



Emoji-Integrated Polyseme Probabilistic Analysis Model: Sentiment Analysis of Short Review Texts on Library Service Quality

Xuan Zheng¹, Wei Chen², Haijun Zhou³, Zhe Li^{4*}, Tianfan Zhang⁴, Qi Yuan⁵

¹ Information Service Center, Wenzhou Business College, Wenzhou 325000, China

² Logistics Service Department, Wenzhou Business College, Wenzhou 325000, China

³ Zhejiang College of Security Technology, Wenzhou 325000, China

⁴ School of Computer and Information Science, Hubei Engineering University, Xiaogan 432000, China

⁵ Mechanical and Electrical Engineering, Hubei Polytechnic Institute, Xiaogan 432000, China

Corresponding Author Email: lizhe_lz@hbeu.edu.cn

<https://doi.org/10.18280/ts.390133>

ABSTRACT

Received: 5 December 2021

Accepted: 19 January 2022

Keywords:

framework for emotional classification of library service quality (FECLSQ), short text, polyseme, emoji, probabilistic linguistic term sets (PLTS)

It is a great challenge to understand user evaluation of library service quality based on short review texts. This is because short texts are limited in length and lack context support. What is worse, the polysemes and emojis in short texts make the literal emotions of these texts rather ambiguous and variable. The variability is often overlooked in previous research on service quality evaluation, which reduces the accuracy of automatic analysis methods. Considering the effects of polysemes and emojis in short texts, this paper introduces probabilistic linguistic term sets (PLTS) and support vector machine (SVM) to establish a framework for emotional classification of library service quality (FECLSQ). Every word and emoji were converted into the corresponding PLTS to depict the probability of the word/emoji belonging to each sentiment polarity, making short text sentiment analysis more accurate. Through supervised learning of corpuses, the authors established the PLTSs of polysemes, and context sentiment weight dictionary (CSWD), and coupled them with the FECLSQ for sentiment analysis and application of text sets with various themes. The proposed approach was utilized to correctly evaluate library service quality.

1. INTRODUCTION

As informatization goes deeper, online services cut down the cost of book management, and provide the service objects with a convenient platform to express personal opinions [1]. These opinions, which convey personal emotional tendencies, facilitate the understanding of problems in the service process, and the reasonable evaluation of the services received by users. Different users often have varied views and emotional tendencies of the process and outcome of the same service. With a good understanding of this phenomenon, decision-makers could formulate reasonable management measures to improve service quality and user experience. To understand user ideas, the important information is generally processed through the natural language processing (NLP) technique of text sentiment analysis. Currently, sentiment analysis is widely used in online service evaluation [2], various recommendation systems [3], public opinion analysis [4], and many other fields [5]. For example, the text sentiment analysis of user reviews on the book management platform yields a user-service relationship model, and provides a user-side quantitative model that supports the improvement of service quality. The latter model is particularly important, for the service-side does not always provide services that are really needed by users. Therefore, the service providers would usually set up an evaluation index of the themes, people, and services [6].

As shown in Figure 1, the current methods for text sentiment analysis mainly fall into two categories: dictionary-

based methods and machine learning methods.

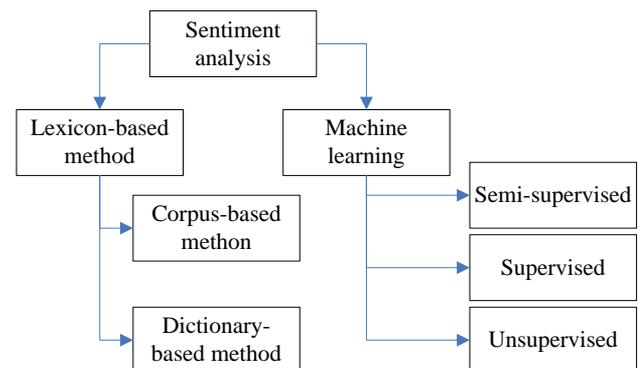


Figure 1. Methods of text sentiment analysis

In most dictionary-based methods, the sentiment polarity score of a word is represented as a real number. If the number is greater than zero, the word is labeled as positive; if the number is smaller than zero, the word is labeled as negative; if the number equals zero, the word is labeled as neutral.

Sentiment analysis methods rely heavily on sentiment dictionaries and related tools. General Inquiry (GI) [7], one of the earliest sentiment dictionaries for artificial classification, has been widely used in sociology, psychology, economics, and anthropology [8]. Each word in this dictionary has one or more positive or negative labels. But GI does not specify the intensity of each sentiment label. Hence, the dictionary is not

suitable for fine-grained sentiment analysis. TextBlob (TB) [9] is an open-source library involving sentiment analysis. It consists of a dictionary of adjectives that often appear in reviews, supplemented with regular expressions. TB calculates the polarity intensity of a given text, and adjusts the polarity score in [-1.0, 1.0] by regulating the word frequency of adjectives and conjunctions. Valence Aware Dictionary and Sentiment Reasoner (VADER) [10], as a rule-based tool for sentiment analysis, is particularly suitable for processing the sentiments in social media. In VADER, the polarity intensity is rated by ten human reviewers, and the ratings fall within [-4, 4]. SentiWordNet (SWN) [11] is an opinion mining dictionary based on WordNet.8, where each word has one or more synonym sets. SWN assigns two sentiment scores to each synonym set: a positive score (PosScore) and a negative score (NegScore). Both scores range from 0 to 1, and are computed by a complex combination of semi-supervised algorithms. Although SWN contains a large vocabulary, most synonyms lack reliable polarity labels.

SenticNet (SCN) provides an open sentiment perception resource [12]. The SCN dictionary incorporates common-sense concepts like anger, adoration, sorrow, and admiration, and differentiates them using four sentiment dimensions (pleasantness, attention, sensitivity, and aptitude). Based on the bag-of-words (BOW) model, the SCN does not simply determine the polarity intensity by co-occurring word frequency. The intensity of sentiment polarity in SCN belongs to [-1, 1].

The artificial construction of sentiment dictionaries costs too much time and manpower. Therefore, many scholars explored the ways to automatically recognize sentiment features [13]. One of these methods is Naive Bayes (NB) classifier [14], which relies on Bayesian probability and assumes that the feature probabilities are independent of each other. Logistic regression (LR) [15] is a general machine learning technique that can be viewed as a generalized linear model. Unlike NB and LR, support vector machine (SVM) [16] is a non-probabilistic classifier that builds a hyperplane in a high-dimensional space, separates data points, and then performs classification. Jalilvand and Salim [17] compared the sentiment classification performance of NB, SVM, and maximum entropy classification (MEC), revealing that SVM is the best performer.

