



Novel feature selection approaches for improving the performance of sentiment classification

Jing-Rong Chang¹ · Hsin-Ying Liang² · Long-Sheng Chen¹ · Chia-Wei Chang¹

Received: 17 April 2020 / Accepted: 8 August 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Text based social media has become one of important communication tools between customers and enterprises. In social media, users can easily express their opinions and evaluation regarding products or services. These online user experiences, especially negative evaluations indeed affect other consumers' behaviors. Consequently, to effectively identify customers' sentiments and avoid these negative comments to bring a great damage to enterprisers has become one of critical issues. In recent years, machine learning algorithms were viewed as one of effective solutions for sentiment classification. But, when the amount of the online reviews arises, the dimensionality of text data rises remarkably. The performances of machine learning methods have been degraded due to the dimensionality problem. But, conventional feature selection methods tend to select attributes from the majority sentiments, which usually cannot improve classification performance. Therefore, this study attempt to present two feature selection methods called modified categorical proportional difference (MCPD) approach that improves conventional CPD method, and balance category feature (BCF) strategy that equally selects attributes from both positive and negative examples, to improve sentiment classification performances. Finally, several real sentiment cases of text reviews will be provided to demonstrate the effectiveness of our proposed methods. Results showed that the combination of proposed BCF strategy and MCPD method can not only remarkably reduce feature space, but also improve the sentiment classification performance.

Keywords Feature selection · Sentiment classification · Social media · Support vector machines · Online reviews

1 Introduction

Social media such as Twitter, Telegram, Facebook, Tik Tok, and so on has changed the communication ways to be more convenient and ubiquitous (Singh et al. 2020). Actually, social media has become one of crucial channels to express personal feelings, opinions, and communicating to other users, to teach students, and even to support mental health specialists to diagnose depression (Denecke and Nejdil 2009; Zhu et al. 2011; Giuntini et al. 2020). Therefore, personal comments regarding products or services have strong

influence on purchasing decision making of other social media participators (Kumar et al. 2006; Zhang et al. 2007; Chang et al. 2020).

Comments in social media could be a major information source of providing suggestions and recommendations regarding commercial products from the customer perspectives (Bai 2011). But, some comments are often harmful and then to reduce purchase intentions. Based on a research online report of Lightspeed company (2011), about 60% of customers will change their purchase decisions after reading 1–3 negative comments. But, these harmful reviews could be viewed as customers' complaints which might provide useful information to improve enterprisers' services. In addition, these online reviews are usually unstructured, subjective, and hard to comprehend in short time. Therefore, how to recognize social media users' sentiments from a huge amount of online reviews has become one of important issues (Chen et al. 2011).

Now, sentiment classification becomes more important when the number of digital text resources rises (Gokalp et al.

✉ Long-Sheng Chen
lschen@cyut.edu.tw

¹ Department of Information Management, Chaoyang University of Technology, 168 Jifong E. Rd., Wufong District, Taichung 41349, Taiwan

² Department of Information and Communication Engineering, Chaoyang University of Technology, 168 Jifong E. Rd., Wufong District, Taichung 41349, Taiwan

2020; Chouchani and Abed 2020). In recent years, sentiment classification which classifies textual sentiment into positive or negative group has attracted lots of attention (Zhao et al. 2020; Kong et al. 2020). Generally speaking, sentiment classification aims to recognize reviewer's sentiments from text comments of customers for specific products or services (Chen et al. 2009; Ye et al. 2009; Mekawie and Hany 2019; Akhtar et al. 2020). Lots of studies have focus on conducting textual sentiment classification. Sentiment classification can also detect social media users' emotions to help enterprises to respond to customers' comments carefully.

From available literatures, machine learning algorithms are widely used to solve this critical problem (Chaovalit and Zhou 2005; Tang et al. 2009; Tan and Zhang 2008; Wu et al. 2006). Machine learning aims to construct classification models from text reviews, and then to recognize the new coming review's sentiment by using the built models. According to published studies, these kinds of methods have been viewed as an effective solution. However, the high dimensionality problem of text data will decrease the classification performance and result in long learning time (Wang et al. 2011). Consequently, to quickly and easily reduce the dimensionality of text data and to retain the performances of classifier have to be solved.

Lots of works attempt to solve high dimensionality problems by integrating dimension reduction techniques into machine learning methods. For examples, Liu (2020) proposed a sentiment analysis model which combines bag of words and convolutional neural network (CNN) to increase the classification performance. Kim (2018) presented a semi-supervised dimension reduction framework which is mainly based on linear feature extraction. Liu et al. (2017) combined feature selection algorithm and machine learning method to propose a framework for multi-class sentiment classification. Khan et al. (2016) introduced a new framework called SWIMS to determine the feature weight based on sentiment lexicon, SentiWordNet. Liu et al. (2017) proposed a framework which combines feature selection algorithm and machine learning method for multi-class sentiment classification. However, traditional feature selection tends to select features from the majority sentiments, which usually cannot improve the performances of classifiers. And, these methods usually need a lot of computational cost. Therefore, we need a feature selection method which can quickly pick up crucial features and then build term-document matrix (TDM) based on them.

Consequently, the major purpose of this work is to develop effective feature selection methods to improve sentiment classification performance and avoid negative sentiments to bring a great damage to enterprisers. This study will propose two feature selection methods called modified categorical proportional difference (MCPD) and balance category feature (BCF) strategy which equally selects features

from both positive and negative sentiments for improving the performance of classifying sentiments. Finally, some real text sentiment cases of customers' comments will be provided to illustrate the effectiveness of our proposed methods.

2 Related works

2.1 Feature selection methods in sentiment classification

Sentiment classification has become very important when the amount of digital text resources remarkably increases (Gokalp et al. 2020; Chouchani and Abed 2020). The purpose of sentiment analysis is to analyze the publics' sentiments, opinions, attitudes, emotions, and so on, towards different elements such as topics, products or services, individuals, or organizations (Liu et al. 2005; Khan et al. 2016; Singh et al. 2020).

According to available works, machine learning method has been report as one of effective solutions. For instances, Dave et al. (2003) used feature selections and scoring methods for sentiment classification for online reviews. Based on extracting and analyzing appraisal groups, whitelaw et al. (2005) proposed a new method containing support vector machines (SVM) to sentiment analysis. Abbasi et al. (2008a) proposed entropy weighted genetic algorithm (EWGA) method with support vector machines (SVM) for recognizing sentiments of movie reviews. Abbasi et al. (2008b) developed SVRCE method to identify emotional states. O'Keefe and Koprinska (2009) used Naive Bayes and SVM in sentiment analysis. But, when using machine learning approaches to deal with text data, we should consider the dimensionality problems. Consequently, feature selection methods which aim to discover important features from the huge amount of candidate attributes, and achieve a goal of dimension reduction in a short term, should be taken into consideration.

Social media data has a curse of dimension problem (Singh et al. 2020), because a large number of text reviews for sentiment analysis entails huge complexity and cost (Kim, 2018). Therefore, for those high dimensional data, it required specific pre-processing and dimension reduction, which leads to improve computational cost (Singh et al. 2020). Xu et al. (2020) also thought the computational efficiency to process a huge amount of text reviews and the ability to continuously learn from increasing reviews are the major problems for sentiment classification. Among dimension reduction techniques, feature selection methods are one of popular used methods.

