

Chapter 3

Knowledge Discovery from Online Reviews



Jiangning Wu and Tong Yang

3.1 Overview of Online Reviews Mining

Given the popularity of online reviews, scholars in various industries and disciplines have focused on exploring business opportunities through mining online reviews, such as tourism, hospitality, manufacturing, etc. Regarding the research topics, it is easy to find a wide range of involvement. Mainstream topics include product ranking systems, user preference analysis, review usefulness analysis, and product competitive analysis.

3.1.1 Product Ranking System Based on Online Reviews

To support consumers' decision-making in choosing products, many studies in recent years have focused on constructing product ranking models based on online reviews or user-generated content (UGC). As per the theories and methods, these studies could be classified into two categories, namely mathematical programming-based methods and multi-attribute decision-making-based methods (Bi et al., 2022). The first methods aim at both optimizing consumer utility and minimizing search costs. By constructing such mathematical programming models, the results are obtained in such a way that the optimization objectives are satisfied as much as possible. For example, Ghose et al. (2012) proposed a mathematical programming model to rank products using user-generated content and crowdsourced content. In this method, the authors posited that the products providing higher surplus (also called consumer utility) should be ranked first. By collecting online surveys, the

J. Wu (✉) · T. Yang
Institute of Systems Engineering, Dalian University of Technology, Dalian, China
e-mail: jnwu@dlut.edu.cn; yangt014@mail.dlut.edu.cn

validity of the proposed model was validated. Similar to Ghose et al.'s (2012) work, Rianthong et al. (2016) constructed a stochastic programming model, which improved the efficiency of product ranking. The authors considered optimizing utility and reducing search costs, while also considering consumer preferences.

The second category of studies considers both product information and consumer preferences to calculate the ranking values. Based on the ranking values, multi-attribute decision-making methods are utilized to determine the optimal results. This method is currently more popular. For example, Tayal et al. (2022) proposed a new multi-criteria decision-making (MCDM) method for the personalized ranking of products based on multiple dimensions. Specifically, the authors first determined customer preferences as input for decision-making. Then aspect-level sentiment analysis was utilized to calculate attribute performance. This model subtly included customer preferences by mapping different customer preferences to Plithogenic degrees of modeling linguistic uncertainty in online reviews to create a personalized product ranking using Plithogenic aggregation. The comparisons with existing MCDM methods have demonstrated its superiority. Essentially, this approach is carried out based on the calculation of aspect-level sentiment in online reviews, and thus the accuracy of sentiment calculation would influence the ranking effectiveness. For this concern, Fu et al. (2020) constructed deep learning models to accurately identify the sentiment orientation embedded in online reviews and then extracted the feature-opinion pairs. Especially, the authors utilized the interval-valued Pythagorean fuzzy-weighted Heronian mean operator to aggregate the attribute information based on the interrelationship between product attributes. Subsequently, product ranking was achieved, and a case study was experimented with to validate the proposed model. Overall, studies on product ranking based on online reviews are gaining attention.

3.1.2 User Preference Analysis

With the rapid development of information technology and social media, more and more consumers are posting their products/services using experiences online, which incorporate rich consumption perceptions and consumers' attitudes to products/services. Such data are characterized by massive, real, insightful, and passionate. Importantly, sentiments embedded in online reviews mirror consumers' preference for products/services. Therefore, numerous scholars from various fields have conducted research on mining online reviews to capture consumer preferences, such as tourism, manufacturing, and hospitality. These studies posit that online reviews are embedded with consumers' detailed perceptions of multiple aspects, and these positive or negative opinions directly reflect consumers' preferences. Hence to enhance customer satisfaction, managers and practitioners should extract consumers' preferences from massive online reviews. For example, Zhang et al. (2021) constructed an aspect-level sentiment analysis model to identify hotel customers' preferences. Vu et al. (2019) presented a method that utilized online

restaurant reviews and text-processing techniques to investigate tourists' dining behaviors. Regarding the manufacturing field, Xiao et al. (2016) explored how to measure mobile phone users' preferences based on the Kano model, and the results demonstrated the superiority of the proposed models.

Similar to the above preference analysis studies based on online reviews, there are also studies that focus on understanding consumer satisfaction through mining online reviews. Customer satisfaction is the compared results of subjective expectations and the actual performance of a product/service. When the actual performance is higher than expectations, consumers are satisfied; otherwise, consumers are unsatisfied. For example, Guo et al. (2017) explored hotel customers' satisfaction by mining sentiment information in online reviews, and the results demonstrated the efficiency of such data. Aiming at revealing why consumers are satisfied, Kim et al. (2022) investigated user opinions via online restaurant reviews. Similarly, Liu, Song, Sun, et al. (2020) analyzed the relationship between food quality and consumers' dining satisfaction using online reviews from a perspective of negative bias. Regarding the tourism field, Park et al. (2020) focused on the asymmetric relationship between attribute performance and customer satisfaction and unveiled the asymmetric effects of online airline reviews. Satisfaction analysis research can also be easily found in the manufacturing field. Imtiaz and Islam (2020), for instance, identified the influential features of smartphones on consumer satisfaction from online reviews and validated how these features determine satisfaction.

3.1.3 Review Usefulness Analysis

Consumers are used to reading online reviews to understand the products before making purchase decisions. But not all online reviews are useful for consumers; especially in the face of massive amounts of data, consumers need to know exactly which reviews are useful and which are not. As per the theory of reasoned action, trust in sellers is one of the determinants of online consumption intention, and such behaviors of consumers are influenced by existing online reviews. In this regard, online reviews could influence product sales and thus are of great importance for sellers and businesses (Choi & Leon, 2020). Especially, useful information might diffuse faster among consumers (Pavlou, 2003), and thus useful online reviews could be efficiently identified to help build consumer trust. For this concern, a lot of academic efforts have been devoted to the usefulness analysis of online reviews.

In general, these studies focus on two aspects of review usefulness, namely reviews and reviewers. Review factors include review content, length, ratings, sentiments, etc.; reviewer factors involve the reviewer's expertise, identity, ranking, etc. Specifically, some scholars investigate how review extremity influences review usefulness. For instance, Siering and Muntermann (2013) explored the relationship between extremity and review usefulness using online reviews from Amazon and found a negative effect; in contrast, Cao et al. (2011) found a positive relationship based on online reviews from CNET. Moreover, many studies have confirmed the

fact that longer reviews reflect more useful information. Regarding the reviewer factors, most studies have reported a positive relationship between reviewers' expertise and review usefulness; as Choi and Leon (2020) explained, experienced reviewers have more knowledge and are more likely to discuss the product both positively and negatively, so other consumers will find such reviews more useful.

3.1.4 Product Competitive Analysis

Product competitiveness is the combined ability of two or more products in the competition. Traditional competitiveness analysis studies were typically conducted by surveys and questionnaires. In recent years, scholars have started to focus on online reviews for competitiveness analysis. These studies argued that the competitiveness of a product (service) could be reflected by its consumers' satisfaction level, and the sentiments hidden in online reviews, as we mentioned above, mirror consumers' satisfaction. More importantly, consumers may make a purchase decision by reading online reviews, and thus such data are very important for merchants to evaluate the market performance of their products (services). Therefore, numerous studies have analyzed product (service) competitiveness using online reviews. In terms of the areas of these studies, both service and manufacturing industries are relevant. For example, in the hospitality literature, Gao et al. (2018) proposed a comparative opinion mining method to identify the competitors of the target restaurant, and the results revealed which attributes of the target restaurant performed worse than others. Similarly, Wang et al. (2017) also utilized the comparative opinion mining algorithm to extract consumers' opinions of the restaurants and clearly pointed out the restaurant's comparative strengths and weaknesses. As for the manufacturing field, Liu et al. (2021) mined the product competitiveness of smartphones by fusing multisource online information and tested how different factors influence product competitiveness. Liu, Jiang, and Zhao (2019) proposed a supervised learning method to identify competitors from user-generated content and aspect-level sentiment analysis to assess consumers' sentiment attitudes.

3.2 Online Reviews Information Mining Techniques

With the development of e-commerce platforms and online reviews, as well as the emergence of natural language processing and deep learning, techniques for processing online reviews have been boosted. Overall, frequently used online review mining techniques include information extraction, sentiment analysis, text categorization, etc.

3.2.1 Information Extraction Technique

Information Extraction (IE) is an important subtask of natural language processing (NLP), which is the process of extracting useful structured information from unstructured data. Nadeau and Sekine (2007) defined IE as “the extraction of instances of predefined categories from unstructured data to construct a structured and explicit representation of entities and their relationships.” It takes as input a collection of documents such as research papers, logs, and online reviews and generates a representation of relevant information that satisfies different predefined categories. IE techniques efficiently analyze different forms of texts by extracting the most valuable and relevant information from unstructured data. Thus, its ultimate goal is to identify particular facts from texts to enrich the database or knowledge base.

IE includes different subtasks, namely, named entity recognition (NER), relationship extraction, event extraction, and significant fact extraction. NER is a very important task for extracting descriptive entities in the IE. It identifies generic or intra-domain entity objects such as organization, business name, location, disease, etc. The relationship extraction task can help in completing the annotation of data by extracting the relationships between different entities. The immediate goal of the event extraction task is to identify specific types of events and to determine the elements of the events that hold a given role. From the perspective of theoretical development, event extraction helps us gain insights into the mechanism of machines to understand data and the world, as well as our own cognitive mechanism; from the perspective of the application, event extraction techniques can help us solve many real-world problems, such as the automatic processing of massive amounts of information mentioned earlier.

3.2.1.1 Named Entity Recognition

The named entity recognition (NER) task is the process of identifying and mapping entities to predefined categories. Its main application scenarios include question-and-answer systems, machine translation, information retrieval, opinion mining, and knowledge base populating. Therefore, the efficiency and accuracy of NER are crucial.

