Information Processing and Management 59 (2022) 102784

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

E-word of mouth sentiment analysis for user behavior studies

Hui Li [a,1], Qi Chen [b,1], Zhaoman Zhong [a,*], Rongrong Gong [a,*], Guokai Han [a]

[a] College of Computer, Jiangsu Ocean University, Lianyungang, China
[b] School of Business Administration, Dalian University of Technology, China

ARTICLE INFO

Keywords:
sentiment analysis
electronic word-of-mouth
online dating service
machine learning

ABSTRACT

Nowadays, online word-of-mouth has an increasing impact on people's views and decisions, which has attracted many people's attention.The classification and sentiment analyse in online consumer reviews have attracted significant research concerns. In this thesis, we propose and implement a new method to study the extraction and classification of online dating services (ODS)'s comments. Different from traditional emotional analysis which mainly focuses on product attribution, we attempted to infer and extract the emotion concept of each emotional reviews by introducing social cognitive theory. In this study, we selected 4,300 comments with extremely negative/positive emotions published on dating websites as a sample, and used three machine learning algorithms to analyze emotions. When testing and comparing the efficiency of user's behavior research, we use various sentiment analysis, machine learning techniques and dictionary-based sentiment analysis. We found that the combination of machine learning and lexicon-based method can achieve higher accuracy than any type of sentiment analysis. This research will provide a new perspective for the task of user behavior.

1. Introduction

As global Internet presentations continue to increase, the number of consumers who provide online comments have increased significantly (Lu & Bai, 2021). If exploited properly, abundant data should produce useful insights. One insight that can be obtained from the statistics is the information of electronic word of mouth (EWOM). EWOM is known for its significant impact on consumer behavior (Tobon & García-Madariaga, 2021). EWOM communication framework demostrates the direct relation of adopting EWOM and consumers' willingness to purchase. EWOM can provide objective information for more and more consumers who trust these communications (Yaniv & Shalom, 2021).Comment mining concerning sentiment analysis is considered to be a suite of proceedings for identifying sentiments, opinions and author's attitudes in texts, transforming them into meaningful information and using them to make business decisions (Siddiqui et al., 2021).

Sentiment classification identifies opinions and arguments in a given text, and it is part of opinion mining. It tries to find statements of agreement or disagreement in comments or reviews that involve positive, negative or neutral statements. Sentiment analysis has attracted widespread attention and has been widely used in many fields (Wang & Zhang, 2020). Up to now, many approaches of sentiment analysis have been proposed, which can be roughly split into document-level, sentence-level and aspect-level(Jiang, Chan, Eichelberger, Ma, & Pikkemaat, 2021).Most of the work of sentiment analysis can be achieved by assessing the document's polarity.

* Corresponding Author
E-mail addresses: 15340368@qq.com (Z. Zhong), 1037252341@qq.com (R. Gong).
[1] Hui Li and QI Chen are co-first authors of this article

https://doi.org/10.1016/j.ipm.2021.102784
Received 23 June 2021; Received in revised form 1 September 2021; Accepted 30 September 2021
Available online 18 October 2021
0306-4573/

Phrase and sentence levels have become common increasingly in recent years.Dictionary-based and machine learning approaches are two of the most common uses of emotion analysis(Ahlem & Khalil, 2020). Building emotional dictionary is the main way to realize sentiment analysis. For example,SentiWordNet (Madani, Erritali, Bengourram, & Sailhan, 2020), SenticNet (Hung, Wu, & Chou, 2020), and OpinionFinder (Wiebe et al., 2005) are famous emotional dictionaries. Sentiment lexicon plays a vital role, because it includes the emotional polarity of a word or phrase which can be used to classify sentiment of textual data.

Most of the available studies is based on sentiment analysis of product reviews. However in this thesis, we focus on discussing the domain of online dating service's (ODS) reviews,because this domain is different from the other product reviews (Wang, 2020). When a person writes down her/his feelings about the usage of the ODS, she/he probably comments not only on website elements (e.g., easy to use, perfect function) but also on the people who have used the website (e.g., dater, matchmaker, friend, parents). Therefore, the review features of the ODS are more complex than the products' reviews, and it's more challenging to analyze the ODS's reviews. This paper describes a method used natural language processing(NLP) and emotion lexicon generation (Chang, Hwang, & Wu, 2021) to classify the ODS's reviews (이득영, 장선우, & 전한종, 2019).

Many opinion mining studies have been conducted to analyze unstructured text data including customer reviews, blogs, tweets, news and online (Al-Mashhadany, Hamood, Al-Obaidi, & Al-Mashhsdany, 2021).However, there are still some problems to be solved. First, there is less research on emotion analysis based on the theory of user's behaviour analysis. Different from traditional sentiment analysis which is based on the products in the customer reviews, the sentiment analysis method based on user behavior analysis is to extract the user's concept and analyze the emotion from the user's point of view.Next, there were few sentiment analysis studies based on multiple category problem. Traditional sentiment analysis is generally based on document-level (Li, H, Kang, Yang, & Zong, 2019), sentence-level (Eng, Nawab, & Shahiduzzaman, 2021) and aspect-level (Lu, Zhu, Zhang, Wu, & Guo, 2020),but this paper is based on the concept-level and multi-sentiment classification problem. The main problem is the extraction of concepts and the classification of multiple emotions (Sznycer & Lukaszewski, 2019).

Our work tries to perform sentiment analysis on users' behaviors influenced by conflict environmental factors obtained by using machine learning models.We selected 4800 controversial context of online dating service (ODS) which had extremely negative/positive rates of popularity in a online dating sites, extracted linguistic features from them using NLP,and compared their performances within accuracy using machine learning models: Naive Bayes (NB), K-Nearest-Neighbor (KNN) and Support Vector Machines (SVM). To sum up, our contributions in this study are as follows:(1) A new method of learning training data in the form of classifier labels is proposed. By combining machine learning with dictionary methods, the automatically generated questionnaire has more value and credibility.(2) The opinion mining of comments and the analysis of sentiment classification are based on user behavior, rather than the analysis method based on product features in traditional research. It provides a new direction for sentiment analysis in the future by exploring opinions and classifying emotions from a new perspective.(3) The automatic lexico-based classifier is used to classify user behaviors according to eight features of social cognitive theory as influencing factors. Four classifiers were constructed according to the eight features, and a new constructed dictionary was used to correct the false prediction results of the four classifiers, and vocabulary and learning methods were fused to make up for the defects of the two.

This is followed by a discussion of 4 areas: Section 2 describes the work related to EWOM, opinion mining and emotion analysis. Section 3 develops the narrative of our approach. Section 4 report on the case study of SOD. In section 5, experiment results of efficiency,coverage and discussion are provided. Conclusions and further tasks are presented in section 6.

