IEEE Access
Multidisciplinary Rapid Review Open Access Journal

Received November 13, 2017, accepted December 7, 2017, date of publication December 12, 2017, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2782668

TCMHG: Topic-Based Cross-Modal Hypergraph Learning for Online Service Recommendations

ZHIKUI CHEN[1,2], (Member, IEEE), FEI LU[1], XU YUAN[1,2], AND FANGMING ZHONG[1]
[1]School of Software Technology , Dalian University of Technology, Dalian 116023, China
[2]Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian 116620, China

Corresponding author: Zhikui Chen (zkchen@dlut.edu.cn)

This work was supported in part by the Nature Science Foundation of China under Grant 61672123, in part by the State Key Program of National Natural Science of China under Grant U1301253, and in part by the Science and Technology Planning Key Project of Guangdong Province under Grant 2015B010110006.

ABSTRACT Online product reviews sentiment classification plays an important role on service recommendation, yet most of current researches on it only focus on single-modal information ignoring the complementary information, that results in unsatisfied accuracy of sentiment classification. This paper proposes a cross-modal hypergraph model to capture textual information and sentimental information simultaneously for sentiment classification of reviews. Furthermore, a mixture model by coupling the latent Dirichlet allocation topic model with the proposed cross-modal hypergraph is designed to mitigate the ambiguity of some specific words, which may express opposite polarity in different contexts. Experiments are carried out on four-domain data sets (books, DVD, electronics, and kitchen) to evaluate the proposed approaches by comparison with lexicon-based method, Naïve Bayes, maximum entropy, and support vector machine. Results demonstrate that our schemes outperform the baseline methods in sentiment classification accuracy.

INDEX TERMS Cross-modal, hypergraph learning, topic model, sentiment classification, product reviews.

I. INTRODUCTION

Recent years have witnessed dramatic increase of shopping websites like Amazon and eBay, due to the rapidly increasing of people shopping online. However, online transactions often create perplexity and ambiguity with regard to consumers' choices, owing to the intangible of quality and the heterogeneity of services [1]. Thus, recommendation information is particularly significant for online services. An excellent recommendation system can reduce the search effort for users. It also brings higher sales, more advertising revenues as well as greater consumer loyalty [1], [5], [13], [40]. Among these multiform of recommendations, customer reviews are the most influential factor in changing behavior of consumers. Therefore, online reviews play an important role for both businesses and purchasers. On the one hand, the sellers expect to follow the tracks of the effect of their products or services, and how is the consumers' feedback on the shopping websites. The gathered information may simulate the businesses' inspiration to promote their quality of commodities or improve their service quality. On the other hand, customers long for reading valuable comments to help them compare products, and make decisions. However, it is usually impossible to read all of them as the volume of product comments highly increased [7], [41]. Therefore, how to effectively extract the sentiment hidden in the reviews is the remaining challenge. Motivated by this, we focus on sentiment classification of product reviews in this paper.

Various methods have been proposed previously, which can be classified into two categories in general, i.e. machine learning based methods [2], [3], [8], [10] and lexicon based methods [4], [6], [7]. Pang *et al.* [2] firstly employed machine learning methods namely Naive Bayes (NB), Maximum Entropy (ME), and Support Vector Machines (SVM) to sentiment polarity classification. Lu *et al.* [6] estimate the sentiment polarity strength of product reviews by multiplying the strength of adjectives and adverbs that are used in the phrases. The former instance belongs to machine learning approaches, which often yield high rate of accuracy on sentiment classification problem while with limited adaptation. On the contrary, the latter example, usually providing better generalization capability but non-ideal classification accuracy, is regarded as lexicon based methods. However, most of existing approaches only focus on single modal feature ignoring other complementary information, which results in unsatisfied sentiment analysis performance. Motivated by this, we intend to take the advantages of both lexicon-based methods and machine learning methods, as well as the multimodal information in reviews. In this paper, we propose a cross-modal hypergraph model to combine textual feature (term frequency-inverse document frequency, TF-IDF) and sentimental feature (sentiment scores) of reviews simultaneously.

Nevertheless, the limited and ambiguous local information are not discriminant enough to sentiment prediction. Therefore, a mixture model based on cross-modal hypergraph and topic model is presented that can take the global information into consideration. There are a number of specific words that are ambiguous according to the domain they appeared. For instance, the word "unpredictable" in the phrases "unpredictable screen" and "unpredictable plot". The first one may indicate negative orientation in an electronic product review, at the same time it may also express positive orientation in the second phrase in a book review. Obviously, sentiment polarities rely on the aspects or domains that phrases emerged. Hence, exploring both sentiment and topic information simultaneously should be conducive to the task of opinion mining on product reviews [9]. There are some researches attempted to recognize the sentiment of a certain aspect in one sentence other than the whole paragraph or document. This kind of simple method is to obtain an opinion score of one definite aspect by the weighted sum of sentiment scores of all sentiment words appeared in the sentence, where the weight is calculated by the inverse of the distance between aspect and sentiment word [19]. This method has been improved by recognizing the aspect-opinion relations employing tree kernel approach [20]. Recently, topic models have become a powerful tool to learn the document collections [16] after Blei *et al.* [15] proposed Latent Dirichlet Allocation model (LDA). In this work, we design a topic mixture model by employing LDA topic model to realize soft clustering, to reduce the ambiguity of some specific words.