Except for SWM, almost all dictionaries ignore the issue of polysemy. Even SWN does not assign the probability of occurrence to each synonym set. Khan et al. [18] regarded the mean sentiment score of SWN synonym set as the sentiment score of the same word. Yet this approach also leads to information loss and bias. In machine learning, vector space models are a common way to represent text documents, and each dimension corresponds to a specific word. The value and weight of the word are calculated according to the word frequency.

The problem is that most methods overlook word variability, and treat the sentiment polarity of words as being constant. Sentiment analysis of short texts is more challenging than the analysis of typical sentences and documents, owing to the limited context and polysemy. Polysemes make it particularly difficult to analyze the sentiment of a short text. As shown in Figure 2, the word “funny” can describe someone or something that makes you laugh or smile, e.g., “a funny story”. However, “funny” can also express negative emotions, as in “my head bad begun to ache and my stomach felt funny.” It is not uncommon for a word to have both positive and negative

connotations. Since the sentiment polarity of words fluctuates with the context, it is not reasonable to express the sentiment information of a word with real numbers. Otherwise, the information will get lost, and the ambiguity and uncertainty of human language will go unnoticed. In machine learning, the BOW model and word embedding models (e.g., word2vec) ignore the issue of semantic variation. Short texts lack contextual information and contain spelling errors that challenge machine learning methods. Therefore, the same word needs to be described by a set of possible sentiment polarities, during sentiment analysis.

There is an abundance of non-verbal cues (intonation, facial expressions, and body movements) in face-to-face conversations. Similarly, many emotional cues are contained in text messages, such as the capitalizing all letters in English words, repeating a letter or modal particle in a word, and ending a sentence with multiple punctuations [19]. The new generations of Internet users are increasingly inclined to use emojis in reviews and messages. Whether it is a small face that directly expresses facial expressions or a graph that only illustrates objects, every emoji is utilized similarly: (1) It is mostly placed at the end of sentences; (2) It is used similarly to express excitement and depression; (3) It is used to express intimacy through jokes; (4) It is a manifestation of social rules [20]. The existing studies have confirmed that the collocation of text contents with different emotional tendencies and facial emojis will lead to different results [21]. As shown in Figure 3, neutral text information + emoji with emotional tendency = emotion induction; text information with emotional tendency + emoji with consistent emotional tendency = a relatively high emotion induction [22]. Nonetheless, positive words plus negative emojis will only reduce the level of positive emotions, rather than change the original meaning of the words and induce negative emotions in readers. The inverse is also true. It is also worth noting that the number of emojis at the end of the sentence makes no difference. As shown in Figure 3, adding three “dozens of banknotes”, or “sleeping” emojis does not significantly enhance sentiment polarity [22].

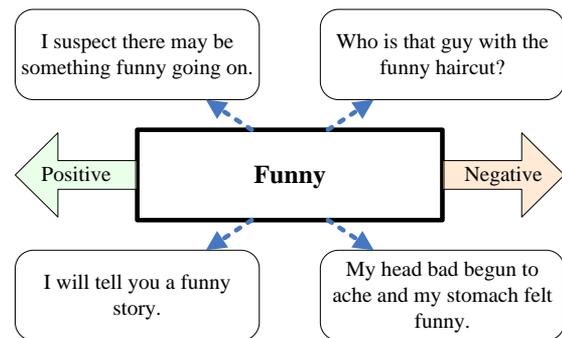


Figure 2. Different uses of the word “funny”



Figure 3. Sentiment enhancement by emojis [22]

As an expression of emotion, emoji is different from other non-verbal cues in conversation in that its use is conscious and active [23]. A special case may arise: The sender’s choice of emojis is not a response to his/her real sentiment, but an intentional behavior called emotion work: the individual efforts of emotional expression management, with the aim to maintain his/her role [24]. The combination of sentiment and text, especially that with polysemes, poses additional challenges to sentiment analysis of short texts.

The ambiguity and uncertainty of languages can be described by probabilistic linguistic term sets (PLTS), whose definition and basic operations were initially given by Pang et al. [25]. However, the operation values of PLTS may be beyond the scope of the given linguistic term set, causing possible loss of language information. Gou et al. solved the problem by redefining richer logical operations [26, 27]. Zhang et al. [28] applied PLTS to evaluate investment risk, and created a new concept called probabilistic linguistic preference relation (PLPR). Peng et al. [29] presented a cloud decision support model named probabilistic linguistic integrated cloud (PLIC) for hotel selection on tripadvisor.com. Focusing on multi-criteria decision-making, Liao et al. [30] put forward a probabilistic linguistic linear programming (PLLP) method to evaluate the level of hospitals in China. Tian et al. [31] came up with a multi-criteria decision-making method based on PLTS and evidence-based reasoning. Luo et al. [32] used PLTS to assess the sustainability of constructed wetlands. Krishankumar [33] developed a decision framework based on PLPR. PLTS has been extensively applied in group decision making (GMD). For instance, Wu and Liao [34] proposed a comprehensive multiple criteria group decision making method (MCGDMC) with PLTS based on consensus measures and outranking methods.

Therefore, this paper introduces PLTS and related theories into sentiment analysis of short texts, and designs a novel PLTS-based framework for emotional classification of library service quality (PLTS-FECLSQ), trying to accurately evaluate library service quality. Under the framework, the sentiment information of a word or an emoji was represented as a PLTS, which was then used to describe the ambiguity and uncertainty of the word/emoji. Next, the sentiment information of each sentence was obtained by aggregating the PLTSs of words and emojis. Finally, the texts were classified by the SVM. The proposed evaluation approach was proved effective on multiple datasets.

The remainder of this paper is organized as follows: Section 2 details the FECLSQ, and its realization process; Section 3 introduces five datasets, including three public ones and two self-built ones, compares our approach with three other typical analysis methods, and demonstrates the effectiveness of our PLTS-FECLSQ; Section 4 sums up the research findings, and looks forward to the future work.