2. Related work

2.1. Electronic word of mouth

Electronic word-of-mouth (EWOM) is defined as an online sharing campaign that includes a wealth of consumer information from experienced consumers' opinions and recommendations on vendors/products (Donthu, Kumar, Pandey, Pandey, & Mishra, 2021). EWOM is becoming an essential part of the online experience both for marketers and customers. EWOM can greatly attract people's attention and raise the topic of discussion. In addition, EWOM is widely used by online visitors as a reference (Wu, Song, Duan, Hong, & Sui, 2021). Many marketers have difficulty in identifying relevant and valuable EWOMs (Liu, Jayawardhena, Osburg, Yoganathan, & Cartwright, 2021). This might be a result of the conceptual inexpedience of sentiment extraction for identifying relevant EWOM.The progress in calculation methods and natural linguistic processing has provided an opportunity to address EWOM's challenges.

Many surveys have investigated the influence of EWOM on sales and marketing.Traditionally, researchers have focused on the meta data (e.g., number of reviews, date of posting) of online reviews,and how it relates to business performance (e.g., future sales) (Gao, 2018).However, the wealth of information embedded in the textual content of online reviews provides the possibility of understanding the dynamics of purchasing decisions. In reality, customers may rely on online reviews to make their decisions. Therefore, understanding the content of online reviews can provide value for understanding/predicting different characteristics of businesses (e.g., consumer perceptions). Many researches show that EWOM affects consumer behaviour and the product sales (Kim, 2019). They found that the properties of online feedback (for example,a wide range of online feedback, the ratings and emotions shared in the comments) affect the sale of the product.Some studies have succeeded in understanding the association between online comments and product sales by combining the Bass model with the predictive approach of EWOM ratings. (Wu et al., 2015). As discussed by Lau et al. (2018), Mining from EWOM can improve the precision of sales predictions across predictive models and ets.

2.2. Opinion mining and sentiment analysis

Opinion mining and sentiment analysis emphasize on extracting the author's perspective and emotional polarity(such as positive and negative) ,and it can be seen as a suite of procedures for identifying the author's emotions, opinions and attitudes displayed in the text.The classification of emotions in textual content with automatic way can be done in two ways: a lexicon-based method and machine learning method (Ahmad, Asghar, Alotaibi, & Khan, 2020). Lexicon-based approach uses the linguistic resource called sentiment dictionary or sentiment lexicon, such as SentiWordNet, SenticNet and OpinionFinder. Machine learning approach using well-known machine learning algorithms can be divided into three categories: supervised, semi-supervised and unsupervised.

2.2.1. lexicon-based approach

Among lexicon-based approaches, public vocabularies like SentiWordNet is often used in many studies because of its broad coverage and the reliability issues with artificial sentiment lexicons (Madani, Erritali, Bengourram, & Sailhan, 2019).The lexicon can be created manually or automatically. It is extremely challenging and time-consuming to create lexicons by manually method, so it is more common to create a dictionary automatically. The general approach is to build a seed collection and then extend it to form a lexicon.Lexicon-based methods can be classified into two main groups: dictionary-based approach and corpus-based approach (Shi, Zhu, Li, Guo, & Zheng, 2019).The dictionary-based method is based on bootstrapping techniques, and the algorithm helps to grow this collection by searching online dictionaries (such as WordNet: http://wordnet.princeton.edu) to find synonyms and antonyms (Vij, Tayal, & Jain, 2020). The corpus-based method relies on syntactic rules proposed by Hatzivassiloglou and McKeown (1997).It has been proved that the dictionary based method is more useful because it is difficult anyway to prepare large entries that can cover all English words. (Ahlem & Khalil, 2020).

Owing to lexicon-based method increases efficiency of text classification tasks significantly, numerous studies have used this method to analyze social media data to understand social media storms (Holthoff, 2020)and product reviews(Britzolakis, Kondylakis, & Papadakis, 2021). LIWC2015(Pennerbaker, Boyd, & Jordan, 2015), a text mining software supporting a sentiment lexicon for positive and negative sentiments, was used to conduct a lexicon-based sentiment analysis of data samples in psychology and linguistics. A contextual significance study also proposes an algorithm that automatically constructs a word-level sentiment dictionary for social sentiment detection and points out that the dictionary generated for a specific purpose is more effective in predicting the emotional distribution of news articles (Wang, Gao, Tao, Mei, & Tong, 2018). However, like sentiment analysis, it is difficult for a dictionary to infer meaning from the co-occurrence of words. In addition, because the process relies on a variety of subjective steps, dictionary-based text analysis often presents the risk of being too specific and missing words. Therefore, it may not be able to adequately reflect the entire data set or be flexibly enough to apply to new texts.

2.2.2. Machine learning approach

People often use machine learning methods to analyze sentiments and mine opinions. This method generates a classification algorithm with linguistic features in the training data set, and tests the performance of the algorithm in the verification data set(Sharma & Sharma, 2020). Machine learning methods can address more complex meanings than dictionary-based methods. In addition, the classifiers can be trained so that they can be used over and over again.

Pang et al. applied three machine learning approaches to identify whether a movie comments were positive or negative,such as NB, SVM and maximum entropy classification (Pang et al., 2002). Many different machine learning approaches, such as SVM and NB, are often used in market research and have been found to be more effective in traditional text categorization. (Kubler et al., 2017).Supervise machine learning has been introduced increasing in the area of marketing and consumer researches. For example, Ordenes et al. (2018) used support vector machines in their study to mine brands' information intent on social media and thus evaluate their impact on the sharing behavior of consumers.Another study used SVM and NB classifiers to test the impact of EWOM on consumer intentions(Sharma & Sharma, 2020). The researchers conducted an intent training on NB, conducted emotional training on the SVM, and verified the EWOM's impact through accuracy and recall rate as a performance indicator. Supervised machine learning is also commonly employed to analyze the affect of UGC. For instances, Homburg et al. (2015) used supervised SVM to survey the impact of company interventions on consumer sentiment in online forums. One study attempted to use machine learning approaches to categorize the sentiment of microblogs and tweets using features of morphemes and n-grams(Ahmad, Asghar, Alotaibi, & Khan, 2020). The classification algorithms they used were NB and SVM and the accuracy of the classification performance was mainly based on the F1 score.

3. Proposed method

Most previous studies on opinion mining and sentiment classification have been analyzed from the perspective of product attributes. The usual method is to determine the product attributes through the statistics of the word frequency, and judge the emotional tendency of the sentence. We proposed a novel method for learning a classifier from labeled training data,which can then be used to classify eight features from the aspect of user's behavior. Our proposed method has novelties on both main factors.First, the analysis of the opinion mining and sentiment classification is based on user's behavior, not on product's characteristics. Second, we propose a novel concept learning and classification through applying the technology of natural language process.