General speaking, feature selection approaches were widely used to decrease computational cost and to delete unimportant features for improving classification

performance (Li et al. 2007). Feature selection could obtain a high-quality minimal feature subset (Yousefpour et al. 2017). In related works, lots of methods have been proposed for dimension reduction in sentiment classification. For instances, In the work of Liu et al. (2017), they compared four feature selection algorithms (document frequency, CHI statistics, information gain and gain ratio) and five machine learning algorithms (decision tree, naïve Bayes, support vector machine, radial basis function neural network and K-nearest neighbor). Results indicated that gain ratio and support vector machine have the best performance. Akhtar et al. (2017) developed a framework of feature selection and classifier ensemble using particle swarm optimization (PSO) for aspect based sentiment analysis. Yousefpour et al. (2017) showed part-of-speech (POS) patterns are more effective in their classification accuracy compared to unigram-based features.

To sum up, feature selection algorithms can result in good performance but they also need lots of computational cost. For text data, we need other feature selection methods to quickly select important terms and then construct term-document matrix (TDM) based on them. To avoid confusing readers, we use “term selection” instead of using feature selection to denote dimension reduction tools for sentiment classification.

2.2 Term selection method

In this work, we separate feature selection into two types, term selection which uses metrics to quickly reduce feature space, and traditional feature selection which needs lots of computational cost and has good classification performance. Term selection aims to extract important and relevant attributes (key words) to describe collected documents from a huge amount of candidate attributes and achieve a goal of dimension reduction in a short term. Usually, unlike conventional feature selection algorithms

which can result in good performance but they also need lots of computational cost, term selection methods in text classification need to quickly select important term to construct TDM.

Usually, we only set a threshold of DF (document frequency) or TF-IDF (term frequency–inverse document frequency) to select important features. If one feature whose DF or TF-IDF is below this threshold, this attribute will be considered as irrelevant. Other studies tried to use POS tagging to pick up crucial features for sentiment classification. But, till now, this kind approach cannot result in significant improvement of performance (Na et al. 2005; Chen and Su 2008).

Other approaches calculate a score for each individual features and then select a predefined amount of feature set based on the rank of scores, such as Chi-square statistic (CHI), information gain (IG) and so on (Keshtkar and Inkpen 2009; O’Keefe and Koprinska 2009; Simeon and Hilderman 2008; Tan and Zhang 2008; Ye et al. 2009). From Table 1, we can know these kinds of methods are effective in some experiments. Zheng et al. (2004) indicated that there are two groups of feature selection methods, one-sided (e.g. correlation coefficient and odds ratios) and two-sided (e.g. IG and CHI).

Among them, IG is the most widely used approach and it has been viewed as effective for classifying documents. For instances, Tan and Zhang (2008) indicated IG outperformed document frequency (DF), MI, and CHI when building SVM classifiers. In the study of Ye et al. (2009), they integrated IG into SVM, Naïve Bayes, and N-gram model to identify sentiments of travellers. An improved Fisher’s discriminant ratio (FLDA) developed by Wang et al. (2011) for feature selection. Zheng et al. (2004) employed signed indexes to handle class imbalance problems in text categorization. Singh et al. (2020) aims to find optimal combination of machine learning (SVM, Navies Bayes, linear regression and random forest) and feature

Table 1 Related works of term selection and machine learning methods in sentiment classification

Related works	Machine learning	Term selection methods
O’Keefe and Koprinska (2009)	Naive Bayes (NB) and SVM	CPD, SWN, subjectivity scores
Abbasi et al. (2008a)	SVM	Entropy weighted genetic algorithm (EWGA), genetic algorithm (GA), IG
Tang and Zhang (2009)	Centroid classifier, KNN, NB, Winnow classifier, SVM	DF, CHI, MI, IG
Ye et al. (2009)	NB, SVM, N-gram model	IG
Bai (2011)	NB, SVM, Maximum Entropy (ME)	Tabu search-Markov blanket classification (TS-MBC), IG
Wang et al. (2011)	SVM	FLDA, IG
Ogura et al. (2011)	KNN	Gini, CHI, IG, Signed IG, Signed CHI
Singh et al. (2020)	SVM, NB, Linear Regression and Random Forest	POS, BOW and HASS tagging
Yousefpour et al. (2017)	SVM, NB, ME and LDF	part-of-speech (POS), unigram

extraction techniques (POS, BOW and HASS tagging). They indicated that random forest and linear regression provide the better result with Hass tagging.

When using two-sided feature selection methods to process data of binary classes, the selected features also have the problem of biasing to a certain class. Therefore, this study proposes the balance category feature (BCF) strategy. It is expected that when two-sided methods are used for feature selection, the class distribution of the features will be taken into account at the same time to further improve the classification efficiency.

2.3 Categorical proportional difference (CPD)

CPD (Simeon and Hilderman 2008) is another easy term selection method for multi-class classification problems. O’Keefe and Koprinska (2009) employed CPD on binary sentiment classification. CPD can be defined in Eq. (1):

$$CPD = \frac{|PositiveDF - NegativeDF|}{PositiveDF + NegativeDF} \quad (1)$$

where ‘Positive DF’ represents the positive document frequency and ‘Negative DF’ means the negative document frequency.

CPD attempts to compute Positive DF and Negative DF of one term individually, and next it calculates the proportional difference of one term in both positive and negative classes. The CPD score will locate into [0, 1] interval. If one feature only appears in positive document or negative document, the CPD score is equal to 1. Then, this feature will be considered as important. On the other hand, if a feature appears equally in positive and negative documents, the CPD score is equal to 0. And this feature will be viewed as unimportant. Practically, CPD can discover the useful attributes. However, when using CPD, the dimensionality space of text data still is too large to be solved.

CPD indeed could consider the class information and select relevant attributes effectively. But, the important attributes might be deleted when the dimensionality space of training data is low. To demonstrate this disadvantage, we take Table 2 for example. In this example, there are six candidate features. It can be found that feature A is more relevant than others, but all features have the same CPD score. We cannot know which one should be selected, if we merely use CPD. Therefore, the important feature A might be removed, if we try to use lower dimension of training documents. That’s the reason why we proposed MCPD to enhance CPD.

Besides, in the work of Zheng et al. (2004), they indicated conventional term selection methods tend to select attributes from majority examples. Therefore, they proposed Sign-IG combining sign metric and IG to classify imbalanced text

Table 2 An illustrative example of drawbacks of CPD

	Positive DF	Negative DF	CPD	MCPD
Feature A	80	0	1	80
Feature B	15	0	1	15
Feature C	0	11	1	11
Feature D	2	0	1	2
Feature E	1	0	1	1
Feature F	1	1	0	0

data. Signed IG and signed CHI also had been employed for imbalanced text data (Ogura et al. 2011). Wang et al. (2011) proposed an improved FLDA and compared to IG. In works of Ye et al. (2009) and Tan and Zhang (2008), they indicated integrating IG into SVM can have an optimal performance. Consequently, this study modified the Sign index to classify candidate features into positive and negative sets, and then equally select important ones from both set according to IG and FLDA. The results will be compared with traditional IG and FLDA.

Therefore, CPD which introduces class information has been employed to select important terms. In practice, CPD is very easy to be used and it can effectively extract crucial features in practice. However, CPD cannot dramatically reduce the size of feature sets when applying it to real world.