Traditionally, NER uses rule-based approaches, learning-based approaches, or hybrid approaches. Rule-based NER systems rely on manually set rules, which are generally designed based on domain-specific dictionaries and syntactic-lexical patterns. Rule-based NER systems can achieve good results when the dictionary size is limited. Due to the domain-specific nature of the rules and the incompleteness of the dictionaries, such NER systems are characterized by high accuracy and low recall, and similar systems are difficult to migrate to other domains (domain-based rules are often not generalized and require reformulation of rules for new domains and different domain dictionaries). Learning-based methods include unsupervised and supervised learning methods. A typical unsupervised learning approach is clustering,

where named entities are extracted from clusters based on semantic similarity, and the core idea is to use lexical resources, lexical models, and statistics obtained from a large corpus to infer the class of named entities. In the supervised learning approach, NER is converted into a multi-categorization or sequence labeling task. Based on the labeled data, researchers apply domain knowledge and engineering skills to design complex features to characterize each training sample, and then apply machine learning algorithms to train the model to make it learn the patterns of the data. Many machine learning algorithms have been applied in supervised NER. With the development of information technology, deep learning algorithms have also been applied in NER (Che et al., 2019). NER utilizes the deep learning nonlinear relationship fitting capability to be able to learn more complex features from data. At the same time, deep learning does not require overly complex feature engineering and is able to learn features from data automatically.

3.2.1.2 Relationship Extraction

Relationship extraction is another subtask of information extraction, which aims to identify the relationships of entities in the text. Relationship extraction is important for knowledge base construction and understanding of the text and plays an important role in application scenarios such as question and answer, text understanding, etc. According to the complexity of the task, relationship extraction can be divided into simple relationship extraction and complex relationship extraction. The purpose of simple relationship extraction is to identify the relationship between two entities from the text. For example, a binary relationship can be extracted from the sentence “Tsinghua University is located in Beijing, the capital of China,” i.e., “Tsinghua University (Entity) is located in Beijing (Entity).” The current methods of simple relation extraction can be divided into supervised, semi-supervised, and weakly supervised methods. The supervised learning-based simple relation extraction method uses high-quality labeled sample data to train learning, which is obtained through manual labeling or crowdsourcing. With the development of deep learning techniques in recent years, many new models of neural network-based relationship extraction have emerged, such as the graph-based neural network model C-CGNN (Zhang et al., 2018), the pre-training-based method EPGNN (Zhao et al., 2019), and the capsule network-based method (Zeng et al., 2018). However, in many specific domains, it is difficult to obtain high-quality labeled data, yet there is a large amount of unlabeled data available. To be able to utilize a large amount of unlabeled data in the training phase, semi-supervised learning-based relationship extraction models attempt to learn from both labeled and unlabeled data together (Luo et al., 2019). Complex relationship extraction is an emerging research direction that attempts to extract more complex relationships involving multiple entities or under specific constraints. Therefore, some scholars have proposed joint extraction methods (Yuan et al., 2020; Zheng et al., 2017) to extract such complex entity relations. However, at present, there is still relatively more room to explore this method.

3.2.1.3 Event Extraction

Event extraction can be divided into open-domain-based event extraction (Liu, Huang, & Zhang, 2019) and closed-domain-based event extraction (Sheng et al., 2021). Open-domain event extraction refers to acquiring a series of events related to a specific topic, which usually consists of multiple events. As to the closed-domain event extraction, its task is to find words belonging to a specific temporal pattern that reflect a change in the action or state that occurred, such as time, person, place, etc. In the open domain event extraction task, events are usually some descriptions related to a certain topic, which can be achieved by clustering or classification. In either task, the purpose of event extraction is to capture the types of events from a large number of texts and present the essential arguments of the events in a structured form.

3.2.2 *Sentiment Analysis Technique*

3.2.2.1 Sentiment Dictionaries

As an automated, unsupervised sentiment analysis method, sentiment dictionaries have attracted a great deal of interest from academics and businesses. Companies attempt to mine user reviews and social media content to understand users' sentiments and opinions about their products and services. This approach uses constructed lexicons that contain words marked as positive, negative, or neutral (sometimes with values that reflect the intensity or strength of the sentiment). The lexicon can be constructed in different ways, such as manually; using a corpus of automatically associated words with known seed words; or semi-automatic acquisition of sentiment values based on WordNet.

The advantage of the sentiment dictionary approach is that it does not require manual annotation of the data. With the proliferation of product review sites with user reviews and ratings, the Internet has seen a large number of domain-specific unlabeled online reviews, so sentiment analysis methods based on sentiment dictionaries can be useful. At the same time, however, there are some problems with this approach, such as the fact that words may have multiple meanings, and meanings that are common in one domain may not be common in another. In addition, words that are not usually considered emotionally charged may also be emotionally charged in a given context. Finally, and the biggest problem, this kind of method has limited accuracy.

3.2.2.2 Machine Learning

Online reviews are often poorly structured and have large amounts of data, which makes manual processing very difficult. Among natural language processing and text mining approaches, machine learning is often used to process unstructured data

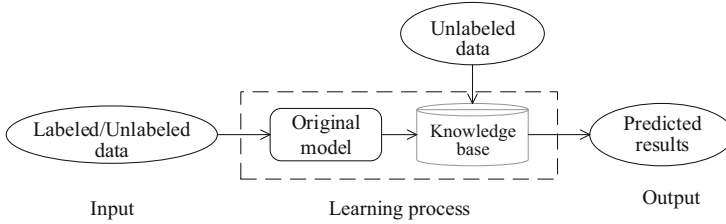


Fig. 3.1 Machine learning method flowchart

and has been widely used to mine the sentiment of online reviews. Machine learning uses information technology to learn patterns from past experiences and then accurately predict existing and future data. The term “experience” refers to past data, which is used to build classification models, which may come from online open platforms, or questionnaire data.

The classical machine learning method flow is shown in Fig. 3.1.

The input to this flow is labeled data or unlabeled data, representing supervised and unsupervised machine learning, respectively. The original model is trained using the input data, and the knowledge base holds the completed trained model and its parameters. The new unlabeled data is then predicted, and the output is the prediction result. Machine learning algorithms can usually be classified into two types: supervised learning and unsupervised learning. Supervised learning refers to the training of classification or prediction models using prior knowledge from the input data. This type of input data is usually labeled datasets. The purpose of supervised learning is to obtain the desired output based on the input data. Common supervised learning models include logistic regression, random forest, and decision trees. In unsupervised learning, model building and training do not require the use of labeled datasets. The purpose of unsupervised learning is to perform dimensionality reduction and exploratory analysis on data with high-dimensional features. Commonly used unsupervised learning models include clustering algorithms such as K -means, SOM, and some optimization algorithms. As shown in Table 3.1, we sort out some common machine learning algorithms used in sentiment analysis.

3.2.2.3 Deep Learning

Traditionally, sentiment analysis based on machine learning algorithms requires first modeling the comment texts to extract features. One commonly used approach is the bag-of-words (BoW) model; however, the BoW model ignores the semantic and word order features of the review texts, which is the core of the texts. Another approach to feature extraction is the N -gram approach, which overcomes the shortcomings of BoW but also creates the new problem of over-sparse high-dimensional vectors. All of these are problems inherent to traditional machine learning, which requires features to be extracted from the data in order to train the model. Unlike machine learning, deep learning solves these problems through deep neural

Table 3.1 Commonly used machine learning algorithms in sentiment analysis

Type	Algorithm name	Description
Supervised	LR (logic regression)	Logistic regression assumes that the data obeys the Bernoulli distribution (0-1 distribution) and uses the gradient descent method to solve the parameters by the method of the great likelihood function for the purpose of dichotomizing the data (Rafeek & Remya, 2017)
	Decision tree	Classification of data into predefined classes according to the “Yes” or “No” rule (Mullainathan & Spiess, 2017)
	SVM (support vector machine)	Determining multidimensional boundaries that separate data points belonging to different classes (Srivastava and Bhambhu 2010)
	Random forest	A “forest” classifier built based on decision tree integration with a typical training process of “bagging” strategy (Probst et al., 2019)
	NB (Naïve Bayes)	Predicting the classification of the output by calculating conditional and prior probabilities using Bayes’ theorem (Jensen & Nielsen, 2007)
Unsupervised	K-means	It divides the set of samples into k subsets, which constitute k classes, and assigns n samples to k classes, each with the minimum distance to the center of the class to which it belongs (Vashishtha & Susan, 2021)
	KNN (K-nearest neighbor)	A sample belongs to a class if most of its k most adjacent samples in the feature space belong to that class and it has the characteristics of samples in that class (El Mrabti et al., 2018)
	Association rules	Association rule mining finds associations or interconnections between sets of items in the data (Jiang et al., 2018)

networks. The depth of a deep learning model is the number of neural network layers between the input and output layers. While shallow neural networks extract abstract features of the data, as learning advances, deep neural networks can extract certain features that are meaningful in the data. This approach is based on pre-trained word vectors, such as W2V (Rong, 2014), GloVe (Glove, 2014), and fastText (fasttext, 2016), which can transform input text into high-dimensional word vectors. Unlike machine learning, which requires manual extraction of data features, deep learning can automatically extract data features. Thus, deep learning can learn and make intelligent decisions on its own.

For the task of sentiment classification, the types of applications of deep learning algorithms are divided into two main categories, namely, convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The convolutional neural network is a feed-forward deep neural network that was originally applied in the field of image processing. Due to its excellent local feature extraction capability, it has also been used in recent years for text sentiment analysis tasks. CNNs consist of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. The convolutional layer extracts local features of the input data, and different convolutional layers can extract different local features, and then the

Table 3.2 Commonly used deep learning algorithms

Type	Algorithm name	Description
Feed-forward	NN (neural networks)	Neural networks reduce the error between truth and prediction by adjusting the weights between neurons (Yang & Sudharshan, 2019). They can be further extended as a multilayer perceptron, back-propagation neural network, and Hopfield network
	CNN	Convolutional neural networks can effectively extract local features from the data (Liu et al., 2015)
Back-propagation	RNN	Recurrent neural networks can learn dependencies in the data (Mikolov et al., 2010)
	LSTM	A special type of recurrent neural network (RNN) that learns long-term dependencies in data (Sherstinsky, 2020)
	GRU	A special type of recurrent neural network (RNN) that learns long-term dependencies in data with less “forget gate” and easier convergence than LSTM (Liu, Wang, Zhu, et al., 2020)
	BiLSTM	A sequence processing model consists of two LSTMs: one receiving input in the forward direction and the other in the reverse direction (Xu et al., 2019)

pooling layer reduces the dimensionality of the features. The feature extraction capability of CNN depends on the number of convolutional and pooling layers. The feature data is then passed into the activation function, Rectified Linear Unit (ReLU), or sigmoid, and the model is trained using a loss function to evaluate the learning process, and the above steps are cycled through.