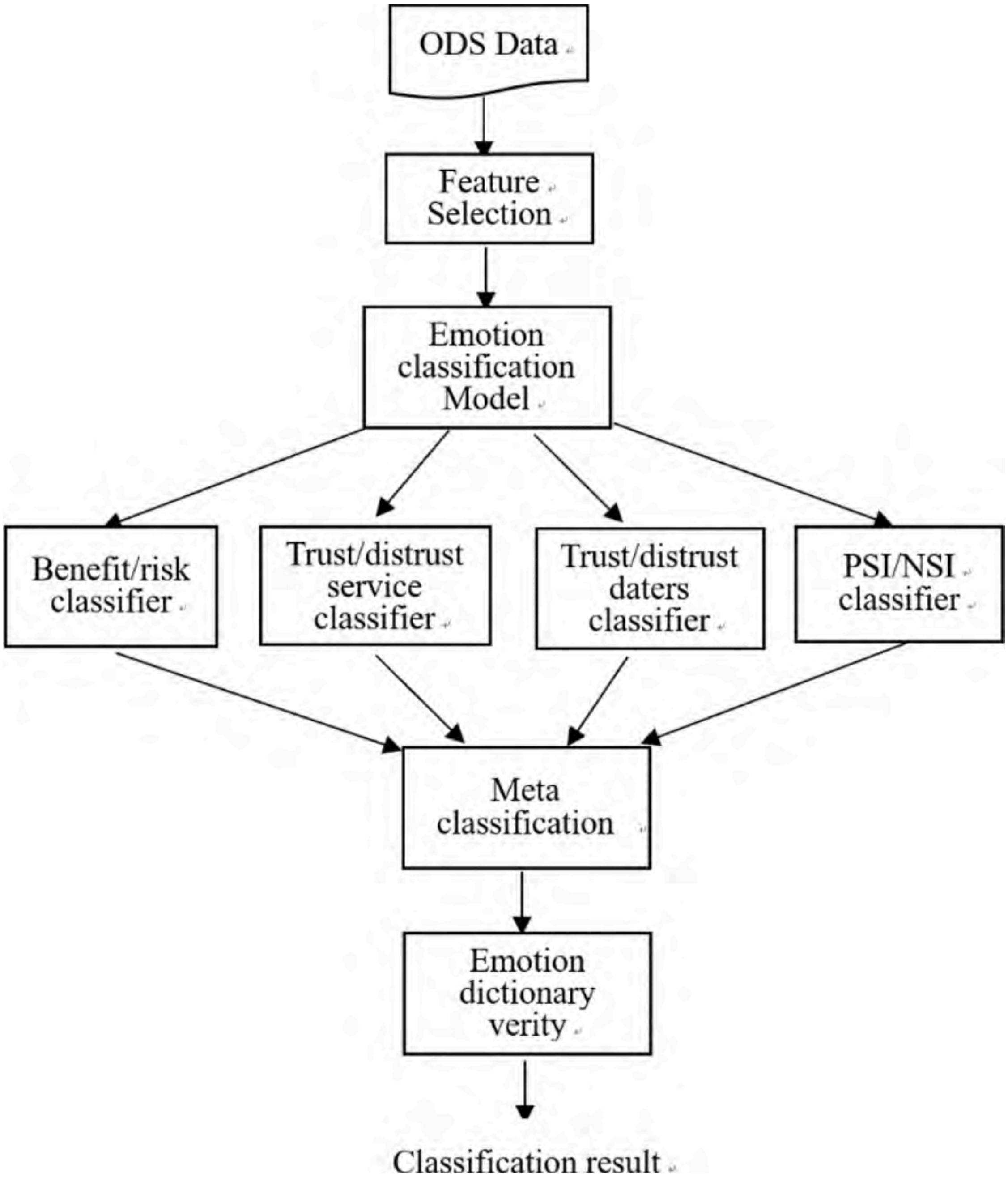


Fig. 1. General workflow of our method

3.1. Sample

Social cognitive theory proposes that the likelihood of a person's willingness to engage in a behavior is affected by cognitive personal and environmental assessments (Bandura, 1986). If these assessments are favorable, a behavioral attention may be developed, and vice versa. We used the social recognition principles to extract concept, which can reflect the people's willingness from different aspects.

Emerging studies have focused on individual perceptions, beliefs and expectations as prospective individuals elements that might be implicated in the determination process, impacting adoption-related decision-making (Benight and Bandura 2004). For example, perceived benefits provide users with the incentives for individual adoption intentions, while perceived risks are proposed to have negative effects on the intentions. Besides, trust and distrust are other key personal factors influencing individual decisions. They may coexist and have different influences on individual adoption intentions (Dimoka 2010). The environmental factors external to the individual (e.g. social influence) also predict one's behaviors (Bandura 1989). Prior studies have confirmed that individuals often adopt a technology because they perceive that important others believe that they should use it or not (Xu et al. 2011).

In summary, we use the eight factors derived from social cognition theory as conceptual features influencing user behavior (Trust and mistrust, sense of interest and risks, positive and negative social impacts on online services and selected dating partners) in the background of internet dating services. In the following, we will consider the eight features as our classification target to classify the review opinions of online dating service.

Since these eight features also describe the influence factors of the individual's intention to adopt or non-adopt the platform from different aspects, we can use it as potential influence factors to predict user' behavior decisions, called Intent Behavior, which is defined as follows.

Definition 1: Intent Behavior: Inspired by the features on intent behavior of users in SOD, users generally adopt/deny the platform only if they think that the platform will bring some benefits/risk points. These factors can be understood as the intent for users to use the platform. we define intent behavior if (1) it contains at least one benefit or risk and (2) explicitly describes the user trust or distrust the online service or selected daters (3) there are some people recommend or prevent to use it.

Definition 2: Intent-Indicator: It comprises a group of terms that are used by users to express behavior intents. It is noun, adjective or compound noun that immediately follows a subject word. For example, in tweet "该平台值得依赖(This platform is trustful.)", "trustful" is an intent-indicator, indicating the user is likely to adopt this SOD platform.

Users' intents exhibited in SOD may belong to different categories, and different categories of intents may be of interest to different applications. To establish taxonomy for intent behavior of user, we have reviewed a large number of reviews in SOD based on social recognition theory above. Finally, we define four types of intent: benefit & risk: In the ODS context, perceived benefits may contribute to individual adoption, although the observed risks prompted the absence of adoption.

- trust & distrust service: In social interaction of ODS, individuals who have strong trust in the service would contribute to individual adoption, while distrust would result to no-adoption.
- trust & distrust dater: In social interaction of ODS, individuals who have strong trust in the selected daters would result to individual adoption, while distrust would result to no-adoption.
- PSI & NSI: In the online environment, positive social influence (PSI) may contribute to individual adoption, while negative social influence (NSI) is associated with non-adoption.