2. METHODOLOGY

2.1 PLTS in service quality reviews

There may be various sentiment polarities in service quality reviews. But it is not very accurate to simply divide the reviews into positive or negative class [35], for a single review may simultaneously contain multiple sentiments [36]. For example, an online review of library service quality may go as: “Although the books are boring, we are impressed by the

service quality of the service staff; Besides, the air-conditioning is poor.” If quantified by the degree of polarity, the review is 50% neutral (the books are boring), 30% positive (service quality is good), and 20% negative (air-conditioning is poor). Then, the subjective review of the library can be expressed as $\text{feel}(\text{library}) = \{\text{terrible}(0.2), \text{ordinary}(0.5), \text{favorable}(0.3)\}$. To quantify the information and proportions, it is important to consider the subjective feelings of users, and other knowledge-based information, such as probabilistic distribution [37, 38], importance [39], and confidence [40]. Ignoring the information may lead to inaccurate analysis and erroneous subsequent decisions.

An effective way to solve uncertainty evaluation is to apply PLTS to decision-making. The sentiment information of each word can be transformed into a PLTS: $\{S_{-\tau}(p), S_{-\tau+1}(p), \dots, S_{-\tau}(p)\}$, where τ is the number of levels, and S_i is the level of sentiment polarity. If $i > 0$, the greater the i , the higher the ranking of positive sentiment polarity; if $i < 0$, the smaller the i , the lower the ranking of negative sentiment polarity. p stands for the probability of S_i . Then, negative, neutral, and positive meanings can be expressed as subsets $\{S_{-\tau}(p), S_{-\tau+1}(p), \dots\}$, $\{\dots, S_{-1}(p), S_0(p), S_1(p), \dots\}$, and $\{\dots, S_{\tau-1}(p), S_{\tau}(p)\}$, respectively. Take $\tau = 4$ for example. The sentiment information of funny can be described by a PLTS: $L(p) = \{S_{-3}(0.1), S_{-1}(0.2), S_3(0.7)\}$. That is, the probability for the word to convey level 3 negative meaning is 10%, that for the word to convey level 1 negative meaning is 20%, and that for the word to convey positive meaning is 70%. Obviously, a large τ value allows thorough description of sentiment information. In this way, PLTS illustrates the sentiment information of words to the maximum possible degree.

2.2 Flow of FECLSQ

To eliminate the effects of uncertain factors (polysemes) on short text sentiment analysis, the FECLSQ framework was divided into three processes: preprocessing, Word to PLTS (Emoji to words), and classification. The flow of this framework is shown in Figure 4.

2.3 Preprocessing

To improve data quality and improve analysis, preprocessing was carried out, including spelling correction, negative word check, and stop word removal.

2.3.1 Negative word check

The sentiment polarity of a text may be changed or reversed by negative words. If negative words like “no” and “not” appear in a text (e.g., not happy, and no success), the sentiment polarity of the positive words (e.g., happy, and success) in the text will be reversed. The reversal time depend on the number of negative words in the text. The negative coefficient λ can be expressed as:

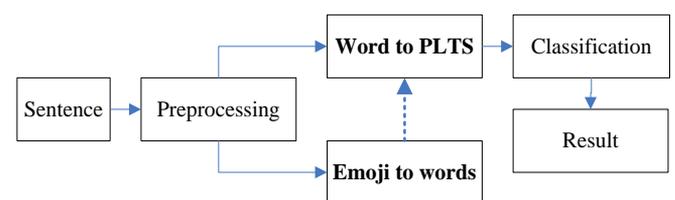


Figure 4. Flow of FECLSQ

Table 1. POS label conversion

POS	Penn Tag	POS	Penn Tag
	RB (adverb)		JJ (adjective)
Adverb	RBR (adverb, comparative)	Adjective	JJR (adjective, comparative)
	RBS (adverb, superlative)		JJS (adjective, superlative)
	VB (verb, base form)		NN (noun, singular or mass)
	VBD (verb, past tense)	Noun	NNP (proper noun, singular)
Verb	VBG (verb, gerund or present participle)		NNPS (proper noun, plural)
	VBN (verb, past form)		NNS (noun, plural)
	VBP (verb, non-3 rd person singular present)	Person	PRP (Personal Pronouns)
	VBZ (verb, 3 rd person singular present)	Determiner	DT (Qualifiers)

Note: Stanford POS Tagger <http://nlp.stanford.edu/software/tagger.shtml>
 Language Technology Platform (LTP) Tagger <http://www.ltp-cloud.com/>

$$\lambda = (-1)^n \tag{1}$$

$$SynsetScore = PosScore - NegScore, SynsetScore \in [-1, 1] \tag{2}$$

where, n is the number of negative words. If $\lambda = -1$, the sentence is negative, and the sentiment polarity is reversed; if $\lambda = 1$, the sentiment polarity remains unchanged. Note that if the negative words in the short text appear with a certain interval, the change of sentiment polarity may cease to be effective, and such a short text should be filtered out.

Next, a one-dimensional array $ArrSS$ is defined to describe the scores of the synonym set for the word:

$$ArrSS(word) = [synset_1, synset_2, \dots, synset_n] \tag{3}$$

2.3.2 Part-of-speech (POS) labeling

POS labeling aims to mark the words as nouns, verbs, adjectives, adverbs, etc. This paper utilizes the Stanford POS Tagger to process English short texts, and the Language Technology Platform (LTP) Tagger of Harbin Institute of Technology to process Chinese short texts. The taggers adopt Penn Treebank POS labels, and convert them into the labels of nouns, verbs, adjectives, and adverbs (Table 1). This paper only tags nouns, verbs, adjectives, and adverbs, without considering prepositions and numbers.

where, n is the size of the synonym set for the word.

To ensure the correspondence to linguistic terms $L^{(k)}$, $SynsetScore$ is mapped from $[-1, 1]$ to $[-8, -7, \dots, 7, 8]$:

$$f_{trans}(SynsetScore) = ([8 * SynsetScore]) = \alpha \tag{4}$$

In this way, $SynsetScore$ is converted into $\alpha(k)$, $\alpha \in [-8, -7, \dots, 7, 8]$. Thus, $ArrSS$ can be converted to a 1D array of the linguistic term $ArrLT$:

$$ArrLT_{(word)} = [S_{\alpha_1}, S_{\alpha_2}, \dots, S_{\alpha_n}] \tag{5}$$

2.3.3 Stop word removal

In a language, the stop words refer to the words that do not have special meanings, such as “it”, “is”, and “the”. These words are useless in text sentiment analysis. To reduce their effects on computing and storage space, these words are filtered out against the stop word list.

