile
Normal	With a bright resume and detailed introduction, they graduated from a well-known university, mostly with a master's degree or above. The career experience is complete, and the work content of each stage has a relatively specific description, and also has important positions in well-known companies
Abnormal	Short length, low education, working in a small or unknown company, work content is not detailed

(2) The interpretability of text style based on style vector and influence function.

Most of the previous studies use doc2vec as a feature representation method directly. Although the precision of text classification using doc2vec is high, and the semantic, grammatical and emotional information of context can be well combined with words, it still ignores the important role of the rich part of speech details of text. Therefore, this paper proposes "style vector" to describe style differences.

For the positive and negative executive profiles, we use doc2vec method to get their 200 dimensional doc2vec features. First, we make a difference between the values of each dimension and take the absolute value. For the 42 dimensional part-of-speech features of the two categories, the average value of each dimension is calculated, and the new 42-dimensional part-of-speech features are concatenated behind the new vector to obtain a new 242-dimensional style vector. Secondly, we use the influence function to get the influence coefficient of each dimension feature. Thus, 242 points are obtained (xi, yi) (i = 1, 2,… 242), where x is the influence coefficient and Y is the value of the style vector. Then the least square method is used to fit the line equation of 242 points. Finally, the slope of the line equation is used to describe the style difference, so as to enhance the interpretability of the model. The principle of the straight line fitting based on the least square method is as follows: if the regression straight line equation is, its slope and intercept can be obtained according to formula (3) and formula (4).

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{3}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \tag{4}$$

4 Experiments and Analysis

4.1 Data Source

We have collected data about 4,554 P2P companies from a third-party platform of network lending, the Home of Network Loan (https://www.wdzj.com). There are four categories graded from 0 to 3, and different categories represent different business status.

Table 3. The status and numbers of the four categories.

Label	Business status	Number
0	Normal	1849
1	Out of business	1263
2	Difficult withdrawing	595
3	Absconded with ill-gotten gains	847

The companies with label 1–3 are all abnormal companies. The significance of all kinds of data is shown in Table 3.

The company profile mainly includes business content, scope of business, operation philosophy, social responsibility, etc. That information can fully describe a company, and the description information of different types of companies varies greatly, which plays a significant role in the subsequent risk category assessment.

4.2 Interpretable Machine Learning Based on Integration of Psychology and NLP

(1) Detailed description.

By counting the frequency of part-of-speech, the company profile text has the following rules: the quantitative relationship of each part-of-speech is noun > adjective > pronoun > adverb > preposition, and the number of these part- of-speech is much higher than other unimportant part-of-speech. And the number of the representative part-of-speech of the company from normal state to Absconded with ill-gotten gains is decreasing, the specific data is shown in the Table 4. By analyzing the text of the company profile, it is found that the company profile that operates normally contains a larger number of representative part-of-speech such as nouns, adjectives, as well as places, organizations, etc., which are rich details in CBCA and IDT theory. Unusual details: some companies operating abnormally include a large number of words such as exposure, rights protection and resolution. Redundant details: the company profile of abnormal company has too much space to publicize the corporate culture.

Table 4. Frequency of representative part-of-speech in some companies.

Company	Label	Noun	Adjective	Preposition	Adverb	Pronoun	Place
Xin**	0	405	62	46	28	25	7
Ren**	0	90	11	13	11	4	2
Qian**	1	33	3	5	1	0	2
Hua*	1	17	1	1	2	0	1
Shuo*	2	14	0	1	0	0	1
Jun***	2	19	0	2	5	3	2
Tai***	3	10	1	1	0	1	1
Shi**	3	6	0	0	1	0	1

In this case, this paper analyzes the representative part-of-speech of each company to verify that these abnormal companies do exist fraud. It not only conforms to the behavior of fraud in psychological fraud theory, but also provides psychological explanation and support for many machine learning and deep learning models.

(2) Importance of Part-of-speech Features.

For normal and abnormal companies, it is still relatively simple to quantify the part-of-speech distribution only from a statistical perspective. Therefore, it is necessary to combine the influence function to analyze the specific importance of each part-of-speech feature. We call the importance of each part-of-speech feature based on the influence function as the "influence coefficient". The larger the influence coefficient, the more important the feature is.