3.2. Our learning method

The basic idea behind our approach is to multi-classify of emotion from user's behavior combining machine learning and lexicon dictionary. As mentioned above, there are several possible characteristics that can be applied to predict the behavior of an individual. We focus on identifying concept which can be used to express user's intention and sentiment analysis from e-word of mouth for user's behavior analysis. We can also consider it as an emotional multi-classification problem for user's behavior analysis. Most previous works of emotion categorization have focused on binary classification, i.e., positive and negative. Yet, it is often more practical to reveal more detailed information of multi-categorization systems. In our research, there are more than one factor that affect whether users adopt or non-adopt the SOD platform, therefore, these influencing factors are separately classified.Finally, different classification methods may be used to construct meta-classifier combining the predictions of different classifiers into a final classification result.

Fig. 1 outlines the proposed method. The data obtained from the ODS is unstructured text data that needs to be converted into structured data by feature selection. Then we construct four classifiers, benefit/risk classifier, trust/distrust service classifier, trust/distrust daters classifier, PSI/NSI classifier, to classify individual's emotions from eight feature aspects respectively. The input features of the meta-classifier are the outputs of the four classifiers. We also construct a lexicon dictionary to modify the wrong prediction result of the four classifiers. Lexical and machine learning methods can be integrated to compensate for each other's shortcomings and deficiencies.

4. Case study: user controversial behaviour in online data service

4.1. Data collection and pre-processing

(1) Data collection. We gathered data from a famous public word-of-mouth website in China, Baidu reputation.com. We selected this website for several reasons. First, it is one of the most popular consumer word-of-mouth platforms in China and founded in 2014 by Baidu.com which is the leading Chinese search engine. In addition, it is a famous UGC (User Generated Content) aggregate interactive platform. We selected the six famous companies in Chinese ODS industry, including Baihe, Youyuan, Zhenai, Shijijiayuan, 58tong-chengjiaoyou and Supei.com. We then used the deep collection strategy to craw to the data of the websites. In order to improve the availability of getting the three leading ODSs' network data, we use the crawler software–Octopus collector (http://www.bazhuayu.com) to obtain the review data. We use the deep collection strategy to craw the data of websites. Finally, a total of 4080 valid reviews were collected.

(2) Preprocessing. Pre-processing also includes removing duplicate, stop words, special characters.In addition, word splitting and negative recognition were performed.In English, spaces are the natural separator of words.Different from English, Chinese does not use any separator of words, each of which may consist of one or more Chinese characters,so we must separate Chinese word into a sequence of words.We use ICTCLAS (http://ictclas.nlpir.org/), a Chinese lexical analysis system, to perform Chinese word separation.

Table 1
several example of manual coding in training model

Example reviews	A	B	C	D	E	F	G	H	I	J
"Baihe Net is a very good, trustful and famous platform. My parents suggested me to use. I found my current boyfriend through it. This site is really great."	1	0	1	0	1	1	1	0	1	0
This website is still quite reliable, and has always been very trustworthy. I have met many good boys in it.	1	0	0	0	1	1	0	0	1	0
There are a lot of members inside, so there are more opportunities to find lovers. I have met a few girls on it, and it is very reliable.	1	0	0	0	0	1	0	0	1	0
Recommended by girlfriends, as he said, there are so many people, and I have met a few good boys, hoping to succeed.	1	0	1	0	0	1	0	0	1	0
Very rubbish, most of the replies are replied by the website robot, Don't believe it, don't charge it!	0	1	0	1	0	0	1	1	0	1
These platforms are useless. In addition to charging fee, there are not many people who really contact you. Even if contact you, it is not reliable.	0	1	0	0	0	0	0	1	0	1
The platform is not bad. there are too many members' fakes, the information is fake, indicating that the staff of the company is too bad.	1	1	0	0	0	0	1	1	0	1
The online reputation is still quite high. However, a lot of mails in the mail are automatically sent by the other party's game system. I was disappointed after opening it and wasting my stamp. I hope to improve a lot at this point.	1	1	1	1	0	0	1	1	0	1

A-J corresponds to the following: Benefit, risk, PSI, NSI, trust-service, trust-dater, distrust-service, distrust-dater, accept, non-accept.

4.2. Feature extraction

The pre-processed comment text is more generalized data, but it is still unstructured data. If you want to classify these unstructured data, it is necessary to turn it into structured data and produce feature collections that can be used for machine learning models. This process is called feature extraction. The frequency-based model (such as TF, TF-IDF)(Jiang, Gao, He, Han, & Zhu, 2021) is a commonly used feature extraction method.The advantages are simplicity and speed, and more accord with the actual situation. The disadvantage is that the importance of measuring a word simply by term frequency which is not comprehensive enough. It is lack of considering the semantic environment of the context. Therefore, we use a new method based on *context coherence* to construct a document-term matrix. There is a difference between the model in this paper and traditional frequency-based models, in which the significance of a phrase in a review is depends on how often it appears.Our model assumes that one word is significant if it contains as many words as possible.We first calculate the similarity W_{ij} between w_i and w_j throughout the reviews of customer D. As shown in Eq. (1),the similarity between w_i and w_j is calculated using point mutual information (PMI) (Yang et al,2010). Where P(w_i, w_j) reduce the likelihood of two words w_i and wj occurring together in a comment. P(w_i) and P(w_j) indicate the likelihood of occurrence of w_i and w_j in a comment, respectively.

$$W_{i,j} = \log\frac{p(w_i, w_j)}{p(w_i)p(w_j)} \tag{1}$$

To compute the context coherence and construct a document of word frequency matrix, we first denote the reviews as a term-term matrix. Each row and each column of the matrix represents all unique terms. Each cell describes the score of word-pair similarity in reviews. The context coherence of a given word wi is calculated by taking the average similarity of the word with other words using the following equation:

$$CC(w_i) = \frac{1}{n}\sum_{j} W_{i,j} \tag{2}$$

Where n represents the number of words in D. Since most word pairs do not occur together frequently in short reviews, our model uses the average PMI to construct the document term matrix.It isimpossible to use all the terms in the reviews as feature sets in machine learning methods. The size of the feature is one of the important factor in machine learning algorithms because it affects the accuracy and processing power of the performance. It is necessary to select an effective feature set that has better classification ability. We thus selected with highest context coherence in the data using a function of NLP functions for data, which eliminates the sparse term and applies it to the training and testing phases. The traditional representation of word frequency matrix will bring serious matrix-sparse problems. In our model, we use context coherence instead of the term frequency, which solves the matrix-sparse problem and improves the efficiency of feature extraction.