After that, it is possible to calculate the frequency of occurrence S of each linguistic term, and estimate the probability p of each term. The sentiment polarity of a word can be converted into a PLTS:

$$L_{(word)}(p) = \{S_{\alpha}(p) | S_{\alpha} \in S, p \geq 0, \sum p = 1\} \tag{6}$$

2.4 Transforming word/emoji to PLTS

Each word/emoji can be transformed into a PLTS in three steps: conversion, update, and calculation.

2.4.2 Updated SWN

2.4.1 Conversion

Every emoji is a binary vector $\langle img_{emoji}, emoji\ label \rangle$ (Figure 5). If the emoji in the original sample is an image img_{emoji} , it should be converted into $emoji\ label$; otherwise, the $emoji\ label$ can be directly used. Through this mapping, all emojis can be converted into words, and then integrated to the subsequent PLTS calculation process.

The preceding subsection extracts sentiment information from SWM, and converts it into PLTS. But the extracted $PosScore$ and $NegScore$ are not fully accurate. Deviation may arise for several reasons:

(1) The scores calculated by the semi-supervised method may be biased. For example, Table 2 lists the sentiment polarity scores of the word “ridiculous” ($L_{ridiculous}(p) = \{S_0(0.333), S_3(0.333), S_5(0.333)\}$). The data show that “ridiculous” conveys a positive meaning, but this word is actually a negative word.

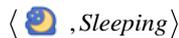


Figure 5. Emoji vector

Then, the difference $SynsetScore$ between positive review score $PosScore$ and negative review score $NegScore$ is calculated, and used to evaluate the sentiment intensity difference between the words in the SWN dataset. The $SynsetScore$ can be expressed as:

Table 2. Sentiment polarity scores of the word “ridiculous”

Synsets	PosScore	NegScore
ridiculous.n.01	0	0
ridiculous.s.01	0.375	0
ridiculous.n.02	0.625	0

(2) The same word appears in different texts at different probabilities. For instance, “good” in SWM has up to ten meanings. Some meanings are common, and some are rare.

The imbalance and sparsity of words and word polarities hinder the polarity judgement. More importantly, SWN does not provide any information about word probability.

(3) The sentiment polarity of words is affected by context. In different contexts, sentiment polarity will transfer or strengthen. The same word may have opposite polarities in different areas.

The above factors may lead to under-fitting. To solve the problem, this paper updates the SWN dictionary by the statistics on the word frequency distribution. On this basis, the context sentiment weight (CSW), which falls between -1 and 1, is introduced to the context:

$$CSW_{word} = \frac{Freq(word|p) - Freq(word|n)}{Freq(word)} \quad (7)$$

where, $Freq(word|p)$ and $Freq(word|n)$ are the frequencies of a word appearing in texts labeled as positive and negative, respectively. Then, the total word frequency can be expressed as $Freq(word) = Freq(word|p) + Freq(word|n)$. For example, $CSW_{best} = 1$ indicates that the word "best" merely appears in texts marked as positive.

Similarly, CSW is mapped from $[-1, 1]$ to $\alpha \in [-8, -7, \dots, 7, 8]$ using f_{trans} , and a word frequency distribution information dictionary is established. The dictionary is named context sentiment weight dictionary (CSWD). Table 3 gives an example of CSWD.

Finally, CSWD is employed to modify the PLTSs extracted from SWN:

$$L_{word}(p) = \left\{ \begin{array}{l} S_{-8} \left(\frac{p}{1+m} \right), S_{-7} \left(\frac{p}{1+m} \right), \dots, \\ \times S_{\alpha}^{word} \left(\frac{p+m}{1+m} \right), \dots, S_8 \left(\frac{p}{1+m} \right) \end{array} \right\} \quad (8)$$

where, $L_{word}(p)$ and S_{α}^{word} are the PLTS and linguistic term of the word, respectively ($Freq(word|p)\alpha$ is looked up in CSWD); $m \in (0, +\infty)$ is an empirical adjustment factor that manually regulates model accuracy. If $m=0$, no adjustment is necessary. The greater the m , the more significant the adjustment. But a large m may result in over-fitting. Here, the PLTS of SWN is modified at $m=1$.

2.4.3 PLTS calculation

Each sentence is a combination of a series of words, which contain (hide) sentiment information. By synthesizing the sentiments of all the words in the series, it is possible to obtain the reduced theme sentiment of the sentence.

Let n be the number of words segmented from a sentence. Then, these words will be converted into n PLTSs. The heavy presence of neutral words adds to the computing load. Thus, the neutral words can be filtered out by the following rule:

$$word \rightarrow L(p) = \{S_{-8}(p), S_{-7}(p), \dots, S_8(p)\}, P_{S_{-1}} + P_{S_0} + P_{S_1} \geq 0.8 \quad (9)$$

where, $P_{S_{\alpha}}$ is the probability of S_{α} .

After the filtering of neutral words, there are m remaining PLTSs. The sentiment polarity of the sentence can be characterized by the mean PL_{avg} of probabilistic language:

Table 3. An example of CSWD

Term	α	Term	α
null	-3	small	-2
uninteresting	-3	often	2
dishonest	-2	sagacious	3
job	0	cheerful	6

Table 4. Input format of SVM

Label	S_0	S_1	...	S_n
-1	0.0751	0.0909	...	0.0880
1	0.0578	0.1680	...	0.2160

$$L_{sentence}(p) = \lambda * PL_{avg}(L(p)) = \lambda * \sum_{i=1}^m L_i(p) = \{S_{\alpha}(p)\} \quad (10)$$

where, $L_{sentence}(p)$ is the PLTS of the sentence. If the sentence is negative, $\lambda = -1$; otherwise, $\lambda = 1$.

Normally, $L_{sentence}(p)$ contains lots of $S_{\alpha}(p)$, and the adjacent S_{α} have a very small difference. To further reduce computing load, S_{α} ($\alpha \in [-8, 8]$) is mapped to S_{α} ($\alpha \in [0, 1, \dots, 100]$), and the $S_{\alpha}(p)$ with the same suffix a is merged.

During the transformation of words into PLTSs, each sentence of a text is expressed as one $L_{sentence}(p)$, which represents the possibility of each sentiment polarity. This approach fully utilizes the sentiment information and word frequency information in the sentiment dictionary.