The main contributions in this paper are summarized as follows:

- We propose a hypergraph model which can integrate advantages between cross modal information (TF-IDF and sentiment score of product reviews) for improving sentiment classification of product reviews. Moreover, it can be extended to fuse multimodal features at will.
- The proposed hypergraph model can show the high-order relations among samples which will contribute to the classification accuracy.
- A mixture model which introduces LDA into the cross-modal hypergraph algorithm has been designed. It not only reduces the impact of ambiguity produced by some specific words, but also lower the running time markedly.

The rest of the paper is organized as follows. Section 2 introduces some previous methods for sentiment classification and a brief description of hypergraph as well as topic model. Section 3 describes our cross-modal hypergraph sentiment classifier. Section 4 explains the topic-based mixture model TCMHG. Section 5 evaluates the results of the experiments. Finally, we conclude our research in Section 6.

II. RELATED WORK

The TCMHG model has been devised for online service recommendation in this work, thus we review prior work in related areas including sentiment classification, hypergraph learning and topic model in this section.

A. SENTIMENT CLASSIFICATION

Sentiment analysis has been regarded as an important role in many fields such as product and restaurant comments [17], [19], [42]. As one of the most significant tasks in sentiment analysis, sentiment classification has been studied extensively. The work presented in [3] employs three prevalent ensemble methods namely bagging, boosting and random subspace on ten different public datasets when using five base learners viz. NB, ME, Decision Tree (DT), K Nearest Neighbors (KNN), and SVM. Turney put forward to use Pointwise Mutual Information (PMI) and Information Retrieval (IR) to calculate the similarity of pairs of words, where "excellent" and "poor" have been regarded as the positive and negative reference words, then sentiment orientation of reviews can be obtained by computing the difference of PMI using "excellent" and "poor" respectively [21]. However, most of the previous work [2]–[4], [21] only consider the text representations and do not take advantage of emotional information [18].

The researchers in [22] concentrate on predicting the sentiment polarity of opinion sentences by utilizing adjectives that are associated with their corresponding sentiment orientation values. Taboada et al. proposed a Semantic Orientation CALculator (SO-CAL) method [4], which uses sentiment lexicon established by linguistic specialist and incorporates linguistic rules like intensification and negation to extract sentiment from texts. Besides, it is worth mentioning that the dictionary not only includes adjectives but also verbs, nouns and adverbs. Nevertheless, the lexicon-based approaches show low level of reliability due to the dictionaries are either built automatically or hand-ranked by humans [32]. In addition, such approaches [4], [6], [22] are considered as restricted by a satisfied sentiment lexicon to a certain degree, where the dictionary is difficult to obtain.

Moreover, most of them mainly focus on unimodal feature extraction, and other complementary modality features are ignored. Hence, how to combine the additional modality information is crucial. To achieve this goal, we construct a cross-modal hypergraph which can take the multimodal information in the reviews into consideration.

B. HYPERGRAPH LEARNING

A hypergraph can be seen as an extension of a simple graph [24], where a hyperedge can connect more than two vertices [23]. There are many ways to construct hyperedges, for example, [23], [25] concatenate vertices with a definite feature, and some others form hyperedges through the centroid vertex and its k-nearest neighbors [26], [27]. Furthermore, the hypergraph model can make full use of unlabeled dataset to showing the high-order information [28]. Thanks to this advantage, hypergraph has been extensively used in various applications, such as partitioning [29], ranking [25], [26] and classification [27], [28]. More specifically, Chen *et al.* [28] construct hypergraph by integrating three modalities (textual, visual, and emoticon) for sentiment classification of microblog. Unfortunately, when it encounters the ambiguity of some specific words, it would fail to obtain the accurate results. Hence, in this paper, we design a mixture model by combining the topic model to address this mentioned problem.

C. TOPIC MODEL

Topic models such as LDA have been widely used in aspect-based opinion mining [19]. The authors in [12] stress that it is significant to possess an unsupervised method for detecting aspect, thus they propose a model which is an extension to the basic prototype of topic model LDA [15] and Probabilistic Latent Semantic Analysis (PLSA) [30]. In order to address the problem of skewed document distribution, the work in [16] allows users to supply several seed words which can represent the corpus appropriately.

Besides, many existing researches have applied topic models to jointly detect both topic and sentiment [7], [9]–[11], [14], [18], [33], [38]. More particularly, MaxEnt-LDA hybrid model [11] was proposed to detect both aspects and aspect-specific opinion words simultaneously, and it can separate sentiment words and aspects through syntactic features. Lin *et al.* [14] present a probabilistic model framework named joint sentiment-topic (JST) model, which is weakly supervised while a majority of existing methods of sentiment classification prefer supervised learning. The work in [7] describes an approach named sentiment-aligned topic model (SATM), which concentrates on the sentiment label alignment problem and aims at predicting the aspect rating of product reviews.

Most recently, the researchers in [8] provide a novel application of the LDA model, where they split the data into multi-fold sub-collections according to the topic distributions. And then sentiment classification models are trained in each sub-collection respectively. Actually, our topic-based mixture model is more closely to the soft clustering by LDA which proposed in [8].

III. CROSS-MODAL HYPERGRAPH SENTIMENT CLASSIFIER

A cross-modal hypergraph is a hypergraph which includes vertices or hyperedges constructed from heterogeneous data source. In this work, we construct cross-modal hypergraph model for sentiment classification, thus each review can be regarded as a vertex, and different kinds of relations among reviews can be treated as different types of hyperedges. After constructing a hyperedge in each modality, the sentiment prediction is then transformed to a ranking problem of relevance score which is computed based on the similarities of reviews.