We used multiple models such as Logistic Regression, SVM [16], CNN [17], LSTM [18]. Among them, the results of the Logistic Regression model are the best. At the same time, the decision tree and the LIME model are used as the comparative experiments of the influence function, and the specific results are shown in Table 5.

Table 5. Importance of part-of-speech features.

Part-of-speech	Influence coefficient	Decision tree	LIME
Suffix	1.21	0.33	0.43
All nouns	0.65	0.24	0.21
Conjunction	0.6	0.09	0.11
All adjectives	0.51	0.11	0.14
Idioms	0.48	0.04	0.06
All pronouns	0.38	0.09	0.08
Quantifier	0.27	0.05	0.09
Idioms and allusions	0.13	0.03	0.06
Preposition	0.11	0.03	0.02
Interjection	0.09	0.05	0.04
Abbreviations	0.09	0.01	0.01
Adverb	0.07	0.03	0.02
Numeral	0.06	0.01	0.01
Prefix	0.06	0.05	0.03

Through the experiment of part-of-speech features based on the influence function, it is found that: nouns, adjectives, pronouns are of high importance, and suffixes, conjunctions, idioms, quantifiers, prepositions and other part-of-speech are of high importance, which fully coincides with the part-of-speech details in CBCA and IDT theories. Companies with good credit usually contain more important details, while companies with fraudulent intent have less quantity and quality of details. The location words and organization groups are also important, which also correspond to the location and organization details in the fraud theory. At the same time, the importance degree of the part-of-speech features based on the influence function is more obvious in the numerical difference, and the importance degree of each part-of-speech is very clear, which shows that the influence function algorithm has a high interpretability.

To sum up, for the three interpretable algorithms of influence function, decision tree and LIME, the influence function is better than the other two algorithms in both the ability to interpret the model and the applicable scope. Meanwhile, influence function also verifies that it is more appropriate to interpret fraud theory in psychology.

(3) Experimental Results.

Psychological fraud theory believes that texts rich in details have stronger authenticity, and reflected in natural language processing are more abundant in terms of nouns and adjectives. In this experiment, the importance of each part-of-speech feature is obtained through influence function, decision tree and LIME. Also, it is proved that part-of-speech features, such as nouns, adjectives and pronouns, which represent details

are indeed very important. Moreover, the precision of classifier based on part-of-speech features can reach 80%. It shows that the details extracted by NLP that can be considered important by machine learning exactly coincide with details that are considered important in psychology, and it is proved by influence function that these important details make the precision of machine learning model higher.

Therefore, this section explains the model behavior based on the psychological fraud theory in the field of social science, and verifies that the machine learning model really learns the important information in the fraud theory by using the NLP technology in the field of natural science, thus providing the interpretable support for the machine learning model of risk fraud.

4.3 The Interpretability of Text Style Based on Style Vector and Influence Function

According to the previous experimental results, the normal company's executive profiles are usually very detailed, complete and convincing, while the abnormal company's executive profiles are usually not detailed. According to the theories of CBCA and IDT related to psychological fraud theory, we believe that companies with low quality executive profiles are more likely to have fraud intentions, and this intuitive difference in text style also provides with a new method of interpretable research.

(1) Interpretable Research Based on Doc2vec.

Most of the previous studies use doc2vec as a feature representation method directly. Although the accuracy of using doc2vec for text classification is high, there is a problem that the specific meaning of each dimension of dec2vec and the relationship between each dimension can not be explained. Therefore, this experiment focuses on the analysis of the interpretability of doc2vec.

For normal and abnormal executive profile texts, firstly, 200 dimension doc2vec are used to represent the features of all positive and negative executive profiles, and then average all positive and negative executive profiles to form a new 200 dimension doc2vec (for example, the blue curve in the Fig. 1 is a normal company). Similarly, all negative executive profiles also obtain a new 200 dimension doc2vec (for example, the orange curve in the Fig. 1 Abnormal company). Then, the influence function is used to analyze the doc2vec itself, and two rules are found, which can understand and explain the behavior of the model: (1) there are positive and negative opposites between the values of different categories in the same dimension of doc2vec. (2) The absolute value of the difference in the same dimension of doc2vec is positively related to the importance of dimension. The larger the absolute value is, the larger the influence coefficient is, which means the more important the features of the dimension are.