2.5 SVM-based sentiment classification

According to the PLTS calculation in 2.4, $L_{sentence}(p)$ contains lots of $S_{\alpha}(p)$. Even if repetitive terms are eliminated by mapping, the resulting $S_{\alpha}(p)$ will affect the sentiment polarity of the sentence. To improve classification performance, this paper constructs a classifier based on SVM, a supervised learning strategy using radial basis function (RBF) kernel. Each linguistic term is expressed as a feature, and the probability P corresponding to that term is regarded as the eigenvalue. Table 4 shows the input format of SVM.

Then, $L_{sentence}(p)$ is converted into a format that can be understood by SVM classifier:

$$\langle Label \rangle \left\{ \begin{array}{l} feature_1[\langle index \rangle, \langle value \rangle], \\ feature_2[\langle index \rangle, \langle value \rangle], \\ \dots, \\ feature_n[\dots] \end{array} \right\} \quad (11)$$

where, $\langle Label \rangle$ corresponds to positive class and negative class; $feature$ corresponds to n linguistic terms; $value$ is the probability p of each term.

3. EXAMPLE OF PLTS-FECLSQ

This section chooses a short text from the sample set, and demonstrates how to analyze it under the FECLSQ. The short text is a review of the service of a library, in which "/anger" is the html code of an emoji:

"In its small heart, librarian goes to ridiculous lengths to duck the many questions it raises/anger."

The short text is processed in the following steps:

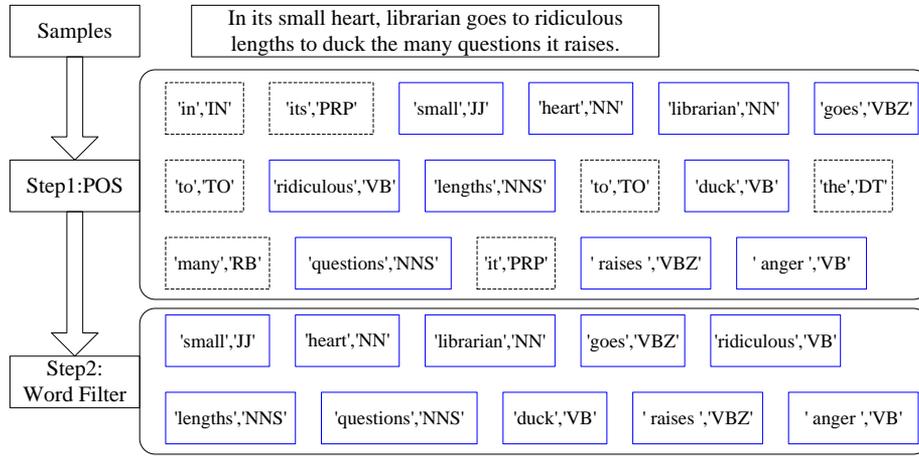


Figure 6. Preprocessing of the emoji-containing short text

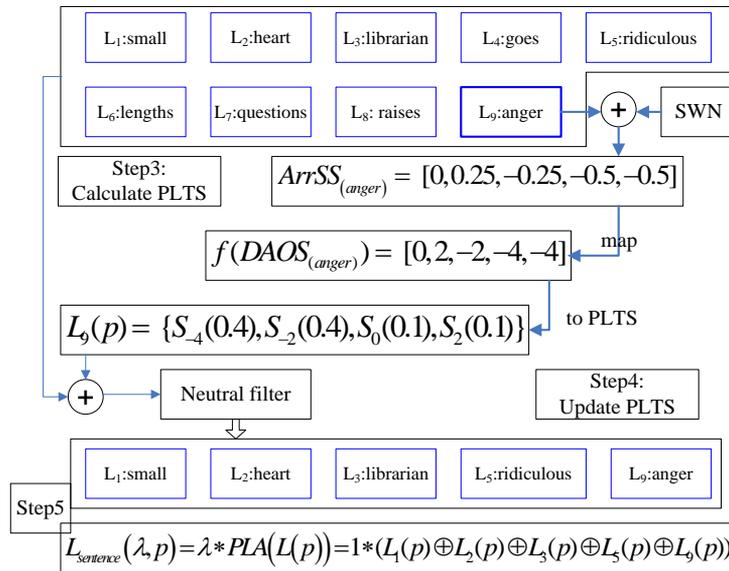


Figure 7. Conversion of words into PLTSs

Step 1. The short text is preprocessed through spelling correction, word segmentation, and POS labeling (Figure 6).

Step 2. The stop words are removed, leaving only the words with labels of verbs, nouns, adjectives, and adverbs.

Step 3. As shown in Figure 7, the sentiment polarity score of the synonym set for each word is extracted, and converted into a PLTS. This step is repeated to compute the PLTS of each word.

Step 4. Referring to CSWD, update the PLTS of each modified word. For simplicity, the neutral words are removed, leaving only five words.

Step 5. The programmable logic array (PLA) operator is called to compute the sentiment polarity $L_{sentence}(p)$ of the sentence, with the negative coefficient being $\lambda = -1$.

Step 6. S_α is mapped into S'_α , $L_{sentence}(p)$ is converted into the input format of SVM, and the results are classified.

4. EXPERIMENTS AND RESULTS ANALYSIS

4.1 Datasets

The performance of PLTS-FECLSQ was tested on three public datasets and two self-built datasets:

(1) Comment Dataset Library Service Quality (CDLSQ)

This dataset of library service quality reviews was established by the authors. There are 8,176 reviews in the datasets. Among them, 4,782 reviews contain emojis.

(2) Movie reviews (MR) [41]

This dataset contains 10,662 movie reviews from Rotten Tomatoes. Half of them is labeled positive, and half are labeled negative. In the original reviews, positive and negative reviews are described by “fresh” and “rotten”, respectively.

(3) Stanford Twitter Sentiment (STS) [42]

The STS contains 1.6 million tweets, which are automatically labeled as positive or negative. This paper randomly selects 20,000 tweets from the dataset.

(4) TripAdvisor reviews (TR) [43]

This paper crawls over 15,000 text reviews from tripadvisor.com, a tourism networking and community network, and obtains 1,100 positive reviews and 950 negative reviews through automatic filtering and artificial evaluation.

(5) DeepMoji [44]

This dataset contains more than 120 million filtered tweets. Referring to the polyseme list of SWN, this paper filters the dataset, and chooses the samples containing only one emoji. After filtering, a self-built dataset DMx was established. Then, DMx was divided randomly into a training set and a test set at the split ratio of 4:1. Table 5 provides the details of the datasets.