4.3. Machine Learning

Machine learning is a sub-topic of artificial intelligence(AI) that focuses on building algorithmic models that can identify patterns and relationships in data. Based on feature set mentioned above, we use several models to examine the effectiveness of machine learning techniques.. We randomly sample 3000 individual reviews, and three annotators with experiences are employed to annotate the category label. Then, every review was coded as the existence of more than one feature described in section 3.1. We present some representative coding results of the manual coding in Table 1.Then we use the manually coded data to build and test our classifiers.

Based on our SOD data, we then created models using four machine learning algorithms to compare performance.We selected the following classifiers frequently adopted in previous research: Naïve Bayes (NB) (Pang et al., 2002), K-nearest neighbor (KNN) (Tsagkalidou, Koutsonikola, Vakali, & Kafetsios, 2011), and Support vector machines (SVM) (Cortes and Vapnik, 1995).

Table 2
"Seed words" of Intent-Indicator sentiment lexicon

feature	Seed words
Benefits	Influence(丰富);Good quality(质量好); free(免费); Convenience(方便); perfect(完善);Popular(有人气);social circle(社交圈)
Risks	nausea(恶心); Leaking secrets(泄密); Harassment(骚扰); waste(浪费);Rubbish(垃圾);Insult(侮辱);Loss(损失)
PSI	Friend recommend(朋友推荐); Good reputation(口碑好); Success case(成功案例); Famous(著名的)
NSI	Don't believe(不要相信); Not recommended(不推荐); Don't be fooled(不要上当); Learn a lesson(引以为戒)
Trust of service	Reliable(靠谱);Trustful(可信的); Safety(安全); Integrity(诚信);Real-name system(实名制);Conscience software(良心软件)
Trust of dater	Reliable person(人靠谱); Good person(人不错);Real membership information (会员信息真实) ;Trustworthy person (可信的人)
Distrust of service	Not reliable(不靠谱); Untrustworthy(不可信); Fraud(诈骗); Deceive(欺骗); False(虚假); Shirk(推卸);Fudge(忽悠人)
Distrust of dater	Con man(骗子); Shill(托儿); Loser(屌丝); Robot(机器人); Mixed dragon(鱼龙混杂);

The model studied in this paper belongs to the classification problem in machine learning, and more specifically belongs to the emotional classification problem of text. There are three techniques used of the machine learning methods are used to classify the sentiments: supervised (Vashishtha & Susan, 2020), semi-supervised (Chen, Feng, Sun, & Liu, 2019), and unsupervised learning methods (Xu & Qiu, 2019).

It is time consuming to label data in traditional classifier. Semi-supervised learning addresses this problem by using large amount of unlabeled data, together with small amount of labeled data, to build better classifiers. In this paper, we use the self-training approach which is a popular learning method among the semi-supervised to construct the training model.In the self-training methods, the classifier should be trained using a small number of labeled training samples firstly, then the trained classifier will be trained with the set of all training samples again and process will be repeated. To obtain the best prediction outcome, three patterns were created based on the output of the self-trained model and the best pattern was selected.

4.4. Meta classification

In our proposed multiple emotion learning method, each base classifier (benefit/risk classifier, trust/distrust service classifier, trust/distrust daters classifier, PSI/NSI classifier) is based on one of the four categories of features. The four classifiers are combined using a logistic regression model (Archer et al., 2006). Logistic regression uses a logistic function to model a statistical model of a binary dependent variable. This is often used to describe the data and it can also be used to illustrate the association between the causal factor and one or more nominal, ordinal, intra-district or ratio-level causal factors.The reason why we chose logistic regression for the meta classifier is that the logistic regression is better choice for meta classifiers (Wolpert, 1992).

4.5. Combine machine learning and lexicon

In section 3.1, we have given the definition of "Intent-Indicator", which expresses the user's explicit evaluation of the SOD's eight features. Therefore, we can use Intent-Indicator to further modify the classification results of machine learning. We first choose the "seed words" manually to form the initial Intent-Indicator sentiment lexicon. Then we use a graph constructed to mine the relationship between words (such as synonyms, antonyms, superior and inferior relations) to realize the lexicon's expansion automatically. Table 2 shows the "seed words" containing 44 words we initially selected.

The initial state of the graph only includes the seed word and in its first iteration all of its synonyms are obtained from HowNet (http://www.keenage.com/)using the library PyDictionary (https://pypi.org/project/PyDictionary/).Through each iteration, a layer of synonyms is generated from the nodes in the previous layer.The generation of the graph is terminated when any antonym of the target word is found. This is because the antonym with opposite meaning to the target word presents a necessary stopping point in the process of graph generation. For efficiency, If there are too many synonyms for a seed word, we will set the maximum number of synonyms to eight.When production of word is complete, there will be formed edges between each of the word nodes.Then put the synonyms into the set of the "seed words". Through the loop iteration, the final Intent-Indicator sentiment lexicon was constituted.

A example of the synonyms generation path for the word "Credible(可信的)" is as follows:

> *Credible*(可信的)->*Trustful* (可信的) -> *Credible* (可信的) -> *Plausible*(合理的)-> *reasonable* (有理的) -> *rational*(合理的)->*feasible* (可行的) ->*probable*(可能的)

Finally, through integrating artificially and deleting the wrong words, the Intent-Indicator sentiment lexicon was composed by 225 words.

In order to optimize the performance of our machine learning classifier, we use a lexicon classifier to optimize its performance.The lexical classifier verified the output label to 1 when the intent-indicators occur in the feature. Such operation will continue until results of the two datasets have the minimum distance. Lexical method and machine learning method can be combined to compensate the disadvantages and drawbacks of each other.

Table 3
The result of classifying sentiment

	benefits	risks	PSI	NSI	TO	TD	DTO	DTD	AC	NAC
Precision	0.905	0.930	0.890	0.750	0.895	0.900	0.870	0.860	0.935	0.890
Recall	0.915	0.935	0.855	0.800	0.935	0.950	0.870	0.745	0.945	0.920
F1-score	0.905	0.935	0.870	0.765	0.915	0.950	0.87	0.775	0.935	0.950
Accuracy	0.900	0.930	0.900	0.780	0.934	0.900	0.868	0.819	0.934	0.950

5. Experiment and discussion

5.1. Performance metrics and estimation method

Various metrics such as precision, recall and accuracy are important factors in measuring the performance of evaluation algorithms and sentiment analysis, and F1 score (Ahmad, Asghar, Alotaibi, & Khan, 2020). These metrics are founded on the following concepts, which relate to the correct or incorrect classification of events (Esuli and Sebastiani, 2010).