2.4 TF and TF-IDF

After segmenting words, TF and TF-IDF are kinds of weights in describing text data. Every document could be views as an attribute vectors with these weights (Zhang et al. 2007). Using TF or TF-IDF, we can build term-document matrix (TDM). Some term weights are widely used in text classification, including term frequency (TF), inverse document frequency (IDF), term frequency-inverse document frequency weights (TF-IDF), feature presence (FP), and so on. Among these weight methods, TF and TF-IDF are the most popular a widely used in the related areas of text mining (Aizawa 2003; Na et al. 2005; O’Keefe and Koprinska 2009; Tan and Zhang 2008). The definition of TF-IDF can be found in Eq. (2):

$$tf - idf = tf \times idf \quad (2)$$

IDF is defined as Eq. (3):

$$idf = \log \frac{\text{The number of total documents}}{\text{The number of documents include a term } t}. \quad (3)$$

In Eq. (2), TF and IDF mean term frequency and the general importance of a term in overall documents, respectively. If a feature’s score of TF or TF-IDF is higher, it represents that the feature occurs frequently in documents. In this work,

we use the TF-IDF to count the weights of a feature in a document.

Since TF and TF-IDF are the methods for representing attributes' weights in TDM, these two term weighting methods are also the most widely used and simplest techniques for selecting important features in text data. TF indicates the amount of occurrences frequency of a feature. Because TF is easy to be computed, many studies in text mining utilize this method. In this study, we called "FF" method which uses TF as threshold to remove irrelevant features (Keshtkar and Inkpen 2009; Na et al. 2005; O'Keefe and Koprinska 2009; Pang et al. 2002). TF-IDF is another popular term weighting technique. So, using TF-IDF to extract relevant attributes, which was called "TI" method in this study, is also very common. In both methods, we utilize the two methods for extracting attributes by removing unimportant features whose TF or TF-IDF are below the set thresholds. If those features whose weights (TF or TF-IDF) are larger than the pre-defined threshold, they will be kept for further learning and the rests will be removed.

2.5 Support vector machines

SVM is a successful classifier developed by Vapnik (1995). It also has been widely applied to related areas in sentiment classification. For examples, Akhtar et al. (2017) used maximum entropy (ME), conditional random field (CRF) and support vector machine (SVM) for aspect based sentiment analysis. Liu et al. (2017) indicated that support vector machine have the best performance compared to naïve Bayes, decision trees, neural networks and K-nearest neighbor in sentiment classification. SVM have been employed to classify sentiment of online comments regarding travel destinations, product, and movies (Tan and Zhang 2008; Na et al. 2005; O'Keefe and Koprinska 2009). Song et al. (2020) proposed a SVM based sentiment classification model by introducing probabilistic linguistic terms sets.

In the work of Alqaryouti et al. (2019), they attempt to help government entities gain insights on the expectations of their customers from reviews. They found that using lexicons and rules as input features to the SVM model has achieved higher accuracy than other SVM models. To enhance the performance of sentiment analysis, Hassonah et al. (2020) presented a hybrid machine learning approach which integrates two feature selection techniques using the ReliefF and multi-verse optimizer (MVO) algorithms into SVM.

From these published works, it's reported that SVM had a superior performance in sentiment classification. Besides, SVM have several advantages include the use of kernels, the absence of local minima, the sparseness of solution and the generalization capability obtained by optimizing the margin (Cerqueira et al. 2008). For these reasons, SVM has been employed to be the learner in this study.

Briefly speaking, SVM constructs a decision boundary between two classes by mapping the training data onto a higher dimensional space via kernel functions, and then finding the maximal margin hyperplane within that space. This hyperplane can thus be viewed as a classifier (Cortes and Vapnik 1995). A brief introduction of SVM operations have been given as follows.

Giving n examples $S = \{x_i, y_i\}_{i=1}^n$, $y_i \in \{-1, +1\}$, where x_i represents the condition attributes, y_i is the class label, and i is the number of examples. The decision hyperplane of SVM can be defined as (w, b) , where w is a weight vector and b a bias. Let w_0 and b_0 denote the optimal values of the weight vector and bias. Correspondingly, the optimal hyperplane can be written as

$$w_0^T x + b_0 = 0 \tag{4}$$

To find the optimum values of w and b , it is required to solve the following optimization problem.

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i$$

$$\text{Subject to } \begin{aligned} y_i(w^T \varphi(x_i) + b) &\geq 1 - \xi_i \\ \xi_i &\geq 0 \end{aligned}$$

where ξ is the slack variable, C is the user-specified penalty parameter of the error term ($C > 0$), and φ is the kernel function.

SVM can change the original non-linear separation problem into a linear separation case by mapping input vector onto a higher feature space. On the feature space, the two-class separation problem is reduced to find the optimal hyperplane that linearly separates the two classes transformed into a quadratic optimization problem. In addition, several popular kernel functions including linear, polynomial, radial basis function (RBF), and sigmoid have been used in related works. According to suggestions in the work of Hsu et al. (2006), RBF kernel function is employed in this study.

3 Proposed methodology

The main objective of this work is to develop two feature selection methods for increasing the performance of sentiment classification.

3.1 The proposed MCPD feature selection metric

In practice, CPD cannot greatly reduce the size of feature space when applying it to real world, even it can get the important attributes. To enhance CPD, we revise the original CPD by introducing variation of positive document frequency (PDF) and negative document frequency

(NDF). Before defining MCPD, we let $d_{P,i}(1, 2, \dots, m)$ and $d_{N,j}(1, 2, \dots, n)$ represent the i th positive document and the j th negative document respectively. Random variables $d_{P,i}(t_k)$ and $d_{N,j}(t_k)$ defined as Eqs. (5) and (6) denote a specific feature t_k appeared in the i th positive document and the j th negative document, individually.

$$d_{P,i}(t_k) = \begin{cases} 1 & \text{if } t_k \text{ occurs in } d_{P,i} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$d_{N,j}(t_k) = \begin{cases} 1 & \text{if } t_k \text{ occurs in } d_{N,j} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Besides, let m_1 denote the PDF of feature t_k , which means feature t_k 's occurrence frequency in positive documents, and m_2 be NDF of feature t_k . CPD can be defined as Eq. (9).

$$m_1 = \sum_{i=1}^m d_{P,i}(t_k) \quad (7)$$

$$m_2 = \sum_{j=1}^n d_{N,j}(t_k) \quad (8)$$

$$CPD = \frac{|m_1 - m_2|}{m_1 + m_2} \quad (9)$$

After introducing variation of PDF and NDF to the original CPD metric, the proposed MCPD could be defined as Eq. (10).

$$MCPD = \sqrt{\frac{\left(m_1 - \frac{m_1+m_2}{2}\right)^2 + \left(m_2 - \frac{m_1+m_2}{2}\right)^2}{2}} \times \frac{|m_1 - m_2|}{m_1 + m_2} \quad (10)$$

The proposed MCPD will be compared with CPD, IG and FLDA. The implemental procedure follows Fig. 1.