4.2 Metrics and results

The classification models were evaluated comprehensively with four metrics: accuracy (Acc), precision (Prec), recall (Rec), and F-measure (F-M) [45]:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (12)$$

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$F - Measure = 2 * \frac{Precision * Recall}{Recall + Precision} \quad (15)$$

where, TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives, respectively; F-measure is the harmonic mean between precision and recall, and an index of comprehensive evaluation.

Firstly, the performance of five mature sentiment dictionaries, namely, GI, TB, VADER, SCN, and SWN, was compared on the datasets. Table 6 shows the test results of these dictionaries, and Figure 8 displays the mean accuracy of each dictionary.

Among all dictionaries, VADER achieved the best performance, followed by TB. SCN differed very slightly from SWN, while GI realized the highest stability. For the five sentiment analysis tools/techniques, the mean accuracy, recall, and F-measure on MR were 58.50%, 69.39%, and 63.36%, respectively; those on STS were 62.36%, 57.92%, and 59.77%, respectively; those on TR were 63.99%, 78.38%, and 70.41%, respectively. The recall was higher than the precision of dictionaries, suggesting that dictionary-based methods are more sensitive to FP than to FN. On the three datasets, the F-M averaged at only 62.5%, and peaked at 77.67%. Thus, dictionary-based methods are far from satisfactory. In addition, the five dictionaries performed poorly on CDLSQ, which is related to their focus on Chinese language.

Furthermore, PLTS-CLSQF was compared with three machine learning methods: NB, SVM, and LR. The classifiers were trained by the BOW [46]. As shown in Table 6 and Table 7, the highest values on the five datasets were 78%, 76%, 79%, 74%, and 88%, respectively. Therefore, machine learning-based methods far outshine dictionary-based methods.

PLTS-FECLSQ, which combines supervised learning with unsupervised learning, improves the best performance on CDLSQ and DMx. The results indicate that the accuracy and F-measure increased with the scope of the training set. PLTS-FECLSQ can effectively solve various problems, using dictionary information, and word frequency distribution. Firstly, the sentiment information of each word is extracted from the current dictionary and word frequency distribution. Each word/emoji is expressed as a PLTS. Next, the PLTSs of all words are synthesized into the PLTS of the sentence $L_{sentence}(p)$. Finally, SVM is called to enhance the classification performance. Compared with unsupervised classifiers, PLTS-CLSQF introduces word frequency distribution and SVM to overcome the problems of neighborhood dependence and contextual information. Compared with supervised classifiers, PLTS-CLSQF can extract the sentiment information of words from the current dictionaries, which is difficult to acquire through data training alone. Our approach can effectively solve the unavailability and sparsity of data. More importantly, the PLTS and related theories are adopted to treat the concurrence of polysemy and emojis.

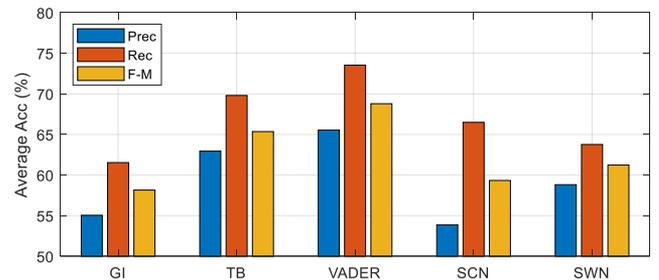


Figure 8. Mean accuracies of the five dictionaries on the datasets

Table 5. Dataset overview

Dataset	Selected dataset	Positive reviews	Negative reviews	Neutral reviews	Mean sentence length
CDLSQ	8176	3116	1666	3394	16
DMx	231938	130251	115038	31063	18
MR	10662	5331	5331	N/A	20
STS	110712	66517	32956	11239	18
TR	2050	1100	950	N/A	22

Table 6. Test results of five dictionaries on the datasets

Dataset	Metric	Sentiment dictionary				
		GI	TB	VADER	SCN	SWN
CDLSQ	Prec	0.4913	0.5549	0.5626	0.4925	0.5064
	Rec	0.581	0.645	0.6851	0.6188	0.5919
	F-M	0.5351	0.5788	0.621	0.5427	0.562
DMx	Prec	0.6277	0.7402	0.7361	0.6045	0.6462
	Rec	0.6943	0.9086	0.8987	0.8163	0.8238
	F-M	0.6283	0.8213	0.8321	0.7086	0.6962
MR	Prec	0.5542	0.6052	0.6419	0.5563	0.5674
	Rec	0.6314	0.7184	0.7551	0.6997	0.6648
	F-M	0.5902	0.6569	0.689	0.6198	0.6122
STS	Prec	0.5674	0.6683	0.7275	0.5071	0.6475

	Rec	0.5969	0.5623	0.6124	0.5902	0.5341
	F-M	0.5818	0.6107	0.665	0.5445	0.5854
	Prec	0.5892	0.6896	0.6897	0.5992	0.6316
TR	Rec	0.6524	0.8667	0.8888	0.7511	0.76
	F-M	0.6192	0.768	0.7767	0.6666	0.6899

Table 7. Comparison between machine learning and PLTS-FECLSQ

Dataset	Metric	Method			
		NB	SVM	LR	PLTS-FECLSQ
CDLSQ	Prec	0.67	0.69	0.68	0.82
	Rec	0.71	0.71	0.71	0.74
	F-M	0.7	0.7	0.71	0.78
DMx	Prec	0.7	0.68	0.7	0.68
	Rec	0.82	0.82	0.83	0.79
	F-M	0.74	0.74	0.74	0.76
MR	Prec	0.75	0.75	0.76	0.72
	Rec	0.78	0.76	0.77	0.76
	F-M	0.76	0.75	0.76	0.79
STS	Prec	0.67	0.65	0.68	0.72
	Rec	0.79	0.82	0.79	0.77
	F-M	0.71	0.72	0.72	0.74
TR	Prec	0.87	0.86	0.87	0.89
	Rec	0.86	0.86	0.88	0.90
	F-M	0.86	0.86	0.87	0.88

5. CONCLUSIONS

This paper combines PLTS and SVM into a short text sentiment analysis framework called PLTS-FECLSQ. Under the framework, the sentiment information of each word is extracted from SWN and a word frequency distribution dictionary, and converted into a PLTS. The PLTSs of words are integrated into the sentiment information of the short text. Finally, the classification performance is enhanced by SVM. Experimental results show that PLTS-FECLSQ improves the performance of short text sentiment evaluation.