- True Positive (TP): the occurrence has been correctly classified as part of the category;
- False positive (FP): the occurrence has been incorrectly classified as part of the category;
- True Negative (TN): the occurrence has been correctly classified as not part of the category;
- False Negative (FN): the eoccurrence has been incorrectly classified as not part of the category

The measures of accuracy, precision, recall and F1 are defined as follows:

Accuracy is the ratio of correct predictions to the total number of predictions. The metric is only valid if there are an equal number of observations in each category. The more accurate it is, the model is better (Esuli and Sebastiani, 2010; Kent et al., 1995).

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

Precision (also called positive predictive value) is used to measure when an occurrence that belongs to the category set is classified as part of category set.Higher precision means that the positive recognition is more correct.(Esuli and Sebastiani, 2010; Kent et al., 1995).

$$precision = \frac{TP}{TP + FP}$$

Recall (also called sensitivity) represents the proportion of true positive cases predicted as positive cases for all positive cases. .High recall indicates that your model is prominent in identifying positives correctly (Esuli and Sebastiani, 2010; Kent et al., 1995).

$$recall = \frac{TP}{TP + FN}$$

Since higher recall can be obtained using lower precision, classification methods have an inherent trade-off between precision and recall. Using the F1-score metric provide a more balanced assessment for performance of classification(Cohen and Singer, 1999).

$$F1 = \frac{2 \ \times prcision \times recall}{precision + recall}$$

5.2. Experiments of classifying sentiment

5.2.1. Effectiveness of classifying sentiment

In the whole data set, we selected 300 comments' data randomly, and then applied our proposed classification model to predict 10 features (benefits, risks, positive social influence (PSI), negative social influence (NSI), trust of online dating services (TO), trust of selected daters (TD), distrust of online dating services (DTO), distrust of selected daters (DTD), acceptance (AC), non-acceptance (NAC)) separately, and then compared their prediction results with the results of manual label. Table 3 shows the result of the model prediction.

Since we use the best of the three classification methods (KNN, NB and SVM) in each feature classification, the accuracy of classification is relatively high. For most features (benefits, TO, TD, AC), classification algorithm has higher precisions while maintain recalls on the same level.In terms of F1-score, which evaluates both precision and recall, our classification also get high performance in most cases.For the feature of NSI, our classification result is worse than others. Our assumption is that it is less likely that people will be willing to use or review the SOD platform when someone gave his negative influence. So the size of training dataset becomes smaller, the efficiency of classification will become less effective.

5.2.2. Effectiveness of merging learning method

It can be seen from Table 3 that the emotional classification's effect of each feature is in line with expectations. What needs to be

Table 4
The performance of different meta mehods

Method	Precision	Recall	F1-score	Accuracy
Proposed method	**0.895**	**0.897**	**0.894**	**0.892**
Random Subspace	0.875	0.875	0.842	0.850
Bagging	0.835	0.837	0.839	0.825
Boosting	0.850	0.850	0.868	0.852

Table 5
The efficiency of machine learning and lexicon-based

Approach	precision	Recall	F1 score	Accuracy	True TP	False FP	True TN
Evaluation of positive classification (total tested reviews N=400)							
Manual	1	1	1	1	261	0	139
Lexicon-based	0.86	0.69	0.77	0.73	180	27	115
Machine learning	0.84	0.72	0.77	0.73	185	35	110
Combined approach	0.82	0.89	0.85	0.80	235	51	85
Evaluation of negative classification (total tested reviews N=400)							
Manual	1	1	1	1	85	0	315
Lexicon-based	0.41	0.94	0.56	0.71	75	110	210
Machine learning	0.59	0.63	0.61	0.85	45	31	298
Combined approach	0.37	0.92	0.53	0.70	76	130	200

verified below is the emotional classification's effect of the entire review, which is the effect of our meta classification proposed in section 4.4. We evaluated the effectiveness of our proposed meta method comparing with other ensemble methods-such as Random Subspace, Bagging, Boosting-as baselines (Yao et al., 2018).Each basic classifier in the random subspace is built on a subset of random features. While in Boosting, all the basic classifiers are based on all the features.

The meta-classifier is determined by logistic regression because it is the better choice for meta classifiers (Wolpert, 1992). Table 4 shows the performance of each meta method. For all three performance metrics, our proposed meta-classifier is superior to any other method. The *t*-test shows that our proposed method is statistically significant for each performance indicator($p<0.01$).

5.2.3. Effectiveness of combing machine learning to lexicon

In another study, we examined the effectiveness of combing machine learning to lexicon. The purpose is not to prevent manual labeling of the training dataset, but to integrate the advantages of combining a lexicon-based approach with a machine learning approach to improve the accuracy of the results. Since the Intent-Indicator sentiment lexicon constructed in section 4.5 can express the user's willingness explicitly, using it to correct the classification's results of machine learning will achieve better classification results theoretically.

Table 5 shows the results of the integrated approach. The F1-score score in the classification of positive emotions obtained a substantial increase, but the F1-score for the classification of negative emotions remained relatively unchanged at around 0.56 when the combined method was used. Combining these two methods resulted in a significant increase in the overall performance of sentiment classification. The fact that positive and negative emotions do not interfere with each other during classification leads to the conclusion that the integrated approach is more valuable in the sentiment analysis of positive word-of-mouth.

Table 5 shows that the number of reviews classified correctly increased significantly due to the combination of the two methods. This suggests that one method is complementary to the other and that the combination of the two methods may produce better results.

6. Conclusion and future work

Sentiment classification in social communication media, particularly online chat services, has attracted considerable research attention.However, existing research largely focus on the product's review. In this paper we focus on the reviews of online dating service (ODS) because this is different from the product's reviews. We tried to conduct sentiment analysis of user' behavior influenced by conflict environmental factors using machine tools to understand availability through comparing of the performance.

The findings of this study have implications for both researchers and practitioners. For research, our work makes several noteworthy theoretical contributions to the IS literature. First, our study enabled by both machine learning and lexicon provides complementary perspectives for traditional questionnaire method, implying some unexpected moderating and asymmetrical effects. Interview data provides a rationale for this result in that viewers perceived these specific elements as prominent and different in influencing their behavior decisions. For practitioners, our study indicates that machine learning may provide a viable alternative that identifies unexpected findings, which is free from reliability problems in the survey method. Thus, the providers may need to pay attention to a multi-method approach to examine individual behavior.

Although the existing research results are valid, there is still room for improvement. First, the accuracy of the model cannot be guaranteed in the case of cross-domain. Secondly, when the information is insufficient, the training of the model is difficult, and a larger and more diverse databases are needed to improve the accuracy of the model. Third, manual annotation of category labels is required, which requires a lot of manual labor.In the future we can continue to extent our work. First, future research can expand our model for different types of controversial IS. Second, To reduce the amount of manual labour required, semi-supervised learning techniques can be used. Both dynamic and incremental learning approaches may be used to extend our learning methods. In the meantime, the findings of this research offer useful insights for providers of ODS that will help them to improve their service performance effectively.