Unlike feed-forward CNNs, RNNs are back-propagation networks. In addition to the current input, it also considers the previous input. RNNs can process data with sequential relationships with the help of internal memory. It is designed on the principle that humans do not start thinking from zero every time, and therefore need to store and remember prior knowledge. Therefore, RNNs can predict subsequent words based on previous words. Commonly used RNNs include the LSTM (Long Short-Term Memory) model and GRU (Gate Recurrent Unit) model, both of which are extensions of RNNs. LSTM can perform long-term memory storage, which overcomes the gradient disappearance problem of classical RNNs. Unlike LSTM, GRU has a more streamlined structure and does not include a “forget gate,” so the model training can converge faster. We summarize the commonly used deep learning algorithms for sentiment analysis tasks, as shown in Table 3.2.

The main application scenarios of sentiment analysis include analysis of online reviews of products (to understand user satisfaction, develop targeted marketing strategies, competitor analysis), analysis of online reviews of special products such as movies (to adjust uptime and marketing strategies), etc.

3.2.3 Text Categorization Technique

Text categorization techniques can be classified into three categories: traditional methods, fuzzy logic-based methods, and deep learning-based methods. In the following, we briefly describe these three methods.

3.2.3.1 Traditional Methods

As mentioned above, text categorization is an important part of text mining. Classifying text from online reviews can help consumers reduce the cost of retrieving information and also help e-commerce platforms obtain effective information. Many studies have been conducted on text categorization using traditional methods (Abrahams et al., 2012; Lan et al., 2009; Liu, Wang, Fan, et al., 2020). A deeper understanding of the feature extraction methods and the correct method of classifier evaluation can ensure the effective operation of traditional text categorization methods. Common feature extraction methods used in text categorization tasks include word frequency analysis, inverse document frequency analysis, N -gram, and other word-embedding models. Many methods have been shown to reduce the complexity of the text categorization process, such as principal component analysis, and information gain. And there is a wide range of classifiers to choose from, such as Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), Multilayer Perceptron (MP), etc.

3.2.3.2 Fuzzy Logic-Based Methods

As a subtask in the field of NLP, text mining inevitably faces the fuzzy nature of natural language. This phenomenon may lead to the misclassification of texts. Therefore, the task of text categorization based on fuzzy logic is born. This approach has been proposed decades ago (Zadeh, 1965) for the study of uncertain knowledge. Unlike traditional mathematics that classifies elements explicitly to a certain set, this approach uses the concept of affiliation for element classification, i.e., fuzzy sets. Fuzzy sets describe the fuzzy concepts embedded in linguistic features, such as “cold” and “a little cold” to describe the weather. Fuzzy logic operates by mimicking the human brain’s processing mechanism for uncertain events. In addition, fuzzy logic can be used to convey knowledge and is also good at building uncertain boundaries, the basic structure of which is shown in Fig. 3.2. In view of these

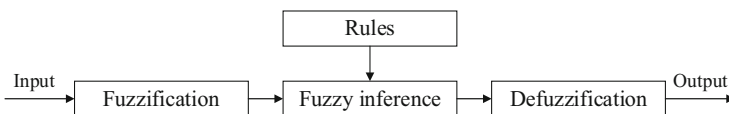


Fig. 3.2 Fuzzy logic diagram (Zadeh, 1965)

advantages of fuzzy logic methods in knowledge representation, it has received increasing attention from researchers in text categorization problems. This approach has been developed in many fields, such as software development, NLP, and cybernetics.

3.2.3.3 Deep Learning-Based Methods

Deep learning-based text categorization methods differ from traditional neural networks in that their core lies in the size of the hidden layers between the input and output layers; the larger the number of hidden layers and the more nodes, the better the fitting ability of the deep learning model. There are many deep learning models applied to text categorization, such as the classical deep neural network model (DNN), recurrent neural network model (RNN), long short-term memory network model (LSTM), convolutional neural network model (CNN), BERT, etc. Neural network models were initially applied to image processing and pattern recognition tasks, and then gradually applied to NLP, including text categorization tasks, due to their excellent feature extraction and autonomous learning capabilities. A deep learning model is trained using a dataset with category labels, and the model automatically learns and extracts features related to the labels from the data and saves the parameters, and then performs category prediction on new text data.

Traditional machine learning methods require explicit data features, and then the classifier learns from the data. But deep learning does not need to specify data features, and it can learn from data automatically. At the same time, deep learning methods are faster and more accurate. As a result, deep learning is now being applied in a wider range of contexts for text categorization.

3.3 Commercial Value Discovery of Online Reviews

This section focuses on the commercial uses of online reviews. Three cases are provided below which cover the topics of product ranking, relationships between retail prices and online reviews, and personalized online reviews ranking.

3.3.1 *Word-of-Mouth Ranking of Products by Using Online Reviews*

Online reviews play an important role as electronic word-of-mouth (eWOM) for potential consumers to make informed purchase decisions. However, the large number of reviews poses a considerable challenge because it is impossible for customers to read all of them for reference. Moreover, there are different types of

online reviews with distinct features, such as numeric ratings, text descriptions, and comparative words, for example, such heterogeneous information leads to more complexity for customers.

To help consumers compare alternative products, there are some studies that describe methods to extract opinions and sentences from text reviews. Some of these studies focus on mining and summarizing customers' opinions and text sentiments from text reviews, whereas other studies directly mine comparative sentences and relationships from text reviews. However, comparative sentences are very rare in text reviews, and they are usually not sufficient to evaluate competitive products comprehensively because there are a limited number of comparisons for some products and even no comparison for many (if not most) products.

To compare multiple products, some websites provide rankings of products according to simple criteria such as the average numeric rating; however, these rankings do not fully consider the voice of customers, such as the text sentiments and comparative sentences. Hereafter, a novel method that integrates heterogeneous information including text sentiments, numeric ratings, comparative sentences, and comparative votes (Yang et al., 2016) is presented. In detail, heterogeneous information is divided into two categories: descriptive information and comparative information. Descriptive information consists of text sentiments and numeric ratings to describe one specific product. Comparative information comes from comparative sentences and online comparative votes that compare more than one product. The flowchart of the proposed method is shown in Fig. 3.3. At first, a crawler is implemented to collect product data from Zol.com.cn. The dataset is from the mobile phone category and comprises three types of consumer reviews: numeric ratings, text reviews, and comparative online votes. By mining text reviews, the sentiment of reviews and comparative sentences can be obtained. Then two types of information can be derived, i.e., descriptive information including numeric rating and sentiment of reviews, and comparative information including comparative sentences and votes.

To consider descriptive information and comparative information simultaneously, a graph structure is applied in which the nodes are the given products, the weights of nodes are derived from descriptive information, and the edges represent pair-wise comparative relationships. A benefit from the graph structure is that the integrated eWOM score can be calculated and an overall ranking of the given set of products can then be generated.

The sentiment embedded in the textual content can be classified as positive or negative. The positive and negative terms in text reviews are assigned explicit polarity values 1 and -1 , respectively. For example, given a product i at time t (denoted as P_{it}) and a text review about P_{it} (denoted as TR_{it}), the text review is labeled with a set of sentiment terms, i.e., $TR_{it} = \{tr_{it1}, tr_{it2}, \dots, tr_{itm}\}$. The overall sentiment score of P_{it} , denoted as $T_Score(P_{it})$, can be computed as follows:

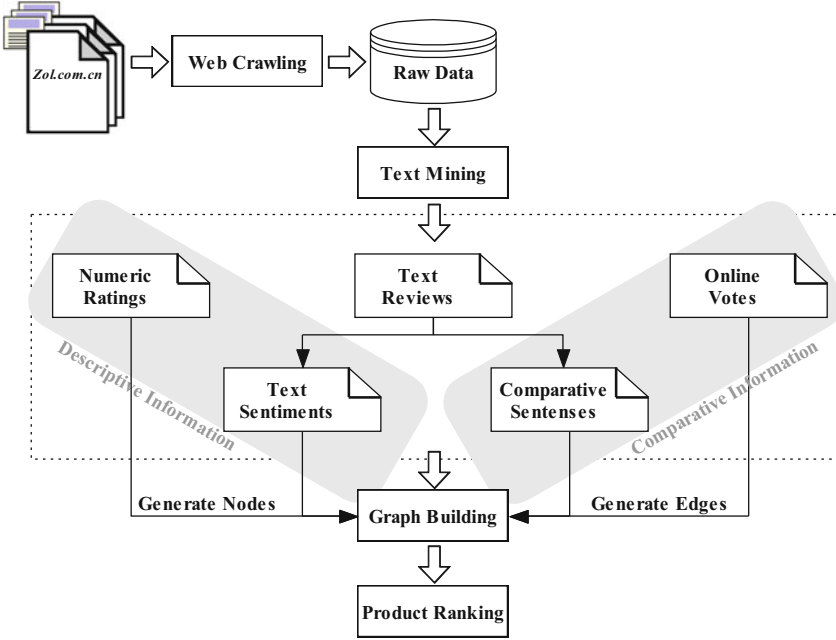


Fig. 3.3 The flowchart of products eWOM ranking

$$T_Score(P_{it}) = \frac{PO_{it}(TR_{it}) - NO_{it}(TR_{it})}{PO_{it}(TR_{it}) + NO_{it}(TR_{it})} \quad (3.1)$$

where the prefix T of $T_Score(P_{it})$ indicates that the score is calculated from a text review, PO_{it} is the number of occurrences of positive terms in all text reviews for P_{it} , and NO_{it} is the number of negative terms.

In addition to the implicit scores in textual contents, another form of the score is readily available, i.e., the ratings in numeric values following each piece of text review. For product i at time t with a set of numeric ratings, denoted as $NR_{it} = \{nr_{it1}, nr_{it2}, \dots, nr_{itm}\}$, the average score rating for P_{it} , denoted as $R_Score(P_{it})$, can be calculated by

$$R_Score(P_{it}) = \frac{\sum_{j=1}^m \text{rate}(nr_{itj})}{m} \quad (3.2)$$

where prefix R of $R_Score(P_{it})$ means that the score is obtained from a rating value, and $\text{rate}(nr_{itj}) \in \{r_{\min}, \dots, r_{\max}\}$ is a discrete rating value within the range of r_{\min} and r_{\max} . The numeric ratings are usually on a 1 ~ 5-star scale on most review websites, i.e., $r_{\min} = 1$ and $r_{\max} = 5$, wherein a 1-star rating shows the least satisfaction, and a 5-star rating indicates the most satisfaction.