Our results show that PLTS-FECLSQ performs well in coarse-grained sentiment classification. However, more works need to be done to extend and apply the framework to fine-grained sentiment classification. Some new analysis tools have been developed to process the ambiguity of human language, namely, intuitive language method [47], gray language method [48], and interval language method [49]. These methods help to improve the analysis accuracy in specific scenarios.

ACKNOWLEDGMENTS

This project was funded by the Humanities and Social Science Research Project of Ministry of Education of China (Grant No.: 20YJCZH081), and Scientific Research Project of Education Department of Hubei Province (Grant No.: D20212701, D20202701, B2020385).

REFERENCES

[1] Wang, Z., Ho, S.B., Cambria, E. (2020). A review of emotion sensing: Categorization models and algorithms. *Multimedia Tools and Applications*, 79(47): 35553-35582. <https://doi.org/10.1007/s11042-019-08328-z>

[2] Wang, L., Wang, X.K., Peng, J.J., Wang, J.Q. (2020). The differences in hotel selection among various types of travellers: A comparative analysis with a useful bounded

rationality behavioural decision support model. *Tourism management*, 76: 103961. <https://doi.org/10.1016/j.tourman.2019.103961>

[3] Ali, F., Kwak, D., Khan, P., et al. (2019). Transportation sentiment analysis using word embedding and ontology-based topic modeling. *Knowledge-Based Systems*, 174: 27-42. <https://doi.org/10.1016/j.knosys.2019.02.033>

[4] Li, J., Fong, S., Zhuang, Y., Khoury, R. (2016). Hierarchical classification in text mining for sentiment analysis of online news. *Soft Computing*, 20(9): 3411-3420. <https://doi.org/10.1007/s00500-015-1812-4>

[5] Tan, S., Yang, L., Sun, H., Guan, Z., He, X. (2014). Interpreting the public sentiment variations on twitter. *IEEE Transactions on Knowledge and Data Engineering*, 26(5): 1158-1170. <https://doi.org/10.1109/TKDE.2013.116>

[6] Vlieger, E., Leydesdorff, L. (2011). Content analysis and the measurement of meaning: The visualization of frames in collections of messages. *Public Journal of Semiotics*, 3(1): 28-50. <https://doi.org/10.37693/pjos.2011.3.8830>

[7] Stone, P.J., Dunphy, D.C., Smith, M.S., Ogilvie, D.M. (1966). *General Inquirer*. Philosophy.

[8] Turney, P., Littman, M.L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. on Information Systems*, 21(4): 315-346. <https://doi.org/10.1145/944012.944013>

[9] Subirats, L., Reguera, N., Bañón, A.M., Gómez-Zúñiga, B., Minguillón, J., Armayones, M. (2018). Mining Facebook data of people with rare diseases: A content-based and temporal analysis. *International Journal of Environmental Research and Public Health*, 15(9): 1877. <https://doi.org/10.3390/ijerph15091877>

[10] Ribeiro, F.N., Araújo, M., Gonçalves, P., Gonçalves, M. A., Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1): 23. <https://doi.org/10.1140/epjds/s13688-016-0085-1>

[11] Baccianella, S., Esuli, A., Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, pp. 2200-2204.

[12] Cambria, E., Speer, R., Havasi, C., Hussain, A. (2010). Senticnet: A publicly available semantic resource for opinion mining. In *2010 AAAI Fall Symposium Series*, pp. 14-18.

[13] Xie, X., Ge, S., Hu, F., Xie, M., Jiang, N. (2017). An improved algorithm for sentiment analysis based on maximum entropy. *Soft Computing*, 23: 59-611. <https://doi.org/10.1007/s00500-017-2904-0>

[14] Xue, J.H., Titterton, D.M. (2008). Comment on "on discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes". *Neural Processing Letters*, 28(3): 169-187. <https://doi.org/10.1007/s11063-008-9088-7>

[15] Walker, S.H., Duncan, D.B. (1967). Estimation of the probability of an event as a function of several