In the following of this section, we describe the proposed cross-modal hypergraph algorithm for sentiment classification. Figure 1 illustrates the procedure of the proposed classifier, yet we neglect the module presented in the red dashed box for the moment, and that will be introduced in particular in the next section.

A. FEATURE EXTRACTION

Assume that there are n product reviews, i.e. $P = \{p_1, p_2, \ldots, p_n\}$, and m different words $W = \{w_1, w_2, \ldots, w_m\}$ in these reviews. We first extract the TF-IDF feature and sentiment scores of each review, then we use p^t and p^s to represent textual and sentimental feature respectively, where these two kind of features are treated as two different modalities in this paper. The TF-IDF value of given p_i can be expressed as $p_i^t = \left\{r_i^1, r_i^2, \ldots, r_i^m | i = 1, \ldots, n\right\}$, where $r_i^j \in R$ represents the TF-IDF value of the j-th word in p_i. And we set $p_i^s = s_i$ to denote the sentiment score of p_i, where the sentiment score can be obtained by employing the method and sentiment dictionaries presented in [4]. For example, "The book arrived as expected and was in great (+4) shape. Thanks (+2)" (the numbers in brackets are the sentiment score of the corresponding words), so the total sentiment score of this sentence is $4 + 2 = 6$.

B. CROSS-MODAL HYPERGRAPH CONSTRUCTION

We use V to denote a finite set of vertices, and E to represent the set of hyperedges, in which a hyperedge e consists of a subset of V. Thus the union of all hyperedges meets the condition that $\cup_{e\in E} = V$. In a hypergraph, each hyperedge e possesses its own weight $w(e)$. Then a weighted hypergraph with the hyperedge set E and the vertex set V can be represented as $G = (V, E, w)$. A hypergraph G can be denoted by using an $|V| \times |E|$ incidence matrix H whose entry $h(v, e) = 1$ if $v \in e$ and 0 otherwise. The degree of a hyperedge $e \in E$ is defined as $\delta(e) = \sum_{v\in V} h(v, e)$, and for a given vertex $v \in V$, its degree is represented by $d(v) = \sum_{e\in E} w(e)\, h(v, e)$. More concretely, $\delta(e)$ is obtained by calculating the sum of a column in incidence matrix H, and $d(v)$ by computing the weighted sum of a row in H. Let D_e and D_v represent the diagonal matrices in which the diagonal entries are the hyperedge and vertex degrees respectively. Similarly, we use W to denote the diagonal matrix whose entries are the weights of hyperedges.

We can calculate the Euclidean distance between each two reviews for textual modal and sentimental modal respectively. Therefore, the cross-modal hypergraph can be constructed

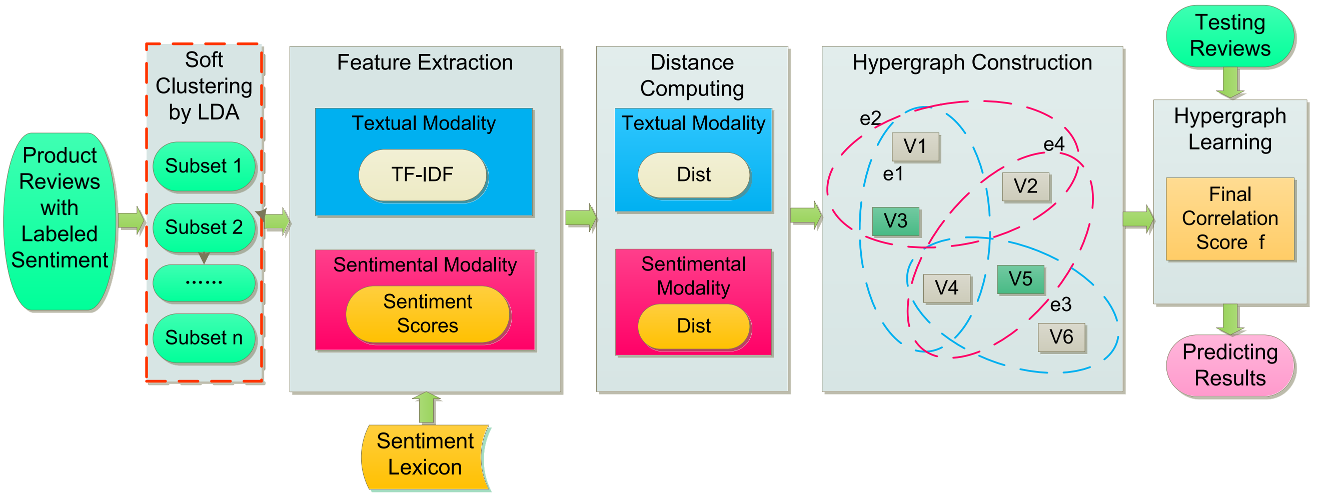


FIGURE 1. **Schematic diagram illustration of topic-based cross-modal hypergraph classifier.**

by each review (treated as centroid vertex) and its k-nearest neighbors on each modality. An example is shown in Figure 1, we take v_3 and v_5 as the centroid vertex, then e_1 and e_3 can be formed by these two vertices and their 2-nearest neighbors in textual modal. Similarly, e_2 and e_4 are formed in sentimental modal. Thus, given a corpus contained N reviews, we can construct a $|N| \times |2N|$ hypergraph. In addition, we derive the similarity for review i and review j from the distance in each modal as follows:

$$Sim_x(i,j) = \begin{cases} exp\left(-\frac{D_x(i,j)}{\bar{D}_x}\right), & if\ i \neq j \\ 0 & else, \end{cases} \tag{1}$$

where D_x denotes the distance matrix calculated on the x^{th} modal, and $\bar{D}_x$ represents the median value of D_x. Then, we can obtain the hyperedge weight on the x^{th} modal by:

$$w_x(e_i) = \sum\nolimits_{v_j \in e_i} Sim_x(i,j). \tag{2}$$

Intuitively, the higher similarity of the reviews within a hyperedge is, the larger weight of a hyperedge will has.