The T_Score from text review, with a range of $[-1, 1]$, will be integrated with the R_Score from numeric rating; however, a problem is that they do not share the same scale. To facilitate the integration, T_Score is scaled up to the range of 1 to 5. For example, if the originally obtained T_Score value is $x = -0.5$, the value transformed into $[1, 5]$ is calculated by $f(x) = 2x + 3 = 2$. For the product without reviews, the sentiment score is set to be a neutral score of 3 by default. The $f(x)$ function is designed from min-max normalization:

$$f(x) = \frac{x - \min}{\max - \min} (\text{new_max} - \text{new_min}) + \text{new_min} \quad (3.3)$$

where \min and \max are the original minimum and maximum values of an attribute x . Min-max normalization maps a value by $f(x)$ into the new range $[\text{new_min}, \text{new_max}]$. In addition to min-max normalization, there are other methods for data normalization, such as z-score normalization and normalization by decimal scaling. However, certain normalization schemes can change the original data quite a bit, particularly the z-score normalization and normalization by decimal scaling. The min-max normalization is the preferred method that preserves the relationship among the raw data values.

The overall score of P_{it} , denoted as $\text{Score}(P_{it})$, is calculated by combining $T_Score(P_{it})$ and $R_Score(P_{it})$:

$$\text{Score}(P_{it}) = \alpha T_Score(P_{it}) + (1 - \alpha) R_Score(P_{it}) \quad (3.4)$$

where $\alpha \in (0, 1)$ is an adjustment factor to balance the effects of T_Score and R_Score .

Comparative information can be expressed as comparative sentences extracted from text reviews, which are identified via comparative keywords, sentence semantics, and sentence structure, as well as fuzzy linguistics. Given a set of sentences comparing two products P_i and P_j , denoted as $S = \{s_1, s_2, \dots, s_n\}$, the information of a comparative relationship derived from one text sentence $s_k \in S$ is described as a quadruple

$$T_Relation_{ij}(s_k) = (P_i, P_j, \text{Vote}_k^T(P_i|P_i, P_j), \text{Vote}_k^T(P_j|P_i, P_j))$$

where the prefix T indicates that the information is obtained from text reviews. $\text{Vote}_k^T(P_i|P_i, P_j)$ equals 1 if the product P_i is commented on as better than P_j in sentence s_k ; otherwise, it is 0. For example, if there is a comparative sentence $s_k = \text{"mobile phone A is less than mobile phone B,"}$ the corresponding quadruple $T_Relation_{ij}(s_k)$ can be written as $(A, B, 0, 1)$. When considering all the comparative sentences, the overall $\text{Vote}^T(P_i|P_i, P_j)$, which denotes the total number of sentences in S preferring P_i to P_j , is computed as follows:

$$\text{Vote}^T(P_i|P_i, P_j) = \sum_{k=1}^h \text{Vote}^T_k(P_i|P_i, P_j) \quad (3.5)$$

The comparative votes can supply direct comparative relationships, also denoted as a quadruple:

$$V_Relation_{ij} = (P_i, P_j, \text{Vote}^V(P_i|P_i, P_j), \text{Vote}^V(P_j|P_i, P_j))$$

where the prefix V means that the information is derived from votes, and $\text{Vote}^V(P_i|P_i, P_j)$ is the number of votes preferring P_i to P_j when comparing both of them.

Roughly speaking, $T_Relation$ focuses more on local comparisons because the limited descriptive words usually concentrate on a few products with some detailed feelings or judgments. $V_Relation$ can provide global information because the comparison is made among a large number of candidate products. The limitation of $T_Relation$ is partially caused by the fact that writing comments are very time-consuming. In contrast, it is much more convenient and simpler to click a button to vote for your favorite products.

The two sorts of comparative information, $T_Relation$, and $V_Relation$ can be combined to provide a more comprehensive description of the relationships between different products. The combined relationship is denoted as follows:

$$\text{Relation}_{ij} = (P_i, P_j, \text{Vote}(P_i|P_i, P_j), \text{Vote}(P_j|P_i, P_j))$$

where $\text{Vote}(P_i|P_i, P_j)$ is the full-scale information measuring the preference for P_i to P_j , calculated by

$$\text{Vote}(P_i|P_i, P_j) = \begin{cases} \frac{\text{Vote}^T(P_i|P_i, P_j) + \text{Vote}^V(P_i|P_i, P_j)}{\text{Vote}^T(P_i|P_i, P_j) + \text{Vote}^V(P_i|P_i, P_j) + \text{Vote}^T(P_j|P_i, P_j) + \text{Vote}^V(P_j|P_i, P_j)}, & i \neq j, \text{Vote}^T(P_i|P_i, P_j) \neq 0 \\ & \text{or } \text{Vote}^V(P_i|P_i, P_j) \neq 0 \text{ or } \text{Vote}^T(P_j|P_i, P_j) \neq 0 \text{ or } \text{Vote}^V(P_j|P_i, P_j) \neq 0 \\ \frac{1}{2}, & i \neq j, \text{Vote}^T(P_i|P_i, P_j) = 0, \text{Vote}^V(P_i|P_i, P_j) = 0, \text{Vote}^T(P_j|P_i, P_j) = 0, \text{Vote}^V(P_j|P_i, P_j) = 0 \\ 0, & i = j \end{cases} \quad (3.6)$$

Now, both descriptive information and comparative information are ready to be integrated by a directed and weighted graph structure. The graph structure is formally defined as a quadruple, $G = (V, E, W^V, W^E)$, where V is the set of vertices or nodes, E is the set of directed edges (i.e., the ordered pairs of vertices), W^V is the weight associated with each node, and W^E is the weight labeled on each directed edge. The weight for node P_i is normalized from $\text{Score}(P_i)$:

$$W^V(P_i) = \frac{\text{Score}(P_i)}{\sum_{k=1}^n \text{Score}(P_k)} \quad (3.7)$$

and the weight on a directed edge from P_j to P_i is normalized from $\text{Vote}(P_i|P_i, P_j)$:

$$W^E(P_i|P_i, P_j) = \frac{\text{Vote}(P_i|P_i, P_j)}{\sum_{l=1}^n \text{Vote}(P_i|P_i, P_j)} \quad (3.8)$$

where n is the total number of candidate products under comparison and P_l is the product that has a comparative relationship with product P_j .

Benefiting from the structure of the weighted digraph, a scalar overall eWOM score is developed with which to rank products. The overall eWOM of each product consists of two elements: inherent eWOM derived from its overall score of the product, and extrinsic eWOM accumulated from its comparative relationships with other products.

Roughly speaking, whether a product P_i should be ranked high is affected by three conditions:

- If a product has a high score from descriptive information (i.e., $\text{Score}(P_i)$ is high), it should be ranked high.
- If a product P_i has many votes from other products, P_i should be ranked high.
- If a product P_i has more votes than P_j whose ranking is high, P_i should be ranked even higher.

Based on the above analysis, the overall eWOM score of product P_i , $W_Score(P_i)$, can be calculated by

$$W_Score(P_i) = (1 - \beta)W^V(P_i) + \beta W_Score^C(P_i) \quad (3.9)$$

where $\beta \in (0, 1)$ is an adjustment factor to balance the effects of $W^V(P_i)$ and $W_Score^C(P_i)$, and $W_Score^C(P_i)$ can be calculated by

$$W_Score^C(P_i) = \sum_{j=1}^n W_Score(P_j)W^E(P_i|P_i, P_j) \quad (3.10)$$

where $W^V(P_i)$ measures the inherent eWOM, whereas $W_Score^C(P_i)$ denotes the extrinsic eWOM. By incorporating Eq. (3.10) into Eq. (3.9), the overall eWOM score $W_Score(P_i)$ can be derived as

$$W_Score(P_i) = (1 - \beta)W^V(P_i) + \beta \sum_{j=1}^n W_Score(P_j)W^E(P_i|P_i, P_j) \quad (3.11)$$

Ranking generation is a calculation of the eigenvector of the matrix to compute the $W_Score(P_i)$ value by using Eq. (3.11). The equation can be expressed as the following matrix function:

$$W_Score = (1 - \beta) \times W^V + \beta \times \mathbf{A} \times W_Score \quad (3.12)$$

where $W_Score = [W_Score(P_1), W_Score(P_2), \dots, W_Score(P_n)]^T$, $W^V = [W^V(P_1), W^V(P_2), \dots, W^V(P_i)]^T$, and \mathbf{A} is a $n \times n$ comparison relationship adjacent matrix:

$$\mathbf{A} = \begin{pmatrix} 0 & W^E(P_1|P_1, P_2) & \cdots & W^E(P_1|P_1, P_n) \\ W^E(P_2|P_2, P_1) & 0 & \cdots & W^E(P_2|P_2, P_n) \\ \vdots & \vdots & \ddots & \vdots \\ W^E(P_n|P_n, P_1) & W^E(P_n|P_n, P_2) & \cdots & 0 \end{pmatrix}.$$

Based on the presented method, rich and more experiments have been conducted on three types of products (mobile phones, laptops, and digital cameras) in Yang's work (Yang et al., 2016). The results demonstrate that with more information integrated, the ranking method can return better performance. In particular, comparative votes, which have attracted little attention in previous studies, contribute significantly to the ranking quality. An effective system is also demonstrated to help customers make informed choices when comparison shopping and assist manufacturers to maintain awareness of the exact positions of their products and to target implicit problems underlying the data.

3.3.2 Mining Relationships Between Retail Prices and Online Reviews

Marketing tools price is an important decision variable in marketing for a product and can affect customers' cognition, feelings, purchase decisions, and post-purchase satisfaction. Some research found that the price could affect consumer reviews. Online retailers are able to adjust their prices more frequently and easily compared to physical retail stores. A survey estimated that Amazon changes retail prices more than 2.5 million times daily for its millions of products.

In this vein, a fundamentally important question to ask is as follows: What effects can be observed regarding the volume and valence of consumer reviews after increasing or decreasing the retail price for a specific product? Here, volume measures the total amount of reviews posted on a product and is an important cue for product popularity. Valence captures the positive or negative nature of reviews, which contains evaluation information on product quality. To answer this question, computable models for describing relationships between prices and volume/valence of reviews should be built.