CRediT authorship contribution statement

Hui Li: Validation. **Qi Chen:** . **Zhaoman Zhong:** Data curation, Funding acquisition, Formal analysis. **Rongrong Gong:** . **Guokai Han:** .

Acknowledgments

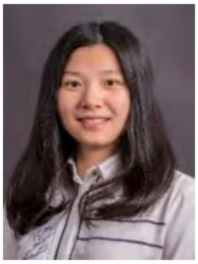
The research is supported byNational Natural Science Foundation of China (No.72174079, No. 1210050123, No. 72101045), Cooperation and Education Project of Ministry Education(201701028110, 201701028011, 201902159041), Natural Science Foundation of the Jiangsu Higher Education Institutions of China(No. 19KJB520004), Jiangsu Province “333” project (BRA2020261), Lianyungang “521 project”, Science and Technology project of Lianyungang High-tech Zone(No. ZD201912), Teaching reform research project of Huaihai Institute of Technology(XJG2017-2-5).

References

Ahlem, D., & Khalil, H. (2020). An opinion spread prediction model with twitter emotion analysis during algeria's hirak. *The Computer Journal, 64*(3), 358–368.
Ahmad, S., Asghar, M. Z., Alotaibi, F. M., & Khan, S. (2020). Classification of poetry text into the emotional states using deep learning technique. *IEEE Access, 8*, 73865–73878.
Al-Mashhadany, A. K., Hamood, D. N., Al-Obaidi, A., & Al-Mashhsdany, W. K. (2021). Extracting numerical data from unstructured arabic texts (enat). *Indonesian Journal of Electrical Engineering and Computer Science, 21*(3), 1759–1770.
Archer, K. J., & Lemeshow, S. (2006). Goodness-of-fit test for a logistic regression model fitted using survey sample data. *The Stata Journal, 6*(1), 97–105.
Bandura, A. (1986). Social Foundation of Thought and Action: A Social-Cognitive View. *Englewood Cliffs)*.
Bandura, A. (1989). Human Agency in Social Cognitive Theory. *American psychologist, 44*(9), 1175.
Benight, C. C., & Bandura, A. (2004). Social Cognitive Theory of Posttraumatic Recovery: The Role of Perceived Self-Efficacy. *Behaviour research and therapy, 42*(10), 1129–1148.
Britzolakis, A., Kondylakis, H., & Papadakis, N. (2021). A review on lexicon-based and machine learning political sentiment analysis using tweets. *International Journal of Semantic Computing, 14*(4), 517–563.
Cohen, W. W., & Singer, Y. (1999). Context-sensitive learning methods for text categorization. *ACM Transactions on Information Systems (TOIS), 17*(2), 141–173.
Cortes, C, & Vapnik, V (1995). Support-vector networks. *Mach Learn, 20*(3), 273–297.
Chang, C. H., Hwang, S. Y., Wu, M. L., et al. (2021). Learning bilingual sentiment lexicon for online reviews. *Electronic Commerce Research and Applications, 47*(2), 101037.
Chen, J., Feng, J., Sun, X., & Liu, Y. (2019). Co-training semi-supervised deep learning for sentiment classification of mooc forum posts. *Symmetry, 12*(1), 8.
Dimoka, A. (2010). What Does the Brain Tell Us About Trust and Distrust? Evidence from a Functional Neuroimaging Study. *Mis Quarterly*, 373–396.
Donthu, N., Kumar, S., Pandey, N., Pandey, N., Mishra, A., et al. (2021). Mapping the electronic word-of-mouth (eWOM) research: A systematic review and bibliometric analysis. *Journal of Business Research, 135*(2021), 758–773.
Eng, T., Nawab, M., & Shahiduzzaman, K. M. (2021). Improving accuracy of the sentence-level lexicon-based sentiment analysis using machine learning. *International Journal of Scientific Research in Computer Science Engineering and Information Technology, 7*(1), 159–165.
Esuli, A., & Sebastiani, F. (2010). Machines that learn how to code open-ended survey data. *International Journal of Market Research, 52*(6), 775–800.
Gao, S., Tang, O., Wang, H., & Yin, P. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *International Journal of Hospitality Management, 71*(2018), 19–32.
Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the European chapter of the association for computational linguistics* (pp. 174–181). Association for Computational Linguistics.
Holthoff, L. C. (2020). The Emoji Sentiment Lexicon: Analysing Consumer Emotions in Social Media Communication. *49th European Marketing Academy (EMAC) Annual Conference*.
Homburg, C., Ehm, L., & Artz, M. (2015). Measuring and managing consumer sentiment in an online community environment. *Journal of Marketing Research, 52*(5), 629–641.
Hung, C., Wu, W. R., & Chou, H. M. (2020). SenticNet. *Sentic Computing, 55*(2015), 585–595.
Jiang, Q., Chan, C. S., Eichelberger, S., Ma, H., & Pikkemaat, B. (2021). Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists. *Current Issues in Tourism, 24*(17), 2501–2522.
Jiang, Z., Gao, B., He, Y., Han, Y., & Zhu, Q. (2021). Text classification using novel term weighting scheme-based improved tf-idf for internet media reports. *Mathematical Problems in Engineering, 2021*(6), 1–30.
Kent, A., Berry, M. M., & Luehrs, F. U. (1995). Operational criteria for designing information retrieval systems. *American Documentation, 6*(2), 93–108.
Kim, K., Yoon, S., & Choi, Y. K. (2019). The effects of eWOM volume and valence on product sales–an empirical examination of the movie industry. *International Journal of Advertising, 38*(3), 471–488.
Kubler, R. V., Colicev, A., & Pauwels, K. H. (2017). Social media's impact on consumermindset: When to use which sentiment extraction tool. *Marketing Science Institute Working Paper Series, 17*(122), 1–99.
Lau, R. Y. K., Zhang, W., & Xu, W. (2018). Parallel aspect-oriented sentiment analysis for sales forecasting with big data. *Production and Operations Management, 27*(10), 1775–1794.
Li, J., H, Li, Kang, X., Yang, H., & Zong, C. (2019). Incorporating multi-level user preference into document-level sentiment classification. *ACM transactions on Asian language information processing, 18*(1), 7.1–7.17.
Liu, H., Jayawardhena, C., Osburg, V. S., Yoganathan, V., & Cartwright, S. (2021). Social sharing of consumption emotion in electronic word of mouth (ewom): a cross-media perspective. *Journal of Business Research, 132*, 208–220.
Lu, J., & Bai, H. (2021). Information Usefulness and Attitude Formation a Double-Dependent Variable Model (DDV) to Examine the Impacts of Online Reviews on Consumers. *Journal of Organizational and End User Computing (JOEUC), 33*(6), 1–22.
Lu, Q., Zhu, Z., Zhang, D., Wu, W., & Guo, Q. (2020). Interactive rule attention network for aspect-level sentiment analysis. *IEEE Access, 2020*, 99.