C. HYPERGRAPH LEARNING STAGE

A definite testing review is represented by a $|V| \times |1|$ testing vector q, in which the item corresponding to the testing vertex is assigned to 1, otherwise 0. Likewise, the final correlation scores are also denoted by a $|V| \times |1|$ vector $\mathbf{f}$. Thus, in order to obtain the correlation scores, we employ the similar method proposed in [23] which is aimed to minimize the following cost function:

$$\arg\min_f \{\Omega(f) + \mu\Phi(f)\}, \tag{3}$$

where $\mu > 0$ is the regularization parameter which can adjust the tradeoff between these two terms. Next, we will explain these two functions respectively in detail. The first term can be written as follows:

$$\Omega(f) = \frac{1}{2}\sum\nolimits_{e\in E}\sum\nolimits_{u,v\epsilon V}\frac{w(e)\,h(u,e)\,h(v,e)}{\delta(e)} \times\left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}}\right)^2, \tag{4}$$

where the right hand part is a constraint item which can compel the vertices sharing many hyperedges mutually to have similar correlation scores. That is to say, two reviews may get a similar correlation score if they are similar to many mutual reviews. Moreover, the function is generally known as a regularizer based on normalized hypergraph Laplacian [28], and it can be expanded as:

$$\begin{aligned} &\frac{1}{2}\sum\nolimits_{e\in E}\sum\nolimits_{u,v\epsilon V}\frac{w(e)\,h(u,e)\,h(v,e)}{\delta(e)} \\ &\quad\times\left(\frac{f^2(u)}{d(u)} - 2\frac{f(u)f(v)}{\sqrt{d(u)\,d(v)}} + \frac{f^2(v)}{d(v)}\right) \\ &= f^T f - f^T D_v^{-\frac{1}{2}} HWD_e^{-1}H^T D_v^{-\frac{1}{2}} f. \end{aligned} \tag{5}$$

The transform procedure in Equation (5) is close to the process of proof proposed in [31]. As we define a matrix $\Theta = D_v^{-\frac{1}{2}}HWD_e^{-1}H^TD_v^{-\frac{1}{2}}$, the Equation (4) can be trimmed as $\Omega(f) = f^T(I-\Theta)f$, where I indicates the identity matrix. Further, we let $\Delta = I - \Theta$ be the hypergraph Laplacian, so the concise form of $\Omega(f)$ can be stated as follows:

$$\Omega(f) = f^T\Delta f. \tag{6}$$

Then the empirical loss $\Phi(f)$ in the second term of Equation (3) is defined by

$$\begin{aligned} \Phi(f) &= \sum\nolimits_{u\epsilon V}(f(u) - q(u))^2 \\ &= (f-q)^T(f-q). \end{aligned} \tag{7}$$

The role of this function is to control the final correlation score should be as close as possible to the initial value of the testing vector. Thus, we use $\Psi(f)$ to denote the whole cost function, and it can be transformed into $\Psi(f) = f^T \Delta f + \mu (f - q)^T (f - q)$. Omitting a series of deriving process after differentiating $\Psi(f)$ with respect to f, the final correlation score can be obtained by:

$$f = \left(\frac{\mu}{1+\mu}\right)\left(I - \frac{1}{1+\mu}\Theta\right)^{-1} q. \tag{8}$$

Intuitively, the $\frac{\mu}{1+\mu}$ can be seen as a constant coefficient without prejudice to the final scores. Therefore, the ultimate form of the final correlation score can be simply expressed as

$$f = \left(I - \frac{1}{1+\mu}\Theta\right)^{-1} q. \tag{9}$$

We summarize the whole procedure of our cross-modal hypergraph model for sentiment classification of product reviews in Algorithm 1.

Algorithm 1 Sentiment Classification on Cross-modal Hypergraph

Input: The product reviews $P = \{p_1, p_2, \ldots, p_n\}$

Output: The final correlation score f for product reviews sentiment classification.

Steps:

1: Extract features of each reviews p_i in each modal (p^t and p^s).

2: Calculate the review distance matrix D_x in each modal respectively.

3: Calculate the similarity matrix Sim_x through D_x.

4: for Each modal do:

5: for Each review p_i do:

6: Construct a hyperedge by connecting its k-nearest neighbors based on Sim_x.

7: end for

8: Generate the incidence matrix H_x.

9: Calculate the weight matrix W_x.

10: end for

11: Generate H and W by concatenating H_x and W_x on the basis of column respectively.

12: Calculate matrix Θ.

13: Calculate the final correlation score f by a given testing review q_i.

IV. TOPIC-BASED HYPERGRAPH MIXTURE MODEL

Topic model has been proven to be effective for relieving the ambiguity of some specific words [7], [9]–[11], [14], [16], [38]. Therefore, we try to employ topic model to further improve the classification accuracy. In this section, we introduce how to encode the topic information into our proposed cross-modal hypergraph.