A GP (genetic programming) method is introduced to exploit functional relationships between retail prices and consumer reviews from a large and unique data set (Yang et al., 2021). In the experiment, a data set is obtained from an online retailer that comprises 321 types of products with retail prices and corresponding reviews. According to statistics, prices change 5431 times during the period of data collection, and 1,738,114 pieces of reviews are crawled in the same period. Experimental results show that for the relationships between retail prices and the volume of reviews, three types of models demonstrate the best performance: the linearly decreasing, asymmetric U-shaped, and asymmetric inverted U-shaped models. For the relationships between retail prices and the valence of reviews, the promising models are the linearly decreasing, asymmetric inverted U-shaped, and linearly increasing models.

Nevertheless, none of the models dominates all the others on the basis of three evaluation metrics: fitness, complexity, and coverage. For example, for the relationships between retail prices and the volume of reviews, the linearly decreasing models feature high coverage, low complexity, and low fitness, whereas the asymmetric U-shaped model features low coverage, high complexity, and high fitness. Instead of simply suggesting the model, comprehensive evaluations have been conducted to examine the performance of each candidate model in various categories of products to show its comparative advantages and disadvantages. The experimental results provide detailed references for the application of relationship models, such as which model is more suitable for a product or how to choose another model to complement this model when it does not model the relationship under a certain metric.

3.3.2.1 Model Building

To find a good model, a classic method is to perform the Generate/Test Cycle by designing alternatives and testing them against constraints (see Fig. 3.4). The traditional Generate/Test Cycle explicitly determines the models by human researchers on the basis of their hypothetical solution space, which becomes an obstacle when discovering the model from the solution space with sheer size. For example, in the relationship between price and reviews in this research, there are many potentially applicable candidate models. The models should be tested with

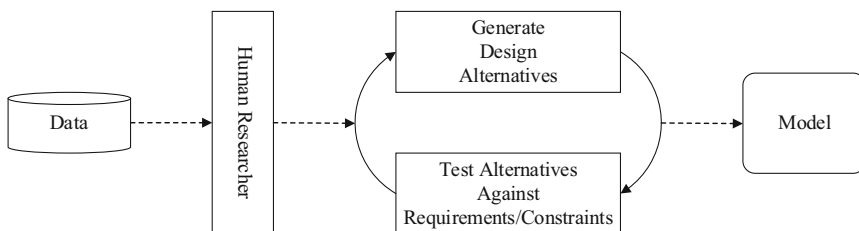


Fig. 3.4 Traditional generate/test cycle

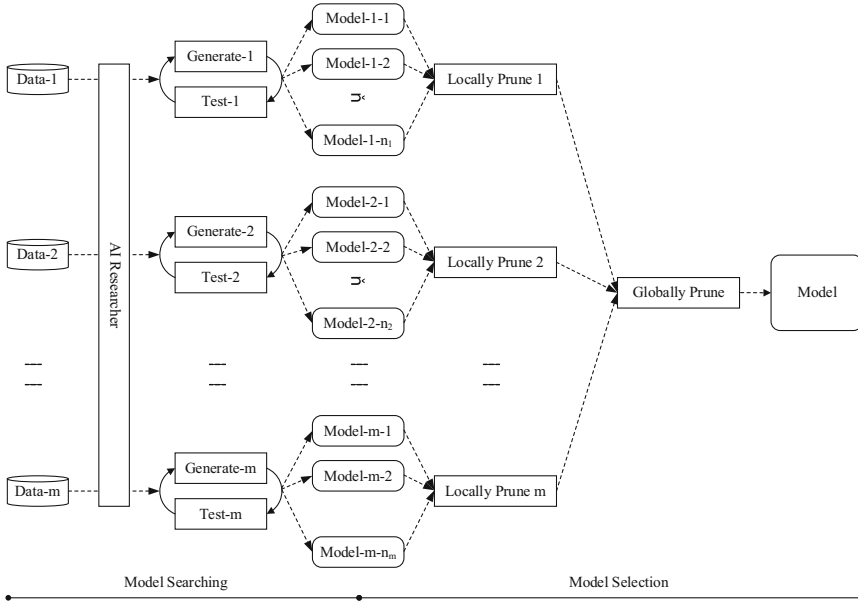


Fig. 3.5 New generate/test cycle

many products associated with frequent price changes and a huge amount of reviews on the selected retail website. Thus, generating and testing many candidate models by human researchers would be time-consuming.

To overcome the above obstacle, a new approach is proposed to the Generate/Test Cycle by incorporating an artificial intelligent (AI) researcher (see Fig. 3.5). Human experts do not have to perform the cycle to determine the proper models, and the AI researcher automatically suggests alternatives to describe the relationships hidden in data. The new approach has two processing stages: model searching and model selection. In the model searching stage, it generates and tests the model by using GP for each subset of data corresponding to one product. In the model selection stage, it first selects the Pareto optimal models for the given products at the individual level, namely, local pruning. Next, it discovers the models with high coverage for all types of products at the group level, namely, global pruning.

The intelligent data-driven generating/testing approach can search for promising models from an extremely large solution space by means of its two-phase operations. The core of the approach is evolutionary algorithms GP with easy transferability, which identifies meaningful analytical links and distills free-form models from data.

To reveal the relationships between price changes and the volume/valence of reviews, GP is adopted to automatically discover the mathematical model. The functional form expresses the nature of the relation, and the substantive meaning could be concretely made by some linear or nonlinear mathematical models. Following Bass's recommendation of simplicity to describe a pattern, the fundamental relationship in terms of reviews and price is defined as follows:

$$\text{REVIEWS} = f(\text{PRICE}) \quad (3.13)$$

Identifying fundamental relationships is a critical pursuit of research. Such a model is more likely to be generalizable and provides a starting point for further replication and extension of research.

Based on the basic model (3.13), the volume and valence of reviews have been seriously considered in the study, and their functional relationships with prices are represented by the following formulas (3.14) and (3.15) respectively.

$$R_{it}^{vo} = f_{it}^{vo}(P_{it}) \quad (3.14)$$

$$R_{it}^{va} = f_{it}^{va}(P_{it}) \quad (3.15)$$

where P_{it} denotes the price for a product i for t period; R_{it}^{vo} and R_{it}^{va} , respectively, denote the volume and the valence of reviews regarding the P_{it} .

3.3.2.2 Model Searching by Using GP

GP is the variant of the genetic algorithm with tree structure encoding and could be applied as a function discovery approach to analyze a multivariate dataset. GP explores the solution space by combining building blocks from a set of mathematical operators and operands (e.g., variables and constants) and searching the space of the mathematical expressions to find the model that best fits a given dataset.

In GP, a candidate solution is encoded as a tree structure. The flowchart of GP is shown in Fig. 3.6. There are several procedures. First, initialize the population; a set of primitive functional operators and variables is selected to integrate into the mathematical models to express the intrinsic relationship. The functional operators commonly used in relationship models include addition (+), subtraction (−), multiplication (×), exponential (*exp*), natural logarithm (*ln*), variable and constant. Second, calculate the fitness of the model until the terminal condition is satisfied. The fitting accuracy of the corresponding model is measured by the *R*-squared value. Last, the structures and parameters of models are evolved by genetic operators, such as reproduction, crossover, and mutation (see Fig. 3.7). Reproduction is used to select better individuals for the next generation directly. Crossover is used to exchange parts of two individuals and generate two new individuals. The mutation is used to alter a small portion of one individual randomly. Genetic operators generate new individuals. The configuration of the GP used for function discovery is in Table 3.3. The computational time in our research is 1000 s; at this point, the results have converged.

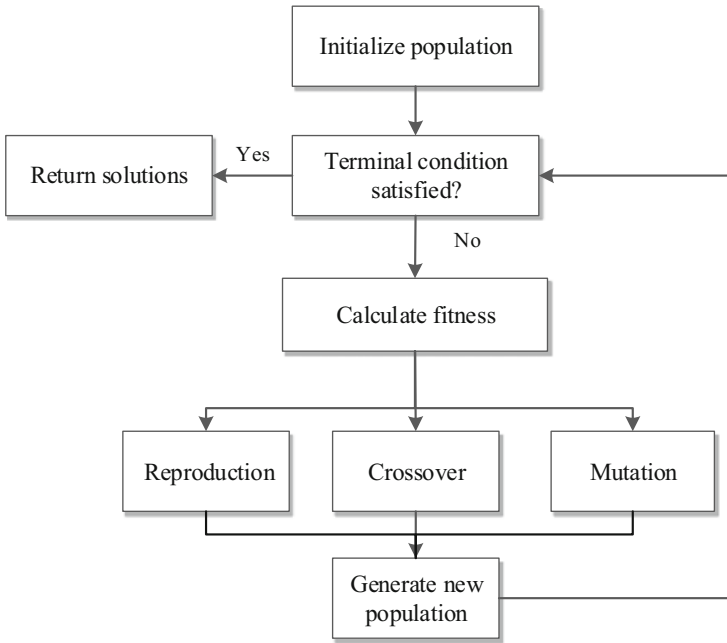


Fig. 3.6 Flowchart of GP

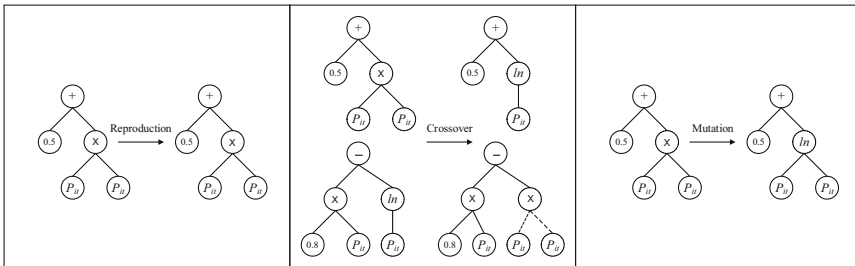


Fig. 3.7 Genetic operations in symbolic regression

Table 3.3 Components of function discovery using GP

Components	Values
Terminals	P_{it}
Functional operators	$+$, $-$, \times , \exp , \ln
Fitness function	R -squared
Genetic operators	Reproduction, crossover, mutation

3.3.2.3 Model Selection

Because GP returns a huge number of candidates, model selection plays a key role in pruning the less-promising candidates. Two steps for pruning are designed: local pruning and global pruning. The main principle of local pruning is to select the Pareto optimal models by considering their complexity and fitness based on Occam's razor, and global pruning considers the coverage of the selected models. After pruning, the models with low complexity, high fitness, and coverage remain for further analysis.