3.2 The proposed BCF strategy

The second objective of this work is to propose a balancing category features (BCF) strategy. Before introducing our BCF strategy, we need to discuss ‘‘positive features’’ and ‘‘negative features’’. A feature’s sign listed in Eq. (11) can be used to determine one feature tends to positive or negative class. In this study, we use F score in Eq. (12) to determine one feature is positive or negative. For example, if one feature’s F score is +1 (-1), then this feature will be considered as ‘‘positive’’ (‘‘negative’’).

$$Sign = m_1(n - m_2) - m_2(m - m_1) \quad (11)$$

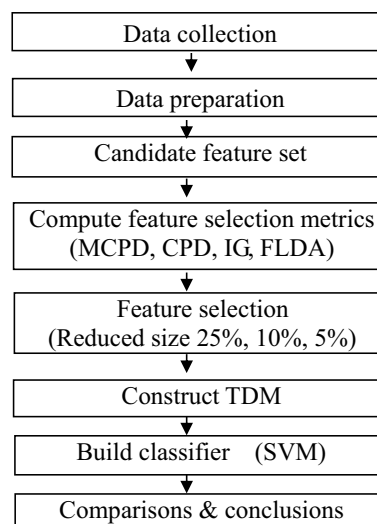


Fig. 1 The procedure of implementing MCPD and comparing with traditional metrics

$$F = \begin{cases} +1, & \text{if Sign} > 0 \\ 0, & \text{if Sign} = 0 \\ -1, & \text{if Sign} < 0 \end{cases} \quad (12)$$

The implemental procedure of our proposed BCF strategy consists of following 5 major steps. And its detailed procedure can be found as Fig. 2.

- Step 1: Construct a candidate feature set.
We use unigram to represent collected documents. After removing some stop words and irrelevant terms, a set of candidate features can be constructed.
- Step 2: Divide candidate features into positive and negative sets.
According to Eqs. (11) and (12), we calculate F value for every feature in candidate set, and then assign those features whose F value is +1 (-1) to positive set P (negative set N).
- Step 3: Feature selection.
Step 3.1: Calculate feature selection metric.
For P and N sets, respectively, we calculate each term’s CPD, MCPD, IG, and FLDA. Then, according to the score of CPD (or MCPD, IG, and FLDA), we rank these features in P and N set, individually.
Next, we define IG and FLDA, respectively. For a term t_k , its IG can be defined as Eq. (13).

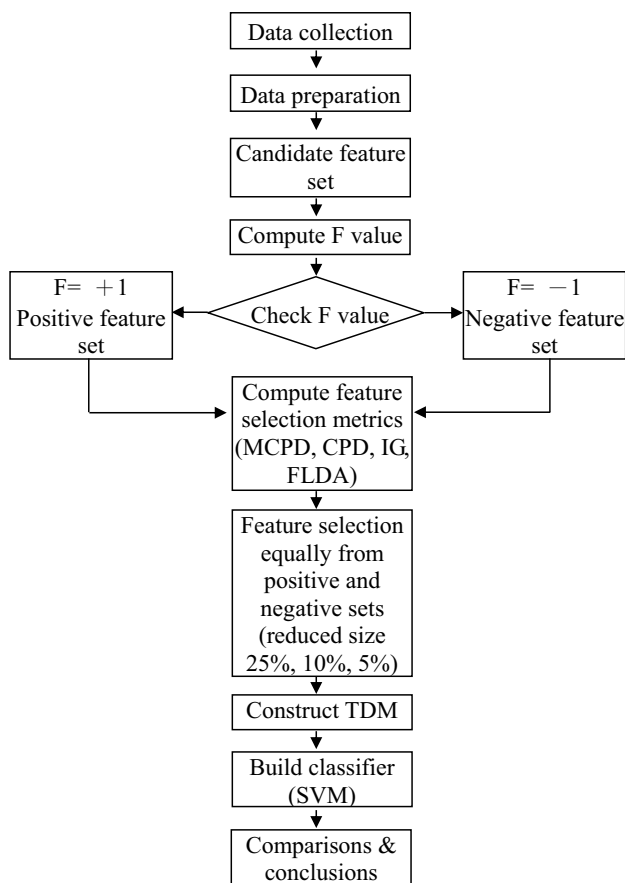


Fig. 2 The procedure of implementing BCF strategy

$$\begin{aligned}
 IG(t_k) &= H(C) - H(C|t_k) \\
 &= - \sum_{i=1}^m p(c_i) \log(p(c_i)) + p(t_k) \sum_{i=1}^m p(c_i|t_k) \log(p(c_i|t_k)) + p(\bar{t}_k) \sum_{i=1}^m p(c_i|\bar{t}_k) \log(p(c_i|\bar{t}_k)) \\
 &= \sum_{i=1}^m \left(p(c_i, t_k) \log \left(\frac{p(c_i, t_k)}{p(c_i)p(t_k)} \right) + p(c_i, \bar{t}_k) \log \left(\frac{p(c_i, \bar{t}_k)}{p(c_i)p(\bar{t}_k)} \right) \right)
 \end{aligned} \tag{13}$$

where $p(c_i)$ is the probability that category c_i occurs, $p(t_k)$ is the probability that term t_k occurs, $p(\bar{t}_k)$ denotes the probability that term t_k does not occur, $p(c_i, t_k)$ means the joint probability of c_i and t_k , and $p(c_i, \bar{t}_k)$ represents the joint probability of c_i and \bar{t}_k . For a certain term t_k , its FLDA can be defined as Eq. (14).

$$FLDA(t_k) = \frac{(E(t_k|P) - E(t_k|N))^2}{D(t_k|P) + D(t_k|N)} \tag{14}$$

where $E(t_k|P)$ and $E(t_k|N)$ denote the conditional mean of term t_k with respect to the categories P and N respectively, $D(t_k|P)$ and $D(t_k|N)$ are the conditional variances of term t_k with respect to the categories P and N respectively.

Step 3.2 :Determine the reduced feature size. Users need to predetermine the feature size they want to reduce. In this project, we will reduce the dimension size from original dimension size to 25%, 10%, 5%, respectively.

Step 3.3: Select features. In this step, based on the pre-determined dimension size, we implement two different feature selection techniques, BCF1 and BCF2. BCF1 is to equally select important attributes from P and N sets based on the computed IG, FLDA, CPD, and MCPD scores. BCF2 is to select candidate positive and negative features according to original proportion of P and N.

Step 3.4 : Construct the feature set for further experiments. This step joins the selected subsets of P and N together to be the employed features of training data.

Step 4: Construct term-document matrix.

Every single comment is converted into a vector of terms (keywords) with term frequency–inverse document frequency (TF-IDF) weights. Then, based on selected features in step3, the collected documents will be transformed to a term-document matrix (TDM).

Step 5: Build SVM model and make conclusion. This step will build support vector machine (SVM) classification model. Then, the constructed model will be validated by test sets built. Moreover, fivefold cross validation (CV) experiment has been employed for these training data.

Based on experimental results, we can make some concluding remarks.

4 Implementation

4.1 The employed data and data preparation

we employs two sentiment data sets including one real cases from real world comments in social media and one famous movie reviews database. Table 3 summarizes the brief background of the employed sentiment data. The first data set is from movie reviews database. They have 1000 positive and 1000 negative comments. After segmenting words and deleting stop words, 4428 words are left for further analysis.