A. TOPIC MODELING

As mentioned above, Latent Dirichlet Allocation model (LDA) has been broadly applied in Natural Language Processing (NLP) as an effective topic model. The basic idea is that each document is represented as a mixture of latent topics [15]. The model supposes that every word is generated by topics and that all these topics are interchangeable within a document. In other words, there is a multinomial distribution named Dirichlet prior $Dir(\alpha)$ over topics for each document, as well as there is another multinomial distribution over words for each topic. Gibbs sampling [39] is the most popular form of Markov Chain Monte Carlo, and is a widely used algorithm for inference and parameter estimation under LDA. Thus, we employ this effective algorithm in our work.

Each product review can be treated as one document. We conduct text preprocessing through removing a list of stop words and a number of low-frequency words appeared in the product review datasets. For a given review p_i, the posterior probability of each topic t_j can be calculated by:

$$P_t\left(t_j|p_i\right) = \frac{N_{ij} + \alpha_j}{\sum_{k=1}^{Z} N_{ik} + Z\alpha_j}, \tag{10}$$

where Z denotes the number of topics for all the product reviews in a corresponding dataset, and N_{ij} represents the total number of times that topic t_j has been assigned to several words in review p_i. Furthermore, N_{ij} is computed by averaging the multiple iterations of Gibbs Sampling in general. In addition, α_j is the j-th dimension of the hyper-parameter of the Dirichlet distribution which can be optimized while inference and parameter estimation.

B. TOPIC-BASED CROSS-MODAL HYPERGRAPH

As mentioned previously, we conduct LDA model to cluster the datasets in terms of the topic probability. Thus, there is no doubt that the whole dataset will be divided into several subsets. Subsequently, we can construct our cross-modal hypergraph classifier in each subset. For clustering, we can employ several methods according to the similarity of the topic probability distribution, such as K-means, hierarchical clustering and so on. However, these approaches are limited by how to choose a reasonable number of clusters. Therefore, we consider employing threshold value ε to split the total product reviews into corresponding subset. More specifically, the formal description is given as follows: Given a product review p_i, we partition p_i into cluster j if and only if $P_t\left(t_j|p_i\right) > \varepsilon$ or $P_t\left(t_j|p_i\right) = \max_{k\in Z} P_t\left(t_k|p_i\right)$. Therefore, we can note that this is a kind of soft clustering, because in some reviews, the $P_t\left(t_j|p_i\right)$ may exceed the threshold value ε in more than one topic within the topic distribution, and that will lead the same product review to be assigned into multi-clusters.

To be brief, the topic-based hypergraph mixture model is to insert a module before feature extraction as shown in Figure 1, where the module conducts LDA to infer the topic distribution and splits the whole dataset into multiple small subsets through topic probability. Then hypergraph sentiment classifiers are constructed in each subset according to Figure 1.

TABLE 1. Classification accuracy of different approaches.

Datasets	Lexicon-Based	NB	ME	SVM(Textual Modal)	SVM(Cross-Modal)	Sentimental Modal Only	Cross-modal
Books	0.716	0.7417	0.7167	0.7307	0.7836	0.7771	**0.7975**
DVD	0.727	0.7633	0.755	0.6829	0.7493	0.7747	**0.7945**
Electronics	0.732	0.7517	0.7433	0.7329	0.7921	0.8021	**0.8065**
Kitchen	0.7385	0.7533	0.7317	0.7057	0.7814	0.8094	**0.8165**

C. COMPLEXITY ANALYSIS

Time complexity analysis of our TCMHG model is given in this subsection. The final correlation score computed in Equation (9) involves an inverse matrix which plays a decisive role in the running time. Suppose an $n \times n$ invertible matrix M, the time required to obtain M^{-1} (the inverse of matrix M) is generally $O(n^3)$. As for LDA model, its computational cost depends on the number of words appeared in corpus, the number of topics and iteration times. Therefore, the time complexity of LDA is $O(Z * |C| * T)$, where Z denotes the number of topics, $|C|$ represents the where Z denotes the number of topics, $|C|$ represents the number of words and T means the number of iterations. Usually, the number of words is approximately 100K in each dataset, thus we can conclude that $Z * |C| * T$ is less than n^3. Hence, the time complexity is not change after appending the LDA topic model. And not only that, the fewer vertices make the process of constructing hypergraph much easier, resulting the running time significantly reduced.

V. EXPERIMENTS

In this section, we conduct extensive experiments to verify two kinds of situations. The first one is to confirm that whether cross-modal hypergraph can contribute to the accuracy (which can be obtained via the number of samples classified correctly divided by the total number of samples) improvement of sentiment analysis. The other one is to validate whether using soft clustering by LDA model can be conducive to the classification result. In the following of this section, we will introduce the preprocessing, experiment settings and the evaluation of experimental results in detail.

A. DATASETS AND PREPROCESSING

We evaluate the effectiveness of our proposed model by performing the experiments on multi-domain sentiment datasets [34], which contain four different domain (book, DVD, electronics, kitchen) product reviews collected from Amazon.com and have been broadly used in the field of sentiment analysis. Each domain is composed of 2000 product reviews with half positive and half negative, we randomly select 700 instances as labeled data and remaining 300 as test data in each class.

For each product review p_i, we first remove the stop words and some frequent words. Then we use Stanford Log-linear Part-Of-Speech Tagger[1] to determine part of speech which aims at eliminating the ambiguity of some words. For example, the word ''plot'' is negative only when it is a verb, and similarly, ''novel'' is positive only when it is an adjective.

The sentiment dictionaries we used are introduced in [4], including 2827 adjectives, 876 adverbs, 1549 nouns and 1142 verbs. Each word has been labeled from +5 for extremely positive to -5 for extremely negative according to their sentiment orientation. In addition, there are 216 adverbs that can strengthen or weaken the sentiment polarity to a certain degree.