The local pruning is performed by that for a specific complexity level, and only the model with maximal fitness is selected:

$$\min(\text{erro}, C) = \min((1 - R^2), C) \quad (3.16)$$

Such a selection leads to a limited number of models for the tradeoff between error and complexity on a Pareto front.

Global pruning is performed to eliminate the less frequently appearing model when considering the whole data set, and each data set corresponds to one product in this study. After local pruning, all the Pareto optimal models can be collected, and the coverage value of each model structure can be counted. By ranking all the models by coverage, the top k models are selected for further analysis. The coverage of model i , denoted by Cove_i , indicates the proportion of the products that the corresponding model fits and is measured by

$$\text{Cove}_i = \frac{n_i}{m} \times 100\% \quad (3.17)$$

where m is the number of products, and n_i is the number of products that model i is selected for.

Based on the proposed approach, a considerable number of experiments on different types of products have been done in Yang's paper (Yang et al., 2021). From a unique dataset, various free-form relationship models with their own structures and parameters have been discovered. Through the comprehensive evaluations of candidate models, a guided map was offered to understand the relationship between dynamic retail prices and the volume/valence of reviews for different types of products. Practitioners could refer to the guide map and choose a proper response model for specific products according to the models provided by this research. If the practitioner wants to acquire more consumer reviews, it should be effective for most products by lowering the price. The experimental results also show that a higher price does not always lead to a decreased number of reviews. If a commercial practitioner wants to boost the average rating, he/she can adjust the price according to the suggestions in this study. For example, for high-involvement products, it is more effective to lower the price.

3.3.3 *Personalized Online Reviews Ranking Based on User Preference*

With the rapid development of the “customer first” service concept and big data technology, personalized services are flourishing day by day. In order to achieve personalized review ranking, consumer preference must be considered. The so-called consumer preference refers to the degree to which the individual prefers the product’s different features; apparently, consumers are willing to read objective reviews that concretely describe the features of their interests. Taking the hotel as an example, consumers may pay different degrees of attention to various features of the hotel (such as facility, service, location, etc.), and accordingly, a review ranking list matching their preferences is supposed to be shown. How can we measure the matching degree between consumer preference and a review ranking list? In which way can we obtain the ranking list with the maximal matching degree? This study tries to work out these issues.

In the study, a mechanism is designed at first to acquire consumer preferences (Luo & Wu, 2019). And then the matching degree between a review subset S and the consumer preference is defined as the product of the usefulness score of S and the cosine similarity between the feature distribution vector of S and the consumer preference vector. By taking account of potential consumers’ behaviors while reading reviews, a ranking list’s matching degree to the preference could be defined. Thus, the ranking issue is formulated as an optimization problem, whose objective is to maximize the expected matching degree. Due to the NP-hardness of the problem, using exact methods to search for the optimal ranking list is generally infeasible in practice. Hence, a heuristic algorithm for solving the consumer preference-based review ranking (CPRR) problem, denoted as $CPRR(\alpha)$, is proposed. The proposed algorithm selects reviews iteratively to add to the ranking list until an approximately optimal result is produced. In the experiment research, the data of all 79,781 pieces of reviews on hotels in Dalian, China, are collected from Meituan.com up to May 2018, and plenty of experimental results have witnessed the outperformance of the proposed method compared to the other baseline methods. In a word, the main contribution of this study is to present a review ranking algorithm based on consumer preference, which could provide consumers with personalized review ranking lists to support their purchasing decisions more effectively and efficiently.

To formulate the consumer preference-based review ranking (CPRR) problem, a research framework is designed as shown in Fig. 3.8. First, some concepts like the matching degree $M(P, S)$ between the given consumer preference P and a review subset S depends on the similarity $sim(P, FD_S)$ between P and S ’s feature distribution (FD_S) and S ’s usefulness (U_S) are defined. Next in the phase of problem modeling, by introducing the probability distribution of breaking positions Pro as the weighting factor, the expected matching degree $expM(P, L)$ between P and a ranking list L could be calculated. At last, the algorithm named $CPRR(\alpha)$ aiming to maximize the expected matching degree is conducted to obtain the final ranking list.

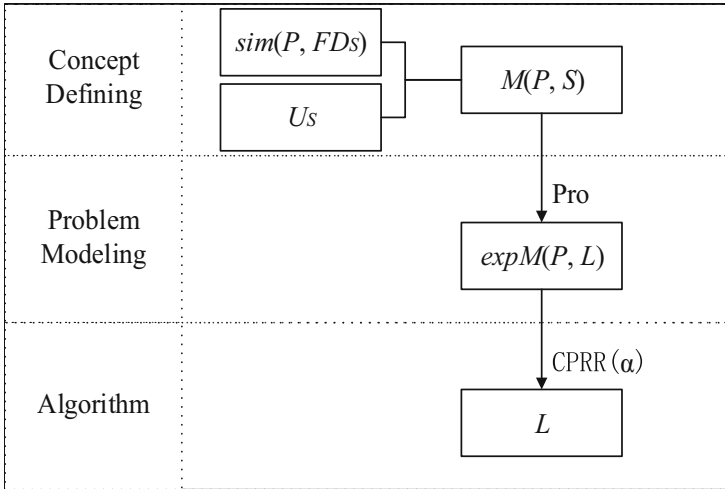


Fig. 3.8 Research framework

3.3.3.1 Concept Defining

As mentioned above, existing consumer preference mining methods always have the problem of cold start; hence a mechanism is designed to obtain consumer preferences directly. Suppose the given product has m features, the interest degree in each feature could be reflected on a five-point Likert-type scale (1 = strongly uninterested; 5 = strongly interested). While doing experiments, the consumer preferences could be generated randomly in the way, so that $5m$ preferences can be obtained as the input of the algorithm, which ensures the universality of the experimental results. When applied in real life or user study, consumers can select their preferences on a radio button group.

In order to measure the matching degree between reviews and consumer preferences for different features, it needs to deeply mine the distribution of features in reviews and represent each review as a feature distribution vector. Given a category of products and their relevant reviews, domain feature dictionary $F = \{f_1, f_2, \dots, f_m\}$ can be built using the available feature extraction method, where f_i denotes a type of feature words including several nouns or noun phrases with similar or related meanings. It is worth mentioning that the feature dictionary could be more accurate by mining as many reviews as possible. Therefore, given a product, its set of reviews $R = \{r_1, r_2, \dots, r_m\}$ and corresponding domain feature dictionary $F = \{f_1, f_2, \dots, f_m\}$, a review $r, r \in R$, can be denoted as a feature distribution vector $r = (r^{f_1}, r^{f_2}, \dots, r^{f_m})$, where r^{f_i} represents the number of feature words in r belonging to f_i .

In reality, consumers are concerned about not only whether a review mentions the feature they are interested in, but also whether the description of the feature in the review is informative and objective. Therefore, in addition to feature distribution, the usefulness of a review should also be considered as a weight coefficient reflecting its

relative significance in matching degree. Review usefulness is online consumers' subjective perception of whether the review information published by previous reviewers is helpful for them to make purchase decisions, which could be influenced by many factors, such as ratings, sentiments of reviews, reviewers' member levels, etc. To measure it, a family of functions could be applied to map a specific review r to a real-numbered usefulness score U_r in the interval $[0, 1]$. Through model training and predicting using some machine learning methods, such as linear regression, U_r could be derived easily.

Given a product and its set of reviews $R = \{r_1, r_2, \dots, r_n\}$, where $r_i = (r_i^{f_1}, r_i^{f_2}, \dots, r_i^{f_m})$, for a set of reviews $S \subseteq R$, the feature distribution vector of S can be denoted as

$$FD_S = \sum_{r \in S} r = \left(\sum_{r \in S} r^{f_1}, \sum_{r \in S} r^{f_2}, \dots, \sum_{r \in S} r^{f_m} \right) \triangleq (S^{f_1}, S^{f_2}, \dots, S^{f_m}) \quad (3.18)$$

Given a consumer preference vector $P = (P_{f_1}, P_{f_2}, \dots, P_{f_m})$, where P_{f_i} denotes the consumer's preference degree towards feature f_i , the cosine similarity between P and FD_S in an m -dimensional vector space can be calculated as follows. It can be seen from the formula that whether to normalize P and FD_S or not does not change the result, so for the accuracy of calculation, normalization is not necessary here.

$$\cos(P, FD_S) = \frac{P \cdot FD_S}{\|P\| \times \|FD_S\|} = \frac{P_{f_1} \times S^{f_1} + P_{f_2} \times S^{f_2} + \dots + P_{f_m} \times S^{f_m}}{\sqrt{P_{f_1}^2 + P_{f_2}^2 + \dots + P_{f_m}^2} \times \sqrt{S^{f_1^2} + S^{f_2^2} + \dots + S^{f_m^2}}} \quad (3.19)$$

Thus, the matching degree between P and S can be defined as

$$M(P, S) = U_S \times \cos(P, FD_S) \quad (3.20)$$

where U_S is a weight coefficient in the interval $[0, 1]$, denoting the average usefulness score of reviews in S .

3.3.3.2 Problem Modeling

Given a review ranking list $L = (r_{l_1}, r_{l_2}, \dots, r_{l_n})$ for the given product, where r_{l_i} denotes the i th review in the list, consumers always read the reviews in sequence and may break at any position. If the consumer stops after reading the i th review, the set of reviews he has read consists of the top i reviews, denoted as $S_i = \{r_{l_1}, r_{l_2}, \dots, r_{l_i}\}$. The information that he obtains from S_i can match his preference P in a degree, which is denoted as $M(P, S_i)$, expressing the matching degree between P and S_i .

In reality, the number of reviews that a consumer will read is related to his behavior habit, the task at hand, and other environmental factors. The probability distribution of breaking positions $Pro = (p_1, p_2, \dots, p_n)$, where p_i denotes the

probability that a consumer stops after reading the i th review, could be observed by eye-tracking, log file analysis, and other various technologies.