The second data set comes from “ReviewCentre (www.reviewcentre.com)”. By focusing on “MP3 product evaluations (MP3)” related issues, we collect 400 comments. There are 200 positive and 200 negative comments in this data set and the amount of attributes is 1384. In addition, because these evaluations have no sentiment information, we use the 5-star rating system in “ReviewCentre” website to define sentiment labels. A comment will be labeled as positive (negative) if the rate is above 4-star (below 2-star). Those comments whose rate is 3-star have been disregarded.

By the way, some frequently used stop words should be removed. Readers can find a useful stop word listed at https://www.dcs.gla.ac.uk/idom/ir_resources/linguistic_utils/stop_words. And, the package software QDA miner has

been utilized to extract key words and construct TDM in this work. Each comment is converted into a vector of terms (keywords) with TF-IDF. In addition, LIBSVM has been employed to build SVM model (Chang and Lin 2001). RBF kernel function was utilized. All optimal parameter settings of SVM could be obtained by grid search.

4.2 Experimental results

4.2.1 Movie review case

Table 4 summarizes the results of movie reviews. In this experiment, we compare three feature selection approaches in different dimensionality. The comparison base is our original data which only remove the stop words without doing feature selection. The original data contains the 4428 attributes. After fivefold CV experiment, the average classification accuracy is 75.75% and the standard deviation is 4.20%.

Next, we implement experiment # 1 by descending the feature number from 4428 to 50. When the dimensionality space descends from 4428 to 1000, the accuracies of both CPD and MCPD raise greatly (from 75.75 to 93.50%). That is because they can keep important attributes. But, TF and TF-IDF have a significant performance loss. In fact, from Fig. 3, we can easily find that the performances of FF and TI went down during the dimension reduction process. Therefore, we merely compare CPD and MCPD.

There are 415 attributes which have the highest CPD scores (CPD = 1). No selection criteria could be followed if we use smaller size of feature set than 415. Consequently,

Table 3 The employed textual data sets

No.	Data set	Notation	Source	No. of attributes	Data size	Class distribution
1	Movie review	Movie	https://www.cs.cornell.edu/people/pabo/movie-review-data/	4428	2000	Positive: 1000 Negative: 1000
2	MP3 product evaluation	MP3	https://www.reviewcentre.com	1382	400	Positive: 200 Negative: 200

Table 4 Results of movie reviews

Methods	FF+SVM		TI+SVM		CPD+SVM		MCPD+SVM	
	Mean (%)	SD (%)	Mean (%)	SD (%)	Mean (%)	SD (%)	Mean (%)	SD (%)
<i>Dimensions</i>								
1000	75.00	2.65	75.00	5.23	93.50	3.24	93.50	4.28
700	71.00	5.11	71.25	4.76	92.00	3.38	93.00	2.74
400	69.75	4.37	70.50	6.03	88.00	5.63	92.00	4.01
200	69.75	2.85	69.00	4.09	79.00	4.37	88.75	4.68
100	68.50	4.28	67.75	2.85	72.00	3.38	83.00	5.35
50	63.75	3.19	67.50	2.34	68.75	4.42	81.25	3.54

The original size of feature set is 4428 features. And the mean classification accuracy is 75.75% and the mean standard deviation is 4.20%, if we use raw data to build a SVM classifier

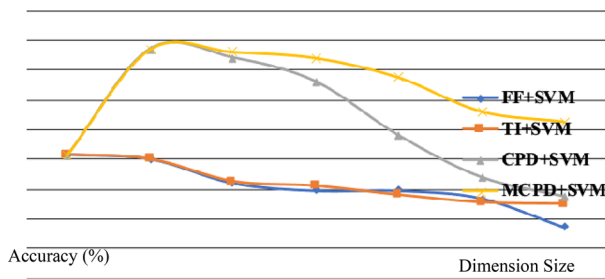


Fig. 3 Results of term selection methods with SVM in six dimensions (movie review case)

only thing we can do is to randomly select attributes from those who have the same CPD score. In dimensions 1000, 700, and 400, both CPD and MCPD could have good classification performances. However, when dimension size keeps going down, we find a performance gap from the result of CPD. The classification performances drop dramatically to 79.00%, 72.00%, and 68.75%, when the dimension size decreases from 400 to 200, 100, and 50, respectively. That’s the drawback of CPD mentioned above. In contrast, when the dimension decreases from 400 to 200, 100, and 50, our method still outperforms others. Their accuracies are 88.75% (dimension = 200), 83.00% (dimension = 100), and 81.25% (dimension = 50).

In order to have statistical evidence, we implement three hypotheses listed in Table 5. The results as shown in Table 5 indicated that the p values of all hypotheses are far less than 0.05. We can reject the null hypotheses (H_0). Consequently, we have 95% confidence to believe the proposed MCPD based SVM is much better than FF, TI, and CPD in movie review case.

The ranges of decreasing for FF and TI methods are stable. However, it cannot get important attributes effectively. Therefore, from results shown in Fig. 3, we can find that MCPD is superior to CPD even when the dimensionality space is low. In addition, both CPD and MCPD are superior to the widely used methods, FF and TI.

Table 6 summarizes the results of the 2nd experiment. We set 8 MCPD thresholds from 1 to 8 select crucial attributes.

Table 5 Hypothesis testing for verifying experiment results (movie review case)

No.	Hypothesis	P value	Decision
H1	$H_0 : \mu_{MCPD+SVM} \leq \mu_{CPD+SVM}$	0.004	Reject H_0
	$H_1 : \mu_{MCPD+SVM} > \mu_{CPD+SVM}$		
H2	$H_0 : \mu_{MCPD+SVM} \leq \mu_{TI+SVM}$	0.000	Reject H_0
	$H_1 : \mu_{MCPD+SVM} > \mu_{TI+SVM}$		
H3	$H_0 : \mu_{MCPD+SVM} \leq \mu_{FF+SVM}$	0.000	Reject H_0
	$H_1 : \mu_{MCPD+SVM} > \mu_{FF+SVM}$		

Table 6 Results of movie reviews

MCPD thresholds	Number of attributes	Accuracy	
		Mean	SD (%)
None (original data)	4428	75.8	4.2
< 1	3950	81.0	4.5
< 2	1816	91.0	2.1
< 3	959	93.3	4.2
< 4	608	92.8	4.1
< 5	314	89.8	4.5
< 6	143	84.5	6.4
< 7	86	84.3	6.8
< 8	48	81.3	3.6

“< 1” means that we remove those attributes whose MCPD values is less than 1

While the values of MCPD are decrease from less 1 to less 3, the number of attributes is decreasing. However, the classification performances are rising greatly. When the features whose MCPD values are from less 4 to less 8 have been deleted, although the classification performances descend eventually with the decrease of attribute amount, the accuracies still could be acceptable. Consequently, our MCPD can keep important attributes and screen un-crucial attributes. If one attribute’s MCPD is small, it represents that the frequency of this attribute in both classes is same. It can’t effectively identify the class labels. On the other hand, when one attribute’s MCPD is large, it represents that the frequency of this factor is high in parts of classes. It could effectively identify the class labels.

4.2.2 MP3 product evaluation case

Table 7 summarizes the results of MP3 product reviews. This employed data has 1,382 attributes. After fivefold CV experiment, the average classification accuracy is 81.50% and the standard deviation is 8.59%. It’s our comparison base.