- independent variables. *Biometrika*, 54(1-2): 167-179. <https://doi.org/10.1093/biomet/54.1-2.167>
- [16] Long, T., Long, W.B. (2011). Test generation algorithm for analog systems based on support vector machine. *Signal, Image & Video Processing*, 5: 527-533. <https://doi.org/10.1007/s11760-010-0168-6>
- [17] Jalilvand, A., Salim, N. (2012). Sentiment classification using graph based word sense disambiguation. *International Conference on Advanced Machine Learning Technologies and Applications*, Cairo, Egypt, pp. 351-358. https://doi.org/10.1007/978-3-642-35326-0_35
- [18] Khan, F.H., Qamar, U., Bashir, S. (2016). eSAP: A decision support framework for enhanced sentiment analysis and polarity classification. *Information Sciences*, 367: 862-873. <https://doi.org/10.1016/j.ins.2016.07.028>
- [19] Li, M., Ch'Ng, E., Chong, A., See, S. (2018). Multi-class twitter sentiment classification with emojis. *Industrial Management & Data Systems*, 118(9): 1804-1820. <https://doi.org/10.1108/IMDS-12-2017-0582>
- [20] Illendula, A., Manohar, K., Yedulla, M.R. (2018). Which emoji talks best for my picture? In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Santiago, pp. 514-519. <https://doi.org/10.1109/WI.2018.00-44>
- [21] Derks, D., Bos, A., Grumbkow, J.V. (2007). Emoticons and social interaction on the internet: the importance of social context. *Computers in Human Behavior*, 23(1): 842-849. <https://doi.org/10.1016/j.chb.2004.11.013>
- [22] Riordan, M.A. (2017). Emojis as tools for emotion work: Communicating affect in text messages. *Journal of Language and Social Psychology*, 36(5): 549-567. <https://doi.org/10.1177/0261927X17704238>
- [23] Sakariassen, H. (2021). Women's emotion work on Facebook: Strategic use of emotions in public discourse. *Computers in Human Behavior Reports*, 4: 100148. <https://doi.org/10.1016/j.chbr.2021.100148>
- [24] Zapf, D., Kern, M., Tschan, F., Holman, D., Semmer, N.K. (2020). Emotion work: A work psychology perspective. *Annual Review of Organizational Psychology and Organizational Behavior*, 8(1): 139-172. <https://doi.org/10.1146/annurev-orgpsych-012420-062451>
- [25] Pang, Q., Wang, H., Xu, Z. (2016). Probabilistic linguistic term sets in multi-attribute group decision making. *Information Sciences*, 369: 128-143. <https://doi.org/10.1016/j.ins.2016.06.021>
- [26] Gou, X., Xu, Z. (2016). Novel basic operational laws for linguistic terms, hesitant fuzzy linguistic term sets and probabilistic linguistic term sets. *Information Sciences*, 372: 407-427. <https://doi.org/10.1016/j.ins.2016.08.034>
- [27] Gou, X., Liao, H., Xu, Z., Herrera, F. (2017). Double hierarchy hesitant fuzzy linguistic term set and multimooora method: a case of study to evaluate the implementation status of haze controlling measures. *Information Fusion*, 38: 22-34. <https://doi.org/10.1016/j.inffus.2017.02.008>
- [28] Zhang, Y., Xu, Z., Wang, H., Liao, H. (2016). Consistency-based risk assessment with probabilistic linguistic preference relation. *Applied Soft Computing*, 49: 817-833. <https://doi.org/10.1016/j.asoc.2016.08.045>
- [29] Peng, H.G., Zhang, H.Y., Wang, J.Q. (2018). Cloud decision support model for selecting hotels on tripadvisor.com with probabilistic linguistic information. *International Journal of Hospitality Management*, 68: 124-138. <https://doi.org/10.1016/j.ijhm.2017.10.001>
- [30] Liao, H., Jiang, L., Xu, Z., Xu, J., Herrera, F. (2017). A linear programming method for multiple criteria decision making with probabilistic linguistic information. *Information Sciences*, 415: 341-355. <https://doi.org/10.1016/j.ins.2017.06.035>
- [31] Tian, Z.P., Nie, R.X., Wang, J.Q. (2020). Probabilistic linguistic multi-criteria decision-making based on evidential reasoning and combined ranking methods considering decision-makers' psychological preferences. *Journal of the Operational Research Society*, 71(5): 700-717. <https://doi.org/10.1080/01605682.2019.1632752>
- [32] Luo, S.Z., Zhang, H.Y., Wang, J.Q., Li, L. (2019). Group decision-making approach for evaluating the sustainability of constructed wetlands with probabilistic linguistic preference relations. *Journal of the Operational Research Society*, 70: 2039-2055. <https://doi.org/10.1080/01605682.2018.1510806>
- [33] Krishankumar, R., Ravichandran, K., Ahmed, M., Kar, S., Tyagi, S. (2018). Probabilistic linguistic preference relation-based decision framework for multi-attribute group decision making. *Symmetry*, 11(1): 2. <https://doi.org/10.3390/sym11010002>
- [34] Wu, X., Liao, H. (2018). A consensus-based probabilistic linguistic gained and lost dominance score method. *European Journal of Operational Research*, 272: 1017-1027. <https://doi.org/10.1016/j.ejor.2018.07.044>
- [35] Schuckert, M., Liu, X., Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, 32(5): 608-621. <https://doi.org/10.1080/10548408.2014.933154>
- [36] Proksch, S.O., Lowe, W., Wäckerle, J., Soroka, S. (2018). Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. *Legislative Studies Quarterly*, 44(1): 97-131. <https://doi.org/10.1111/lsq.12218>
- [37] Dong, Y., Wu, Y., Zhang, H., Zhang, G. (2015). Multi-granular unbalanced linguistic distribution assessments with interval symbolic proportions. *Knowledge-Based Systems*, 82: 139-151. <https://doi.org/10.1016/j.knsys.2015.03.003>
- [38] Wu, Z., Xu, J. (2016). Possibility distribution-based approach for MAGDM with hesitant fuzzy linguistic information. *IEEE Transactions on Cybernetics*, 46(3): 694-705. <https://doi.org/10.1109/TCYB.2015.2413894>
- [39] Liu, H., Rodríguez, R.M. (2014). A fuzzy envelope for hesitant fuzzy linguistic term set and its application to multicriteria decision making. *Information Sciences*, 258: 220-238. <https://doi.org/10.1016/j.ins.2013.07.027>
- [40] Yang, J.B. (2001). Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties. *European Journal of Operational Research*, 131(1): 31-61. [https://doi.org/10.1016/S0377-2217\(99\)00441-5](https://doi.org/10.1016/S0377-2217(99)00441-5)
- [41] Pang, B., Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *arXiv preprint cs/0506075*.
- [42] Sahni, T., Chandak, C., Chedeti, N.R., Singh, M. (2017). Efficient Twitter sentiment classification using subjective distant supervision. In 2017 9th International Conference on Communication Systems and Networks (COMSNETS), Bengaluru, India, pp. 548-553.

- <https://doi.org/10.1109/COMSNETS.2017.7945451>
- [43] Cenni, I., Goethals, P. (2017). Negative hotel reviews on TripAdvisor: A cross-linguistic analysis. *Discourse, Context & Media*, 16: 22-30. <https://doi.org/10.1016/j.dcm.2017.01.004>
- [44] Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. arXiv preprint arXiv:1708.00524.
- [45] Pastor-Pellicer, J., Zamora-Martínez, F., España-Boquera, S., Castro-Bleda, M.J. (2013). F-measure as the error function to train neural networks. In *International Work-Conference on Artificial Neural Networks*, Puerto de la Cruz, Tenerife, Spain, pp. 376-384. https://doi.org/10.1007/978-3-642-38679-4_37
- [46] Larlus, D., Verbeek, J., Jurie, F. (2010). Category level object segmentation by combining bag-of-words models with Dirichlet processes and random fields. *International Journal of Computer Vision*, 88(2): 238-253. <https://doi.org/10.1007/s11263-009-0245-x>
- [47] Wang, J.Q., Han, Z.Q., Zhang, H.Y. (2014). Multi-criteria group decision-making method based on intuitionistic interval fuzzy information. *Group Decision & Negotiation*, 23(4): 715-733. <https://doi.org/10.1007/s10726-012-9316-4>
- [48] Tian, Z.P., Wang, J., Wang, J.Q., Chen, X.H. (2018). Multicriteria decision-making approach based on gray linguistic weighted Bonferroni mean operator. *International Transactions in Operational Research*, 25(5): 1635-1658. <https://doi.org/10.1111/itor.12220>
- [49] Wang, J.Q., Peng, J.J., Zhang, H.Y., Tao, L., Chen, X.H. (2015). An uncertain linguistic multi-criteria group decision-making method based on a cloud model. *Group Decision & Negotiation*, 24(1): 171-192. <https://doi.org/10.1007/s10726-014-9385-7>