B. CROSS-MODAL HYPERGRAPH MODEL

An important factor that affects the results of our hypergraph learning model is the size of hyperedge, which depends on the number of k-nearest neighbors. In this group of experiments, we set k equals to 45, based on our empirical observations. Besides, we fix the value of $\frac{1}{1+\mu}$ listed in Equation (9) to 0.1, that is we choose μ to be 9.

We compare our cross-modal hypergraph model with the following sentiment classification approaches:

- Lexicon-based method: Taboada *et al.* [4] build a lexicon-based classifier which employs linguistic rules to detect the polarity strength of reviews.
- SVM (Textual Modal): The TF-IDF features and SVM are widely-used baseline approaches to build sentiment classifiers. In this work, the SVM classifiers are trained with LibSVM,[2] a popular toolkit proposed in [35].
- SVM (Cross-Modal): Besides the textual information, we also apply sentiment score as an additional feature to train the classifiers.
- In addition to the above, we experimented with other two classifiers: NB and ME, which are widely applied in the field of text classification.

In Table 1, we show the experimental results of the proposed cross-modal hypergraph classifiers against the baselines. In detail, the fifth and sixth columns illustrate the results of SVM classifier by using textual modal only and combining both textual and sentimental feature, respectively. The rightmost two are the results computed by hypergraph model,

[1] http://nlp.stanford.edu/software/tagger.shtml
[2] http://www.csie.ntu.edu.tw/~cjlin/

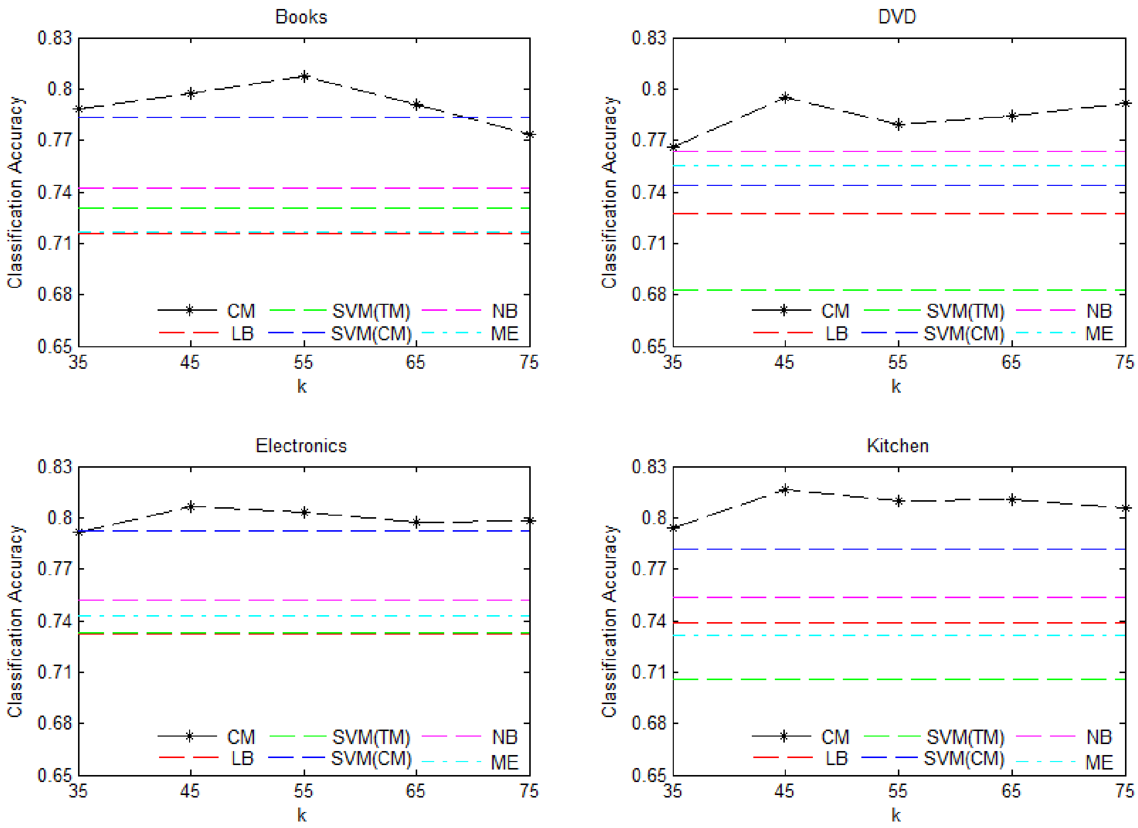


FIGURE 2. Sensitivity of k values for number of nearest neighbors on different datasets.

which denote the accuracy of constructing hypergraph by sentimental modal only and the cross-modal hypergraph model with textual modality and sentimental modality as inputs. More specifically, the results of baseline methods behave differently on various data sets. The improvements of constructing hypergraph are quite prominent. The accuracy has been increased in various degrees compared to the baseline methods. To be specific, the hypergraph classification model with only sentimental modal outperforms lexicon-based method by about 6% on the average accuracy. With regard to the cross-modal hypergraph model, it achieves nearly 2% on average improvements in comparison with the single-modal hypergraph. Moreover, the SVM classifier with cross-modal features input outperforms the method with single-modal feature input obviously. Therefore, we can conclude that combining different modal features can contribute to sentiment classification.