When dimension space descends from 1382 to 1000, and 700, the performances of CPD are 84.75% and 90.75%, respectively. And MCPD has the performances of 86.5% and 87.5%. Compared with the result of raw data, both CPD and MCPD have better performances than the benchmark (81.5%). About FF and TI, we also find that both of them have descending trends in classification when the dimension size of feature set going down. That’s could be confirmed from Fig. 4.

In MP3 product evaluation case, There are 616 attributes who have the highest CPD scores (CPD = 1). Therefore, we can find CPD can have good performance when dimension size is larger than 616 (1000 and 700). However, when dimension size descending from 700 to 400 (less than

Table 7 Results of the first experiment (mp3 product evaluation case)

Methods	FF+SVM		TI+SVM		CPD+SVM		MCPD+SVM		
	Mean (%)	SD (%)	Mean (%)	SD (%)	Mean (%)	SD (%)	Mean (%)	SD (%)	
<i>Dimensions</i>									
1000	79.75	8.72	79.75	8.02	84.75	6.21	86.50	7.36	
700	79.00	7.15	81.00	9.07	90.75	4.56	87.50	4.59	
400	81.00	6.58	79.75	8.26	81.00	9.24	87.75	4.71	
200	81.75	5.05	82.00	4.47	71.75	4.20	85.25	4.95	
100	78.75	7.07	77.75	5.26	60.75	4.01	86.50	7.15	
50	75.50	5.84	74.75	5.18	61.25	4.76	86.50	8.90	

The original size of feature set is 1382 features. And the mean classification accuracy is 81.50% and the mean standard deviation is 8.59%, if we use raw data to build a SVM classifier

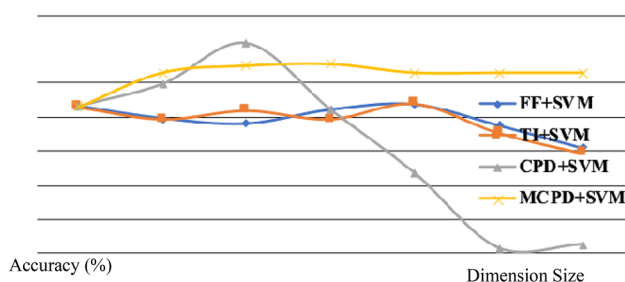


Fig. 4 Results of term selection methods with SVM in six dimensions (mp3 product evaluation case)

Table 8 Hypothesis testing for verifying experiment results (MP3)

No	Hypothesis	P value	Decision
H1	$H_0 : \mu_{MCPD+SVM} \leq \mu_{CPD+SVM}$	0.000	Reject H_0
	$H_1 : \mu_{MCPD+SVM} > \mu_{CPD+SVM}$		
H2	$H_0 : \mu_{MCPD+SVM} \leq \mu_{TI+SVM}$	0.000	Reject H_0
	$H_1 : \mu_{MCPD+SVM} > \mu_{TI+SVM}$		
H3	$H_0 : \mu_{MCPD+SVM} \leq \mu_{FF+SVM}$	0.000	Reject H_0
	$H_1 : \mu_{MCPD+SVM} > \mu_{FF+SVM}$		

616), the classification accuracy decreases remarkably from 90.75% to 81%. It's almost 10% performance loss. When dimension size keeps dropping to 200, 100, and 50, the CPD performances 71.75% (dimension = 200), 60.75% (dimension = 100), and 61.25% (dimension = 50) are even worse than those of FF and TI methods.

Contrarily, our proposed MCPD has stable performances. In dimensions 400, 200, 100, and 50, MCPD could have excellent performance for classifying bloggers' sentiment. They are 87.75% (dimension = 400), 85.25% (dimension = 200), 86.50 (dimension = 100), and 86.50% (dimension = 50). Even when dimension size reduces from 1382 to 50, the performance of MCPD (86.50%) is better than benchmark (81.5%).

Table 9 Results of the second experiment (mp3 review)

MCPD thresholds	Number of attributes	Performance	
		accuracy (%)	SD (%)
None (original data)	1382	81.5	8.6
< 1	1215	84.0	6.6
< 2	982	87.0	5.8
< 3	452	88.0	5.9
< 4	278	87.8	4.3
< 5	186	85.8	2.6
< 6	123	85.3	5.2
< 7	87	87.5	8.2
< 8	67	85.8	8.7

In order to have statistical evidence, we implement three hypotheses listed in Table 8. The results as shown in Table 8 indicated that the p values of all hypotheses are far less than 0.05. Therefore, we can reject all null hypotheses (H_0). So, we have 95% confidence to believe the proposed MCPD based SVM is much better than FF, TI, and CPD in MP3 review case.

Table 9 lists the results of the 2nd experiment in MP3 reviews. Eight thresholds of MCPD have been set from 1 to 8 for selecting crucial attributes. When the features whose MCPD values are from < 1 to < 3 have been deleted, the number of attribute descends, but the classification performances arise remarkably. When removing the features those MCPD values are from < 4 to < 8, the performances of classification decrease slightly. Even in worst situation that remove those whose MCPD scores are < 8, only 67 features are left. But, the accuracy is 85.75% that is also greater than the benchmark (81.50%). Consequently, our MCPD method can extract the useful attributes for classifying sentiment.

4.3 Results of BCF strategy

BCF strategy has been developed for two-sided feature selection methods such as MCPD, IG and FLDA. Table 10

Table 10 Results of implementing BCF strategy to MCPD

Dimension	75%		50%		25%		10%		5%	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>Method</i>										
<i>Movie review</i>										
MCPD	62.80	14.18	59.20	2.17	68.60	1.67	80.80	4.82	81.20	5.76
BCF1-MCPD	54.00	2.74	57.80	1.92	76.00	4.64	85.20	4.76	88.20	3.49
BCF2-MCPD	57.80	2.28	59.20	1.92	68.40	2.88	82.00	4.00	85.00	5.10
<i>mp3 product evaluation case</i>										
MCPD	71.75	12.14	75.50	10.63	81.75	9.79	82.50	7.29	85.50	8.18
BCF1-MCPD	72.00	10.81	77.00	8.18	81.50	10.09	83.50	8.81	85.50	8.51
BCF2-MCPD	71.75	12.67	75.50	10.18	81.25	9.14	83.75	6.19	83.50	8.90

shows the experimental results of MP3 product reviews and movie reviews using BCF strategy combined with MCPD. In the results of MP3 product review, the classification efficiency of the MCPD method combined with the BCF indicator is not significant. But, in the movie review data, when using low dimensions (dimension size reduced to the original 25%, 10%, and 5%), BCF combined MCPD can significantly improve classification performance. And the BCF1-MCPD method has the best classification performance.

From the experimental results, BCF1-MCPD and BCF2-MCPD are generally superior to MCPD. Among them, the

classification performance of BCF1-MCPD is significantly better than BCF2-MCPD and original MCPD.