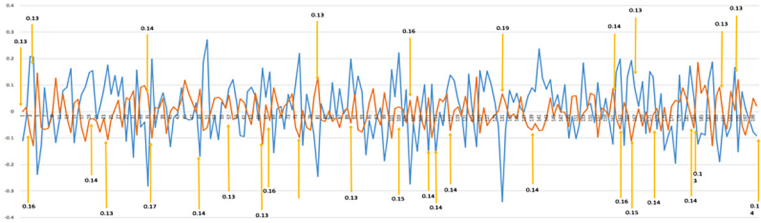

Fig. 1. Doc2vec-200 dimensions (Color figure online)

(2) Research on the Interpretability of Text Style Based on Style Vector and Influence Function.

From the typical content of the executive profile (Table 6), it is found that the executives of class a text graduated from famous universities with complete professional experience and have held important positions in well-known companies; while the executives of class B and C text have poor educational background and the companies they have held are not well-known, and they have not held important positions at the same time. From the perspective of human intuitive feelings, there is a big difference in the style of class A, B and C texts, and a small difference in the style of class B and C texts. At the same time, it also gives people the feeling that the operating risk of A-class companies may be lower than that of B-class and C-class companies, and the probability of A-class companies constituting fraud to users may also be lower.

Table 6. Typical contents of executive profile.

Text	Executive profile	State
A	***, Bachelor of science, Peking University. He has successively held CEO, CTO, vice president and other senior positions in Founder group of Peking University, China interactive media group, with profound technical management, team management ability and rich experience in industries	Normal
B	***Graduated from the school of management, Qingdao University of science and technology. I worked as an administrative assistant in Qingdao priority Export Co., Ltd in 2004–2006. I worked as an administrative manager in Ningbo Aksu Nobel Chemical Co., Ltd in 2007–2013	Abnormal
C	***, worked in 2005, engaged in real estate projects since 2008 in small loan business. Rich experience in business management	Abnormal

For the data in Table 6, we use the style vector in Sect. 3.2 to construct feature. Through the method of Sect. 3.2, the scatter diagram in Fig. 2 is obtained by combining style vector formed by the text of class A and class B with influence function, and the scatter diagram in Fig. 3 is obtained by combining style vector formed by class B and class C text with influence function. It is found from the figure that the slope of straight line fitted by each point in Fig. 2 is larger than that in Fig. 3. It shows that the larger the

difference of text style is, the larger the slope of line based on style vector and influence coefficient is.

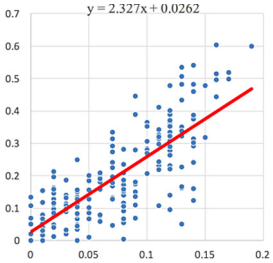

Fig. 2. Class A and B

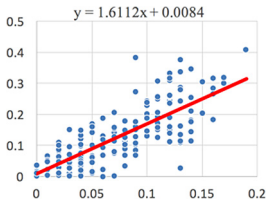

Fig. 3. Class B and C

Then we use the text with a large style difference between class A and class B to retrain the machine learning model. At this time, the accuracy of risk identification reaches 100%; while using the text with a small style difference between class B and class C to retrain the machine learning model, the accuracy is only 74.03%. It can be seen that in the actual task of risk identification of text classification, due to the existence of noise data, using all the data to train the model often can not achieve the highest accuracy. The training data can be effectively filtered by using style vector to analyze the interpretability of text style, so as to improve the accuracy of machine learning model.

In this case, where the text style of executive profiles varies widely, through the above experiments, it is found that using style vector to describe the differences of text style is the same as human's intuitive feeling, which shows that style vector can describe the differences between text styles. Moreover, the larger the slope of the line based on the style vector and the influence coefficient is, the greater the style difference of the text is, and the higher of accuracy in risk identification is. It shows that the machine learning model really learns the text style difference in human cognition and uses it in risk judgment, which further shows that the model has certain machine learning interpretability.