Therefore, all cases should be considered while forming a review ranking list L , in which any subset of the list, i.e., S_i , $i = 1, 2, \dots, n$, may be read by consumers. Thus, the expected matching degree between P and L can be calculated as

$$\exp M(P, L) = \sum_{i=1}^n p_i M(P, S_i) \quad (3.21)$$

Then the consumer preference-based review ranking problem could be formally defined as follows:

Problem: The consumer preference-based review ranking (CPRR) problem. Given a consumer preference P and an original set of reviews R for a product, rank all these reviews to form a ranking list L such that the expected matching degree between P and L , i.e., $\exp M(P, L)$, is maximized.

According to Formula (3.20), the matching degree between consumer preference P and S_i (i.e., the set of the first i reviews in the ranking list L , $i = 1, 2, \dots, n$) could be calculated. Hence, the CPRR problem can be mathematically formulated as

$$\begin{aligned} \max \exp M(P, L) &= \sum_{i=1}^n p_i M(P, S_i) = \sum_{i=1}^n p_i U_{S_i} \cos(P, FD_{S_i}) \\ s.t. S_i &= \{r_{l_1}, r_{l_2}, \dots, r_{l_i}\}, \quad i = 1, 2, \dots, n \end{aligned} \quad (3.22)$$

3.3.3.3 Algorithm

The CPRR problem is NP-hard. Since a well-known NP-hard problem, the maximum coverage problem is reducible to it. For example, 24(4!) review ranking lists can be produced among 4 reviews; in this case, it is necessary to calculate their $\exp M(P, L)$ respectively and select the one with maximal $\exp M(P, L)$ as the resultant ranking list. While the amount of reviews is huge, it cannot be solved using the exact enumeration method in polynomial time; thus approximate methods should be considered. Intuitively, the expected matching degree between consumer preference and the ranking list could be maximized stepwise and iteratively, that is, certain reviews that perform well on the current $\exp M(P, L)$ should be preserved in each iteration until all reviews have been added to a ranking list. Based on this heuristic idea, an approximation algorithm named CPRR(α) is proposed, where α is a parameter controlling the accuracy of the algorithm.

The input of the algorithm CPRR(α) includes the given consumer preference $P = (P_{f_1}, P_{f_2}, \dots, P_{f_m})$, the review set $R = \{r_1, r_2, \dots, r_n\}$ with each review structured as a feature distribution vector $r = (r^{f_1}, r^{f_2}, \dots, r^{f_m})$ and a usefulness score U_r , the probability distribution $\text{Pro} = (p_1, p_2, \dots, p_n)$ where p_i denotes the probability that a consumer stops reading at the i th review, and a controlling variable α ($\alpha \in [0, 1]$) to

help to control the number of candidate lists in each iteration. At the beginning of the algorithm, an empty ranking list L_0 and its corresponding set that only contains the empty list SL_0 are initialized. In the i th iteration, each list L_{i-1} belonging to the list set SL_{i-1} preserved at the previous iteration is extended with a new review to generate possible list set SL_i . Later, the maximal and minimal expected matching degree values of the lists in the list set SL_i are calculated and denoted as maxValue and minValue respectively. The lists with $\text{expM}(P, L_i)$ greater than $\text{maxValue} - (1 - \alpha)$ ($\text{maxValue} - \text{minValue}$) $(\sum_{j=i}^n p_j)$ are preserved as the lists for the next iteration, where

the sum of probabilities for the remaining positions $(\sum_{j=i}^n p_j)$ is multiplied to $1 - \alpha$, for the purpose of further shrinking the number of candidate lists in each iteration. After n iterations, the list in SL_n with the maximal expected matching degree is the resultant ranking list L .

In summary, due to the information overload of reviews and the prevalence of personalized services, the consumer preference-based review ranking (CPRR) problem is raised. The goal of the CPRR problem is to provide a review ranking list L to match with the consumer preference P , which is formulated as maximizing the expected matching degree between P and L . Because of the limitations of the exact solutions in practice, an approximate optimization algorithm named CPRR (α) has been proposed to achieve a ranking list in an efficient manner. Specifically, the parameter α is used to determine the value range of the expected matching degree for controlling the amount of candidate ranking lists in each iteration. After n iterations, the list with the maximal expected matching degree is finally chosen as the resultant ranking list. Furthermore, CPRR(α) has been evaluated with intensive experiments on real data from Meituan.com, whose results demonstrate its sensitivity to different consumer preferences and its advantage compared with other algorithms of concern, especially the default ranking on the website.

3.4 Expected Future of Techniques for Online Reviews

Based on the literature review of online reviews' mining techniques and their commercial value discovery, in this section, we focus on discussing the expected future techniques for online reviews. The first potential topic is deep migration learning within the information extraction field. Subsequently, the processing of multimodal data is also discussed regarding its valuable directions. Then the expected research topics about text categorization are summarized.

3.4.1 *Deep Migration Learning*

Currently, deep learning is extensively applied in massive NLP tasks, due to the development of neural networks and artificial intelligence. Common techniques include sentiment analysis, relationship extraction, event extraction, etc.; and especially, these advanced methods are investigated in numerous research fields such as tourism, manufacturing, and the hospitality industry using online customer reviews. However, such deep learning-based methods mandatorily require labeled training data, as well as a pretty long training time. In the current e-commerce context, there are an increasing number of unstructured consumer online reviews. How to deal with the low-quality and unlabeled data due to a large amount of noise is one of the main challenges, which has reduced the effectiveness and performance of deep learning and negatively affected a variety of NLP tasks, such as semantic association recognition between entities and terms, extraction of contextually relevant information, data modeling, and structuring of data.

Fortunately, deep migration learning may address the above issues well. Scholars of computer vision first proposed and applied migration learning to cope with the unlabeled data issue (Fang & Tao, 2019), and their results demonstrated desirable accuracy. Recently, some researchers of NLP have also noticed the advantages of migration learning. For example, using online patient reviews, Xie and Xiang (2022) effectively identified the discussed topics. Encouragingly, these studies provide preliminary evidence of the effectiveness of applying transfer learning to online reviews. Overall, there might be several aspects worth exploring for future research as follows.

As we mentioned about NER, supervised NER systems rely on a large amount of annotated data, and the annotation of online review data is a time-consuming and expensive task with quality and consistency issues. In addition, entity nesting is widespread. Therefore, there is a need to develop generic annotation schemes to handle nested entities and fine-grained entities.

Another notable problem to be solved is the migration learning-based relationship extraction. There are many datasets available for relationship extraction, but almost no problem-specific datasets. For example, cross-sentence relationship extraction, where two different entities are mentioned in two different sentences. There are no such datasets available for researchers to analyze.

Besides, the event extraction task is complex and deep learning-based event extraction models can bring better results, but these methods require a large amount of annotated data. Currently, the event extraction task has only a relatively small amount of annotated data, and manual annotation is very costly and time-consuming. Therefore, building an automated approach to annotating event extraction data is the future direction of development.

3.4.2 *Multimodal Data Processing*

Motivated by the multimodal way of thinking of humans, more and more consumers are posting their products using experiences in complex and diverse forms on online platforms, including text, images, and even videos. Especially, many e-commerce platforms officially encourage consumers' such behaviors, in order to present their products to potential customers more comprehensively. As stated above, existing studies on text reviews have been very extensive, involving multiple mature text mining techniques such as sentiment analysis, opinion mining, topic discovery, etc. On the other hand, numerous studies have investigated the application and techniques of deep learning in image and video processing. Regarding the exploration of multimodal data in online reviews, however, it is still in its infancy, despite many studies emphasizing its importance to e-commerce and consumers. Overall, there might be several future directions to expand the current literature systems.

The first potential area is analyzing multimodal data with text, images, or videos. By fusing or synthesizing different types of data, sentiments, opinions, or emotions in online reviews could be extracted. A simple way to achieve these tasks is separately processing the content in text, images, or videos, and finally fusing the extracted information to output. Despite its simplicity, however, such a method may lose valuable information in online reviews since it ignores the inside associations that existed between different modal data. Therefore, a more robust way is to transform the input features with different forms into an effective vector representation, and then data mining techniques could be employed to extract useful information.

The second valuable research topic might be the handling of class imbalance issues of multimodal data. Typically, multimodal data may face serious data imbalance (skewed) problems. Regarding text mining, there are several mature and standard steps to preprocess the review texts, such as removing stop words and stemming words. However, little is known about the prior knowledge for images or videos to preprocess, and as a result, much noise would be contained in the dataset.

Another issue worth exploring is the attribute space downscaling. Generally, online reviews are pretty massive; meanwhile, deep learning and neural network are also characterized by high dimensionality. In this regard, the application and processing of multimodal data shall require adequate computing resources and equipment. Therefore, to effectively explore the valuable information in such multimodal data, it is urgent to seek out ways of reducing the high dimensionality. In doing so, the subsequent analysis would be more efficient.

3.4.3 *Text Categorization Tasks*

In the past few years, the development of deep learning and artificial intelligence techniques has led to the rapid development of text categorization. Many novel ideas

and models have been proposed by different scholars, such as the Attention Mechanism, Transformer, Bert, etc. Despite the great progress that has been made, the current text categorization task still faces some difficulties and challenges that need to be continued to be explored by later generations. We believe that the solution to these problems will further promote the rapid development of text categorization tasks, and even natural language processing tasks.

First, *the interpretability of deep learning models in text categorization tasks*. Although a variety of deep learning models have achieved good results in different datasets, the interpretability of the parameters and results of these models is still highly problematic. On the one hand, some models perform better on dataset A and worse on dataset B. Why does this result occur? Is it because of the variability of the models or the variability of the datasets? Moreover, what is the meaning of the parameters of the trained deep learning models? Is there any correlation with the characteristics of the data? In addition, is the larger the structure of the deep learning model, the better? If not, how large should the model structure be set in order to achieve similar classification results? While attention mechanisms have provided some insights to answer these questions, detailed studies of the underlying behavior and dynamics of these models are still lacking. A deeper understanding of these questions could help researchers better design and build deep learning-based text categorization methods.

Second, *the approach of data annotation*. While plenty of large-scale datasets have been collected for common text categorization tasks in recent years, new datasets are still needed for more challenging tasks, such as QA for multi-step inference and text categorization of multilingual documents. Having large-scale labeled datasets for these tasks can help accelerate progress in these areas. The annotation of large-scale data is not very realistic to be achieved manually, so there is a need to promote further development of automatic data annotation techniques.