When the dimension size reduced to 25%, 10%, and 5% of original dimensionality, BCF strategy combined with MCPD has the better performance. Therefore, we also combined the traditional CPD, IG, and FLDA methods with the BCF strategy to conduct experiments in reduced 25%, 10%, and 5% dimension size. Table 11 summarizes experimental results in MP3 product reviews. From the results, the classification efficiency of the CPD method combined with the BCF strategy has only been significantly improved in the dimension of 10%. IG and FLDA methods have significantly improved the

Table 11 BCF strategy combined with CPD, IG, FLDA experimental results (MP3 product reviews)

Method	CPD		BCF1-CPD		BCF2-CPD	
	Mean	SD	Mean	SD	Mean	SD
<i>Dimension (%)</i>						
25	81.50	12.23	80.50	13.33	81.25	12.47
10	81.50	12.23	80.50	13.33	81.25	12.47
5	81.50	12.23	80.50	13.33	81.25	12.47
Method	IG		BCF1-IG		BCF2-IG	
	Mean	SD	Mean	SD	Mean	SD
<i>Dimension (%)</i>						
25	81.75	10.37	83.00	10.41	83.75	9.84
10	83.75	10.90	85.00	9.56	86.25	8.57
5	85.75	7.98	84.50	7.94	85.50	9.08
Method	FLDA		BCF1-FLDA		BCF2-FLDA	
	Mean	SD	Mean	SD	Mean	SD
<i>Dimension (%)</i>						
25	82.50	9.88	83.25	11.03	82.25	9.90
10	85.00	10.19	86.50	9.66	85.50	9.95
5	84.00	7.97	84.50	7.37	85.25	8.02

classification efficiency under the three feature dimensions. In addition, from the experimental results of IG and FLDA, we can see that the classification performance using the BCF1 indicator is generally better than the BCF2 indicator.

Table 12 shows the experimental results of CPD, IG and FLDA methods combined with BCF strategy in movie reviews. The experimental results indicate that the CPD, IG, and FLDA methods combined with the BCF strategy have improved the classification performance in the three dimensions. When BCF2-CPD uses the feature dimension of 25%, all three evaluation indicators show the best classification performance. In addition, the classification performance of the IG and FLDA methods combined with the BCF1 indicator is generally better than the BCF2 indicator, and the best classification performance is achieved when the dimension is 10%.

4.4 Concluding remarks

In addition to comparing the proposed MCPD and traditional CPD, TI, and FF, this section also conducts experiments on the BCF strategy combining MCPD, traditional IG, CPD, and FLDA methods. Based on the results, some concluding remarks could be given as below.

1. The proposed MCPD method significantly improves the shortcomings of poor classification performance when CPD method uses lower feature space for classification.
2. From the evaluation results, we can see that MCPD generally has better classification performance when using fewer features.
3. MCPD combined with BCF strategy can improve the classification performance, of which BCF1-MCPD has better classification results. However, the classification performance improvement in MP3 product reviews is less obvious.
4. The CPD, IG, and FLDA methods combined with the BCF index can improve the classification performance, and the BCF1 index generally has the best classification results.

5 Conclusions

To tackle dimensionality problems when dealing with the huge amount text reviews in social media, we proposed MCPD method and BCF strategy. Results indicated that MCPD outperforms other traditional one-sided term selection methods, CPD, TI and FF. In addition, we also found that BCF strategy combined with MCPD could have the better performance. Consequently, we should use BCF and MCPD together, then we can get the best performance for sentiment classification.

From the experimental results, we can draw some concluding remarks. First, it's confirmed that CPD has drawback in low dimensionality, and MCPD indeed can enhance CPD. In classification problems, both CPD and MCPD outperform

Table 12 BCF strategy combined with CPD, IG, FLDA experimental results (movie reviews)

Method	CPD		BCF1-CPD		BCF2-CPD	
	Mean	SD	Mean	SD	Mean	SD
<i>Dimension (%)</i>						
25	81.40	9.94	81.60	1.52	91.20	6.53
10	75.00	12.81	83.00	13.73	76.00	13.29
5	72.80	5.97	76.40	8.47	87.40	5.03
Method	IG		BCF1-IG		BCF2-IG	
	Mean	SD	Mean	SD	Mean	SD
<i>Dimension (%)</i>						
25	79.20	4.32	82.20	3.96	79.60	2.07
10	81.60	3.13	83.40	1.82	81.60	2.19
5	81.00	4.24	82.20	3.83	81.60	4.62
Method	FLDA		BCF1-FLDA		BCF2-FLDA	
	Mean	SD	Mean	SD	Mean	SD
<i>Dimension (%)</i>						
25	80.80	3.56	82.60	3.91	82.00	3.32
10	79.80	3.42	84.00	2.35	80.80	3.35
5	81.00	3.46	82.80	3.77	82.00	

FF and TI methods which are widely used term selection techniques because they are very easy to be calculated. But, like CPD, FF and TI, MCPD also has the same characteristic of being easy employed. It's very important for sentiment classification, because with the increasing amount of the online reviews, the feature space of textual data increases dramatically. If the feature selection methods cannot reduce the dimensionality with lower computational cost, they might be impractical for applications in real world. Second, the optimal interval of MCPD scores locate at [2, 4]. It means users of MCPD can set a threshold from 2 to 4, and then select important attributes based on this threshold. They can use fewer attributes to obtain better performance in sentiment classification.

This study proposed an easy and simple term selection technique called MCPD to extract crucial features for sentiment classification. Experimental results indicated that our proposed MCPD based SVM learning scheme can improve the drawback of CPD in lower dimensions. In addition, even if we reduce the dimension size from 4428 and 1382 to 50 features, MCPD still has better performances than the performances of using the original dimension size of raw data. Therefore, our method can not only increase the performance of classifying sentiment data, but also dramatically reduce the dimensionality.

Moreover, as mentioned above, when using two-sided feature selection methods, they have the problem of biasing to a certain class. Therefore, this study proposes the BCF strategy. Results indicated that BCF1 + MPCD and BCF1 + FLDA could have the best performance when reduce feature space to extreme low.

With the popularity of the Internet, the amount of text comments in social media is going to increase remarkably. Consequently, our method is very suitable not limited to apply to real-world data of sentiment classification, but also text classification problems. In addition, we use TF-IDF to be our term weights in TDM. Using different term weighting methods could be one of the potential directions of future works.

Acknowledgements This work was supported in part by Ministry of Science and Technology, Taiwan (Grant no. MOST 108-2410-H-324-009).