5 Conclusions

To sum up, in view of the evaluation and analysis of the fraud risk of P2P Internet loan companies, this paper proposes to explain the model behavior based on the psychological interpersonal fraud theory in the field of social science, and uses NLP and influence function in the field of natural science to verify that the machine learning model does learn the details of the fraud theory, which also provides the heart for the machine learning model of P2P risk evaluation The support of the interpretability of Neo Confucianism. On the other hand, machine learning interpretability research based on text style not only improves the accuracy of the model, but also provides interpretability of text style for P2P risk assessment model from the perspective of text style. In the future, we will do further research on the risk assessment of P2P companies, especially the analysis of more model interpretability, extraction of better semantic features and optimization of classification model.

Acknowledgements. This work was supported by Beijing Municipal Commission of Science and Technology [grant number Z181100001018035]; National Social Science Foundation of China [grant number 16ZDA055]; National Natural Science Foundation of China [grant numbers 91546121, 71231002]; Engineering Research Center of Information Networks, Ministry of Education; Beijing BUPT Information Networks Industry Institute Company Limited; the project of Beijing Institute of Science and Technology Information.

References

1. Guo, Y., Zhou, W., Luo, C., et al.: Instance-based credit risk assessment for investment decisions in P2P lending. Eur. J. Oper. Res. **249**(2), 417–426 (2016)
2. Luo, C., Xiong, H., Zhou, W., et al.: Enhancing investment decisions in P2P lending: an investor composition perspective. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August. DBLP, pp. 292–300 (2011)
3. Zhao, T., Li, L., Xie, Y., et al.: Data-driven risk assessment for peer-to-peer network lending agencies. In: 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS). IEEE (2018)
4. Hendricks, D., Roberts, S.J.: Optimal client recommendation for market makers in illiquid financial products. In: Altun, Y., et al. (eds.) ECML PKDD 2017. LNCS (LNAI), vol. 10536, pp. 166–178. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71273-4_14
5. Jin-Qun, H.E., Liu, P.J.: The documents classification algorithm based on LDA. J. Tianjin Univ. Technol. **4**, 28–31 (2014)
6. Mikolov, T., Chen, K., Corrado, G., et al.: Efficient estimation of word representations in vector space. Computer Science (2013)
7. Le, Q.V., Mikolov, T.: Distributed representations of sentences and documents. Eprint Arxiv, vol. 4, pp. 1188–1196 (2014)
8. Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv: 1702.08608v2 (2017)
9. Wang, X., He, X., Feng, F., et al.: TEM: tree-enhanced embedding model for explainable recommendation. In: The 2018 World Wide Web Conference (2018)
10. Melis, D.A., Jaakkola, T.: Towards robust interpretability with self-explaining neural networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp. 7775–7784. Curran Associates Inc., Red Hook (2018)
11. Koh, P.W., Liang, P.: Understanding black-box predictions via influence functions. In: Proceedings of the 34th International Conference on Machine Learning, vol. 70, pp. 1885–1894 (2017). JMLR.org
12. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why should I trust you?": explaining the predictions of any classifier. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. ACM (2016)
13. Lombrozo, T.: Causal–explanatory pluralism: how intentions, functions, and mechanisms influence causal ascriptions. Cogn. Psychol. **61**(4), 303–332 (2010)
14. Byrne, R.M.J., Mceleney, A.: Counterfactual thinking about actions and failures to act. J. Exp. Psychol. Learn. Mem. Cogn. **26**(5), 1318–1331 (2000)
15. Wu, S., Jin, S.H., Cai, W.: Detecting deception by verbal content cues. Progress Psychol. Sci. **20**(3), 457–466 (2012)
16. Das, S.P., Padhy, S.: A novel hybrid model using teaching–learning-based optimization and a support vector machine for commodity futures index forecasting. Int. J. Mach. Learn. Cybernet. **9**(1), 97–111 (2018)
17. Kim, Y.: Convolutional neural networks for sentence classification. Eprint Arxiv (2014)
18. Minami, S.: Predicting equity price with corporate action events using LSTM-RNN. J. Math. Financ. **08**(1), 58–63 (2018)