In addition, *small sample learning techniques*. Most deep learning models are supervised models, which require a mass of domain labels. When the data is relatively small, the effectiveness of deep learning-based text categorization methods is limited. In fact, it is expensive to collect such labeled data for each new domain. Small-sample learning technique is a deep learning technique that can be used empirically on a specific task to classify text using a small amount of labeled data. This problem is not contradictory to the previous one but is an attempt to solve the “big data dependency” problem of deep learning from two perspectives. A solution to either of these two problems will advance the development of deep learning-based text categorization methods.

Finally, *incorporating prior knowledge into deep learning algorithms*. Integrating a priori common-sense knowledge into deep learning models has the potential to significantly improve model performance, just as humans use common-sense knowledge to perform different tasks. For example, QA systems equipped with a common-sense knowledge base can answer questions about the real world. In the absence of information, common sense can also help solve problems. Using people’s general knowledge of everyday objects or concepts, AI systems can reason like humans, based on “default” assumptions about the unknown.

References

- Abrahams, A. S., Jiao, J., Wang, G. A., et al. (2012). Vehicle defect discovery from social media. *Decision Support Systems*, 54(1), 87–97.
- Bi, J. W., Han, T. Y., Yao, Y., & Li, H. (2022). Ranking hotels through multi-dimensional hotel information: A method considering travelers' preferences and expectations. *Information Technology & Tourism*, 24(1), 127–155.
- Cao, Q., Duan, W. J., & Gan, Q. W. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511–521.
- Che, N., Chen, D., & Le, J. (2019). Entity recognition approach of clinical documents based on self-training framework. In *Recent developments in intelligent computing, communication and devices* (pp. 259–265). Springer.
- Choi, H. S., & Leon, S. (2020). An empirical investigation of online review helpfulness: A big data perspective. *Decision Support Systems*, 139, 113403.
- El Mrabti, S., Al Achhab, M., Lazaar, M., et al. (2018). Comparison of feature selection methods for sentiment analysis. In Y. Tabii et al. (Eds.), *Big Data, cloud and applications. BDCA 2018. Communications in computer and information science* (Vol. 872, pp. 261–272). Springer.
- Fang, X., Tao, J. (2019). A transfer learning based approach for aspect based sentiment analysis. *Proceedings of the 6th international conference on social networks analysis, management and security (SNAMS)* (pp. 478–483).
- fasttext. (2016). <https://fasttext.cc/docs/en/support.html>
- Fu, X. L., Ouyang, T. X., Yang, Z. L., et al. (2020). A product ranking method combining the features–opinion pairs mining and interval-valued Pythagorean fuzzy sets. *Applied Soft Computing*, 97(3), 106803.
- Gao, S., Tang, O., Wang, H. W., et al. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *International Journal of Hospitality Management*, 71, 19–32.
- Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3), 493–520.
- Glove. (2014). <https://nlp.stanford.edu/projects/glove>
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.
- Imtiaz, M. N., & Islam, M. K. B. (2020). Identifying significance of product features on customer satisfaction recognizing public sentiment polarity: Analysis of smart phone industry using machine-learning approaches. *Applied Artificial Intelligence*, 33(11), 832–848.
- Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian networks and decision graphs*. Springer.
- Jiang, H., Kwong, C. K., Park, W. Y., et al. (2018). A multi-objective PSO approach of mining association rules for affective design based on online customer reviews. *Journal of Engineering Design*, 29(7), 381–403.
- Kim, J., Lee, M., Kwon, W., et al. (2022). Why am I satisfied? See my reviews - price and location matter in the restaurant industry. *International Journal of Hospitality Management*, 101, 103111.
- Lan, M., Tan, C. L., Su, J., et al. (2009). Supervised and traditional term weighting methods for automatic text categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4), 721–735.
- Liu, T. Y., Fang, S. S., Zhao, Y. H., et al. (2015). Implementation of training convolutional neural networks. arXiv preprint arXiv:1506.01195.
- Liu, W., Wang, Q., Zhu, Y. L., et al. (2020). GRU: Optimization of NPI performance. *The Journal of Supercomputing*, 76(5), 3542–3554.
- Liu, X., Huang, H. Y., & Zhang, Y. (2019). Open domain event extraction using neural latent variable models. *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 2860–2871).

- Liu, X. M., Wang, G. A., Fan, W. G., et al. (2020). Finding useful solutions in online knowledge communities: A theory-driven design and multilevel analysis. *Information Systems Research*, 31(3), 731–752.
- Liu, Y., Jiang, C. Q., & Zhao, H. M. (2019). Assessing product competitive advantages from the perspective of customers by mining user-generated content on social media. *Decision Support Systems*, 123(7), 113079.
- Liu, Y. Z., Song, Y. X., Sun, J. S., et al. (2020). Understanding the relationship between food experiential quality and customer dining satisfaction: A perspective on negative bias. *International Journal of Hospitality Management*, 87, 102381.
- Liu, Z., Qin, C. X., & Zhang, Y. J. (2021). Mining product competitiveness by fusing multisource online information. *Decision Support Systems*, 143(5), 113477.
- Luo, D., & Wu, J. N. (2019). Ranking online reviews based on consumer preferences. *Proceedings of 2019 IEEE international conference on software quality, reliability, and security companion* (pp. 342–347).
- Luo, F., Nagesh, A., Sharp, R., et al. (2019). Semi-supervised teacher-student architecture for relation extraction. *Proceedings of the third workshop on structured prediction for NLP* (pp. 29–37).
- Mikolov, T., Karafiát, M., Burget, L., et al. (2010). Recurrent neural network based language model. *Proceedings of Interspeech*, 2(3), 1045–1048.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106.
- Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1), 3–26.
- Park, S., Lee, J. S., & Nicolau, J. L. (2020). Understanding the dynamics of the quality of airline service attributes: Satisfiers and dissatisfiers. *Tourism Management*, 81, 104163.
- Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce*, 7(3), 101–134.
- Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3), e1301.
- Rafeek, R., & Remya, R. (2017). Detecting contextual word polarity using aspect based sentiment analysis and logistic regression. *2017 IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM)* (pp. 102–107).
- Rianthong, N., Dumrong Siri, A., & Kohda, Y. (2016). Improving the multidimensional sequencing of hotel rooms on an online travel agency web site. *Electronic Commerce Research and Applications*, 17, 74–86.
- Rong, X. (2014). Word2vec parameter learning explained. eprint arXiv:1411.2738.
- Sheng, J., Guo, S., Yu, B., et al. (2021). CasEE: A joint learning framework with cascade decoding for overlapping event extraction. *Findings of the association for computational linguistics* (pp. 164–174).
- Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404, 132306.
- Siering, M., & Muntermann, J. (2013). What drives the helpfulness of online product reviews? From stars to facts and emotions. *Wirtschaftsinformatik*, 7, 103–118.
- Srivastava, D. K., & Bhambhu, L. (2010). Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*, 12(1), 1–7.
- Tayal, D. K., Yadav, S. K., & Arora, D. (2022). Personalized ranking of products using aspect-based sentiment analysis and Plithogenic sets. *Multimedia Tools and Applications*, 82, 1261–1287. <https://doi.org/10.1007/s11042-022-13315-y>
- Vashishtha, S., & Susan, S. (2021). Highlighting keyphrases using senti-scoring and fuzzy entropy for unsupervised sentiment analysis. *Expert Systems with Applications*, 169, 114323.

- Vu, H. Q., Li, G., Law, R., et al. (2019). Exploring tourist dining preferences based on restaurant reviews. *Journal of Travel Research*, 58(1), 149–167.
- Wang, H. W., Gao, S., Yin, P., et al. (2017). Competitiveness analysis through comparative relation mining: Evidence from restaurants' online reviews. *Industrial Management & Data Systems*, 117(4), 672–687.
- Xiao, S. S., Wei, C. P., & Dong, M. (2016). Crowd intelligence: Analyzing online product reviews for preference measurement. *Information & Management*, 53(2), 169–182.
- Xie, Y. T., & Xiang, F. (2022). An improved approach based on dynamic mixed sampling and transfer learning for topic recognition: A case study on online patient reviews. *Online Information Review*, 46(6), 1017–1033.
- Xu, G. X., Meng, Y. T., Qiu, X. Y., et al. (2019). Sentiment analysis of comment texts based on BiLSTM. *IEEE Access*, 7, 51522–51532.
- Yang, X., Yang, G. F., & Wu, J. N. (2016). Integrating rich and heterogeneous information to design a ranking system for multiple products. *Decision Support Systems*, 84, 117–133.
- Yang, X., Yang, G. F., Wu, J. N., et al. (2021). Modeling relationships between retail prices and consumer reviews: A machine discovery approach and comprehensive evaluations. *Decision Support Systems*, 145(3), 113536.
- Yang, Z. G., & Sudharshan, D. (2019). Examining multi-category cross purchases models with increasing dataset scale – An artificial neural network approach. *Expert Systems with Applications*, 120, 310–318.
- Yuan, Y., Zhou, X., Pan, S., et al. (2020). A relation-specific attention network for joint entity and relation extraction. *Proceedings of the 29th international joint conference on artificial intelligence (IJCAI-20)* (pp. 4054–4060).
- Zadeh, L. A. (1965). Fuzzy sets. *Information Controls*, 8(3), 338–353.
- Zeng, X. R., Zeng, D. J., He, S. Z., & et al. (2018). Extracting relational facts by an end-to-end neural model with copy mechanism. *Proceedings of the 56th annual meeting of the association for computational linguistics* (pp. 506–514).
- Zhang, J., Lu, X. C., & Liu, D. (2021). Deriving customer preferences for hotels based on aspect-level sentiment analysis of online reviews. *Electronic Commerce Research and Applications*, 49, 101094.
- Zhang, Y. H., Qi, P., & Manning, C. D. (2018). Graph convolution over pruned dependency trees improves relation extraction. *Proceedings of the 2018 conference on empirical methods in natural language processing (EMNLP)* (pp. 2205–2215).
- Zhao, Y., Wan, H. Y., Gao, J. W., et al. (2019). Improving relation classification by entity pair graph. *Proceedings of the 11th Asian conference on machine learning research (PMLR)* (pp. 1156–1171).
- Zheng, S. C., Wang, F., Bao, H. Y., et al. (2017). Joint extraction of entities and relations based on a novel tagging scheme. *Proceedings of the 55th annual meeting of the association for computational linguistics* (pp. 1227–1236).