References

- Abbasi A, Chen H, Salem A (2008a) Sentiment analysis in multiple languages: feature selection for opinion classification in web forums. *ACM Trans Inf Syst* 26(3):article 12
- Abbasi A, Chen H, Thoms S, Fu T (2008b) Affect analysis of web forums and blogs using correlation ensembles. *IEEE Trans Knowl Data Eng* 20(9):1168–1180
- Aizawa A (2003) An information-theoretic perspective of tf-idf measures. *Inf Process Manag* 39(2003):45–65
- Akhtar S, Gupta D, Ekbal A, Bhattacharyya P (2017) Feature selection and ensemble construction: a two-step method for aspect based sentiment analysis. *Knowl-Based Syst* 1251:116–135
- Akhtar S, Garg T, Ekbal A (2020) Multi-task learning for aspect term extraction and aspect sentiment classification. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2020.02.093>
- Alqaryouti O, Siyam N, Monem AA, Shaalan K (2019) Aspect-based sentiment analysis using smart government review data. *Appl Comput Inform*. <https://doi.org/10.1016/j.aci.2019.11.003>
- Bai X (2011) Predicting consumer sentiments from online text. *Decis Support Syst* 15(4):732–742
- Cerqueira AS, Ferreira DD, Ribeiro MV, Duque CA (2008) Power quality events recognition using a SVM-based method. *Electr Power Syst Res* 78(9):1546–1552
- Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines, Software. <https://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chang JR, Chen MY, Chen LS, Chien WT (2020) Recognizing important factors of influencing trust in O2O models: an example of OpenTable. *Soft Comput* 24:7907–7923. <https://doi.org/10.1007/s00500-019-04019-x>
- Chaovalit P, Zhou L (2005) Movie review mining: a comparison between supervised and unsupervised classification approaches. In: *Proceedings of the 38th Hawaii international conference on system sciences*
- Chen LS, Su CT (2008) Using granular computing model to induce scheduling knowledge in dynamic manufacturing environments. *Int J Comput Integr Manuf* 21(5):569–583
- Chen LS, Hsu CC, Chen MC (2009) Customer segmentation and classification from blogs by using data mining: an example of VOIP phone. *Cybernet Syst* 40(7):608–632
- Chen LS, Liu CH, Chiu HJ (2011) A neural network based approach for sentiment classification in the blogosphere. *J Inform* 5(2):313–322
- Chouchani N, Abed M (2020) Enhance sentiment analysis on social networks with social influence analytics. *J Ambient Intell Human Comput* 11:139–149
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Dave K, Lawrence S, Pennock DM (2003) Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: *The 12th WWW*, pp 519–528
- Denecke K, Nejd W (2009) How valuable is medical social media data? Content analysis of the medical web. *Inf Sci* 179:1870–1880
- Giuntini FT, Cazzolato MT, dos Reis MdJD (2020) A review on recognizing depression in social networks: challenges and opportunities. *J Ambient Intell Human Comput*. <https://doi.org/10.1007/s12652-020-01726-4>
- Gokalp O, Tasci E, Ugur A (2020) A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. *Expert Syst Appl* 14615:Article 113176
- Hassonah MA, Al-Sayyed R, Rodan A, Al-Zoubi AM, Faris H (2020) An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter. *Knowl Based Syst* 19215:Article 105353
- Hsu CW, Chang CC, Lin C-J (2006) A practical guide to support vector classification. <https://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html>.
- Keshtkar F, Inkpen D (2009) Using sentiment orientation features for mood classification in blogs. In: *IEEE international conference on natural language processing and knowledge engineering*
- Khan FH, Qamar U, Bashir S (2016) SWIMS: semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis. *Knowl Based Syst* 10015:97–111

- Kim K (2018) An improved semi-supervised dimensionality reduction using feature weighting: application to sentiment analysis. *Expert Syst Appl* 1091:49–65
- Kong L, Li C, Ge J, Zhang F, Feng Y, Li Z, Luo B (2020) Leveraging multiple features for document sentiment classification. *Inf Sci* 518:39–55
- Kumar V, Venkatesan R, Reinartz W (2006) Knowing what to sell, when, and to whom. *Harvard Bus Rev* 20:131–137
- Li B, Xu S, Zhang J (2007) Enhancing clustering blog documents by utilizing author/reader comments. In: *Proceedings of the 45th annual southeast regional conference*
- Lightspeed Research (2011) Consumer reviews and research online. <https://www.lightspeedresearch.com/press-releases/consumers-rely-on-online-reviews-and-price-comparisons-to-make-purchase-decisions/>
- Liu B (2020) Text sentiment analysis based on CBOW model and deep learning in big data environment. *J Ambient Intell Human Comput* 11:451–458
- Liu B, Hu M, Cheng J (2005) Opinion observer: analyzing and comparing opinions on the web. In: *Proceedings of the 14th international conference on World Wide Web*
- Liu Y, Bi JW, Fan ZP (2017) Multi-class sentiment classification: the experimental comparisons of feature selection and machine learning algorithms. *Expert Syst Appl* 801:323–339
- Mekawie N, Hany A (2019) Understanding the factors driving consumers' purchase intention of over the counter medications using social media advertising in Egypt: (A Facebook advertising application for cold and Flu products). *Proced Comput Sci* 164:698–705
- Na JC, Khoo C, Wu PHJ (2005) Use of negation phrases in automatic sentiment classification of product reviews. *Lib Collect Acquis Tech Serv* 29(2):180–191
- O'Keefe T, Koprinska I (2009) Feature selection and weighting methods in sentiment analysis. In: *Proceedings of the 14th Australasian document computing symposium*
- Ogura H, Amano H, Kondo M (2011) Comparison of metrics for feature selection in imbalanced text classification. *Expert Syst Appl* 38(5):4978–4989
- Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. In: *Proceedings of the ACL-02 conference on empirical methods in natural language processing*, vol 10, pp 79–86
- Simeon M, Hilderman R (2008) Categorical proportional difference: a feature selection method for text categorization. In: *Proceedings of the 17th Australasian data mining conference*, pp 201–208
- Singh NK, Tomar DS, Sangaiah AK (2020) Sentiment analysis: a review and comparative analysis over social media. *J Ambient Intell Human Comput* 11:97–117
- Song C, Wang XK, Cheng PF, Wang JQ, Li L (2020) SACPC: a framework based on probabilistic linguistic terms for short text sentiment analysis. *Knowl-Based Syst*. <https://doi.org/10.1016/j.knsys.2020.105572>
- Tan S, Zhang J (2008) An empirical study of sentiment analysis for Chinese documents. *Expert Syst Appl* 34(4):2622–2629
- Tang H, Tan S, Cheng X (2009) A survey on sentiment detection of reviews. *Expert Syst Appl* 36(7):10760–10773
- Vapnik VN (1995) *The nature of statistical learning theory*. Springer, Berlin
- Wang S, Li D, Song X, Wei Y, Li H (2011) A feature selection method based on improved fisher's discriminant ratio for text sentiment classification. *Expert Syst Appl* 38(7):8696–8702
- Whitelaw C, Garg N, Argamon S (2005) Using appraisal groups for sentiment analysis. In: *Proceedings of the 14th ACM conference on information and knowledge management*, pp 625–631
- Wu CH, Chuang ZJ, Lin YC (2006) Emotion recognition from text using semantic labels and separable mixture models. *ACM Trans Asian Lang Inf Process* 5(2):165–182
- Xu F, Pan Z, Xia R (2020) E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Inf Process Manag*. <https://doi.org/10.1016/j.ipm.2020.102221>
- Ye Q, Zhang Z, Law R (2009) Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Syst Appl* 36:6527–6535
- Yousefpour A, Ibrahim R, Hamed HNA (2017) Ordinal-based and frequency-based integration of feature selection methods for sentiment analysis. *Expert Syst Appl* 751:80–93
- Zhang W, Yu C, Meng W (2007) Opinion retrieval from blogs. In: *Proceedings of the sixteenth ACM conference on conference on information and knowledge management*, pp 831–840
- Zhao P, Hou L, Wu O (2020) Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. *Knowl Based Syst* 1936:Article 105443
- Zheng Z, Wu X, Srihari R (2004) Feature selection for text categorization on imbalanced data. *ACM SIGKDD Explor Newsl* 6(1):80–89
- Zhu L, Sun A, Choi B (2011) Detecting spam blogs from blog search results. *Inf Process Manag* 47(2):246–262

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.