On the other hand, to evaluate the parameter sensitivity, we conduct experiments on how the hyperedge size (i.e., the number of k nearest neighbors) will affect the classification accuracy. We change k from 35 to 75 with a step of 10. As shown in Figure 2, the four line charts represent four domain datasets (books, DVD, electronics and kitchen, respectively), the performances are varying with the increasing of k. For the purpose of facilitating the comparison of the results produced by whether constructing hypergraph or not, we utilize five different color dotted lines in the corresponding charts to represent the results obtained by the baseline approaches respectively. The classification accuracy of cross-modal hypergraph (CMHG) is evidently superior to the baseline approaches in most cases, and it also shows the robustness of performance while k varies in such a large scale. Moreover, a better level of performance is achieved when k equals to 45 and 55.

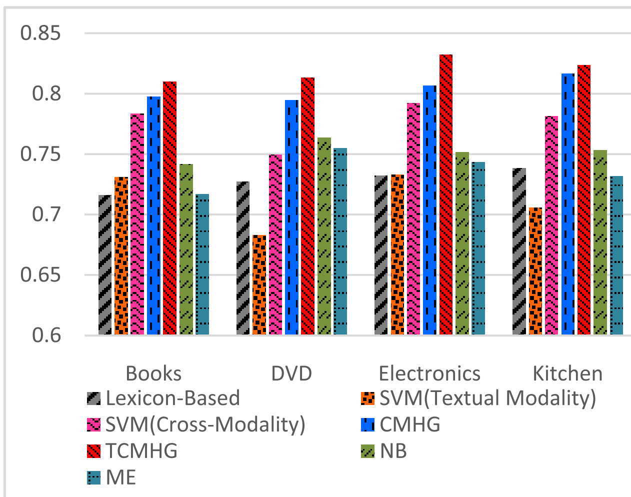


FIGURE 3. Performance comparison of different approaches.

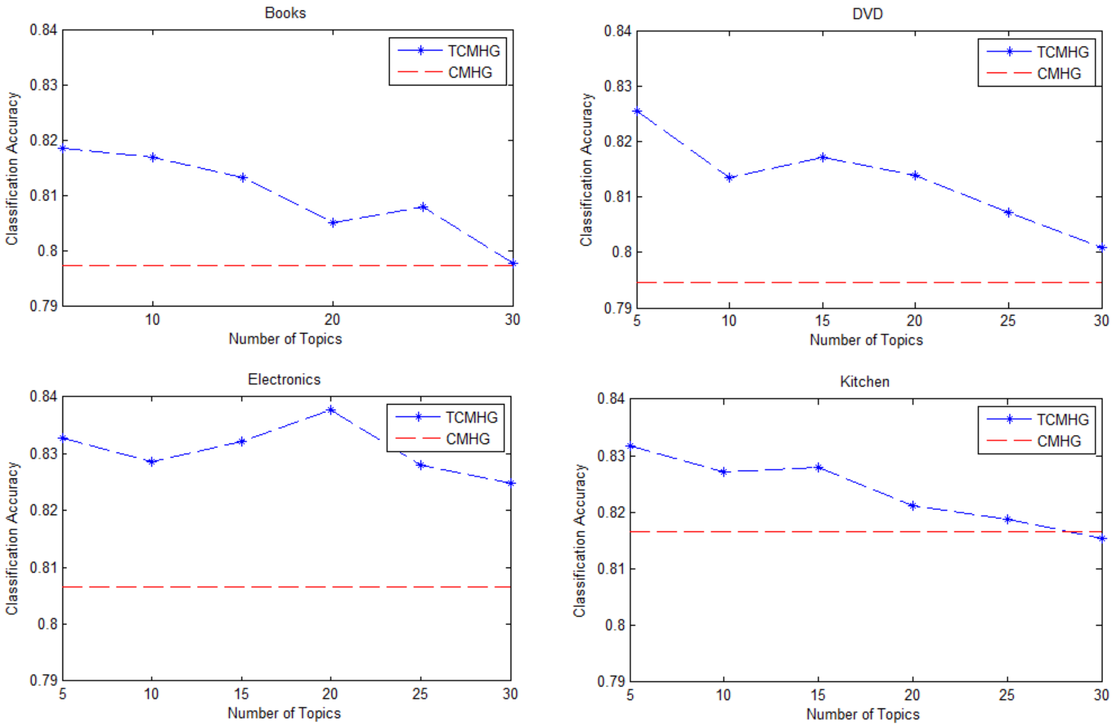


FIGURE 4. **Reaction on changing the number of topics.**

TABLE 2. **Classification accuracy of using diverse *k* and number of topics on Kitchen dataset.**

k/topics	**5**	**10**	**15**	**20**	**25**	**30**
35	0.8109	0.8209	*0.8267*	0.8208	0.8151	0.8151
45	***0.8318***	**0.8272**	**0.8279**	0.8211	**0.8188**	**0.8154**
55	*0.8315*	0.8254	0.8270	**0.8214**	0.8109	0.8021
65	*0.8224*	0.8212	0.8201	0.8094	0.7952	0.7872
75	*0.8305*	0.8212	0.8180	0.7951	0.7947	0.7866

C. TOPIC-BASED MIXTURE MODEL

For the topic-based mixture model TCMHG, we simply set the hyper-parameters of LDA to symmetric Dirichlet prior vectors, and use the default values of them. In addition, based on the experimental observation, we fix the parameter ε mentioned in soft clustering to 0.2. We choose to vary number of topics among 5, 10, 15, 20, 25 and 30.