Madani, Y., Erritali, M., Bengourram, J., & Sailhan, F. (2019). A hybrid multilingual fuzzy-based approach to the sentiment analysis problem using sentiwordnet. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems, 28*(3), 30.
Madani, Y., Erritali, M., Bengourram, J., & Sailhan, F. (2020). A hybrid multilingual fuzzy-based approach to the sentiment analysis problem using sentiwordnet. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems, 28*(3), 30.
Ordenes, F. V., Grewal, D., Ludwig, S., Ruyter, K. D., Mahr, D., & &Wetzels, M. (2018). Cutting through content clutter: Howspeech and image acts drive consumer sharing of social media brand messages. *Journal of Consumer Research, 45*(5), 988–1012.
Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In *Proc. of 2002 Conf. Empir. Methods Nat. Lang. Process* (pp. 79–86).
Use of Deep-Learning based LSTMs Natural Language Sentiment Classification Model for Residents' Sentiment Analysis. 대한건축학회 학술발표대회 논문집*, 39*(2), (2019), 168–171.
Pennerbaker, J. W., Boyd, R. L., & Jordan, K. (2015). The development and psychometric properties of LIWC2015. *University of Texas at Austin,Austin,TX*.
Sharma, P., & Sharma, A. K. (2020). Experimental investigation of automated system for twitter sentiment analysis to predict the public emotions using machine learning algorithms. *Materials Today: Proceedings*, 1–9.
Shi, Y., Zhu, L., Li, W., Guo, K., & Zheng, Y. (2019). Survey on classic and latest textual sentiment analysis articles and techniques. *International Journal of Information Technology & Decision Making, 18*(4), 1243–1287.
Siddiqui, M. S., Siddiqui, U. A., Khan, M. A., Alkandi, I. G., Saxena, A. K., & Siddiqui, J. H. (2021). Creating Electronic Word of Mouth Credibility through Social Networking Sites and Determining Its Impact on Brand Image and Online Purchase Intentions in India. *Journal of Theoretical and Applied Electronic Commerce Research, 16*(4), 1008–1024.
Sznycer, D., & Lukaszewski, A. W. (2019). The emotion–valuation constellation: multiple emotions are governed by a common grammar of social valuation. *Evolution and Human Behavior, 40*(4), 124–132.
Tobon, S., & García-Madariaga, J. (2021). The inf1uence of opinion 1eaders ewom on on1ine consumer decisions a study on socia1 inf1uence. *Journal of Theoretical and Applied Electronic Commerce Research, 16*(4), 748–767.
Tsagkalidou, K., Koutsonikola, V., Vakali, A., & Kafetsios, K. (2011). Emotional aware clustering on micro-blogging sources. *Lecture notes in computer science, 6974* (2011), 387–396.
Vashishtha, S., & Susan, S. (2020). Inferring sentiments from supervised classification of text and speech cues using fuzzy rules. *Procedia Computer Science, 167*, 1370–1379.
Vij, S., Tayal, D., & Jain, A. (2020). A machine learning approach for automated evaluation of short answers using text similarity based on wordnet graphs. *Wireless Personal Communications, 111*(2), 1271–1282.
Wang, B., Gao, L., Tao, A., Mei, M., & Tong, Z. (2018). A method of educational news classification based on emotional dictionary. *2018 Chinese Control And Decision Conference (CCDC)*.
Wang, K., & Zhang, Y. (2020). Topic sentiment analysis in online learning community from college students. *Journal of Data and Information Science, v.5*(2), 35–63.
Wang, Y. S. (2020). The application of netnography to the online dating service experiences of female users. *Behaviour & Information Technology, 39*(2), 226–240.
Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating Expressions of Opinions and Emotions in Language. *Language Resourses and Evaluation, 39*(2), 165–210.
Wolpert, D. H. (1992). Stacked generalization. *Neural Netw, 5*(1992), 241–259.
Wu, M., Wang, L., & Li, M. (2015). An approach based on the bass model for analyzing the effects of feature fatigue on customer equity. *Computational & Mathematical Organization Theory, 21*(1), 69–89.
Wu, W. L., Song, M. M., Duan, Y. C., Hong, Y., & Sui, W. J. (2021). The Influence of Electronic Word-of-Mouth Dispersion on Order Decision from the Perspective of Self-Construal. *Journal of Physics: Conference Series, 1944*(1), 1–8.
Xu, L., & Qiu, J. (2019). Unsupervised multi-class sentiment classification approach. *Knowledge Organization, 46*(1), 15–32.
Xu, Z., Turel, O., & Yuan, Y. (2011b). Online Game Addiction among Adolescents: Motivation and Prevention Factors. *European Journal of Information Systems, 21*(3), 321–340.
Yang, C. C., Tang, X., Wong, Y. C., & Wei, C. P. (2010). Understanding online consumer review opinions with sentiment analysis using machine learning. *Pacific Asia Journal of the Association for Information Systems, 2*(3), 1–10.
Yaniv, G., & Shalom, L. (2021). Consumer engagement in sharing brand-related information on social commerce: the roles of culture and experience. *Journal of Marketing Communications, 27*(1), 53–68.
Yao, L., Cuiqing, J., & Huimin, Z. (2018). Using contextual features and multi-view ensemble learning in product defect identification form online discussion forums. *Decision Support Systems, 105*(2018), 1–12.

Hui Li, female, born in October 1979 in Lianyungang, Jiangsu Province, P.R. China. She received her BE degree in computer science and technology from the Yang Zhou University in 2007. She received her Doctoral degree in the School of Information & Electrical Engineering in China University of Mining & Technology in 2016. Her current research interest includes data mining, information processing, social network. She has published more than 20 papers in various journals.



Qi Chen is an assistant professor in the Department of Management Science and Engineering in the School of Business and Management at Dalian University of Technology (DUT), China. She received her Ph.D., M.S. and B.S. in Management Science and Engineering from Harbin Institute of Technology (HIT) in China. Her research interests are in the areas of security, privacy and trust, management information systems, social media. She has published articles in journals and conferences such as International

Zhao-Man Zhong, male, born in 1977, received his PhD degree in computer science from Shanghai University in 2011. He is currently an associated professor of Jiangsu Ocean University. His research interests are information retrieval and artificial intelligence.

Rongrong Gong, is a graduate student at Jiangsu Ocean University. Her main research direction is intelligent recommendation and artificial intelligence.

Guokai Han, is a junior student at Jiangsu Ocean University, majoring in software engineering. He has won the National Computer Design Competition.