In order to validate the effectiveness of the proposed TCMHG, we compare it against CMHG. As shown in Figure 3, the results of CMHG are computed when k is fixed to 45. For fair comparison, we conduct experiments on TCMHG model by using the same condition. Moreover, the results are obtained by taking the average of all six different number of topics. The results show that the model merging topic distribution soft clustering outperform the model without topic information by a rate of 1.25%, 1.85%, 2.58% and 0.72% better for books, DVD, electronics and kitchen, respectively. In addition, the running time is greatly reduced after the datasets have been divided into several subsets, due to the process of constructing hypergraph is much easier with a smaller number of vertexes. Nevertheless, the improvement of performance is not very remarkable in some datasets. The reason may lie into the unreasonable segmentation of topics.

Therefore, based on the mentioned problem, additional experiments are conducted to evaluate the influence produced by number of topics (which is corresponding to the number of clusters). Experimental results are reported in Figure 4, which is quite similar to Figure 2. For a remarkable comparison, we use the red line to denotes the performance computed by CMHG when k is set to 45 in each specific dataset. Accordingly, the same parameter settings are used in TCMHG except the division of clusters. As shown in Figure 4, the performance has been improved after LDA soft clustering in the overwhelming majority of circumstances. More specifically, the accuracy shows a decreasing trend with the number of clusters becomes larger, and the performances are usually favorable when the number of topics is relatively small (especially when it is set to be 5). We think that is because every dataset is about a specific domain, a small number of clusters is enough to obtain satisfied clustering effect.

Moreover, we also conduct experiments on different k values while the number of clusters is changing, and we report the classification accuracy in Table 2 (We only showed the results on using the Kitchen dataset, because the other three have similar trend). In Table 2, we show how these two parameters will influence the classification accuracy by fixing one and changing another. The greatest value of each column has been marked in bold. Similarly, the italic represents the maximum value in each row. As we can see, the highest classification accuracies usually appear when the hyperedge size is set to be 45 and when the number of topics equals to 5, which is according with what we described earlier.

VI. CONCLUSION

In this paper, a cross-modal hypergraph model has been presented to combine textual information and sentimental information simultaneously for sentiment classification, which is conducive to online service recommendations. Moreover, to take the global higher level information into consideration and alleviate the ambiguity of some specific words, LDA topic model has been further combined into our proposed cross-modal hypergraph model. Experimental results on four domain benchmark datasets clearly demonstrate that: (1) the proposed cross-modal hypergraph model can significantly improve the sentiment classification accuracy by comparison with the baseline methods; (2) the topic-based mixture model can further enhance the classification performance as well as reduce the computational cost. In this work, the adjustment of the parameters may influence the classification accuracy, so we will improve the robustness of our proposed model by tuning the parameters automatically in the future.

ACKNOWLEDGMENT

The authors thank Dr. Maite Taboada for providing the sentiment dictionaries.

REFERENCES

[1] D.-M. Koo, "Impact of tie strength and experience on the effectiveness of online service recommendations," *Electron. Commerce Res. Appl.*, vol. 15, pp. 38–51, Jan./Feb. 2016.

[2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. ACL Conf. Empirical Methods Natural Lang. Process.*, vol. 10. 2002, pp. 79–86.

[3] G. Wang, "Sentiment classification: The contribution of ensemble learning," *Decision Support Syst.*, vol. 57, pp. 77–93, Jan. 2014.

[4] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Comput. Linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[5] A. Ansari, S. Essegaier, and R. Kohli, "Internet recommendation systems," *J. Marketing Res.*, vol. 37, no. 3, pp. 363–375, 2000.

[6] Y. Lu, X. Kong, X, Quan, W. Liu, and Y. Xu, "Exploring the sentiment strength of user reviews," in *Proc. 11th Int. Conf. Web-Age Inf. Manage. (WAIM)*, Jiuzhaigou, China, Jul. 2010, pp. 471–482.

[7] H. Wang and M. Ester, "A sentiment-aligned topic model for product aspect rating prediction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2014, pp. 1–11.

[8] B. Xiang and L. Zhou, "Improving Twitter sentiment analysis with topic-based mixture modeling and semi-supervised training," in *Proc. Meeting Assoc. Comput. Linguistics*, 2014, pp. 1–6.

[9] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2009, pp. 375–384.

[10] Q. Mei, "Topic sentiment mixture: Modeling facets and opinions in weblogs," in *Proc. Int. World Wide Web Conf.*, 2007, pp. 171–180.

[11] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid," in *Proc. Conf. Empirical Methods Natural Lang. Process. Assoc. Comput. Linguistics*, 2010, pp. 56–65.

[12] I. Titov and R. McDonald, "Modeling online reviews with multi-grain topic models," in *Proc. Int. Conf. World Wide Web ACM*, 2008, pp. 111–120.

[13] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "An improved deep computation model based on canonical polyadic decomposition," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2017.2701797.

[14] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly supervised joint sentiment-topic detection from text," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 1134–1145, Jun. 2011.

[15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[16] J. Jagarlamudi, H. Daumé, III, and R. Udupa, "Incorporating lexical priors into topic models," in *Proc. Conf. Eur. Chapter Assoc. Comput. Linguistics Assoc. Comput. Linguistics*, 2012, pp. 204–213.

[17] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "PPHOPCM: Privacy-preserving High-order possibilistic c-means algorithm for big data clustering with cloud computing," *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2017.2701816.

[18] T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Proc. Meeting Assoc. Comput. Linguistics Int. Joint Conf. Natural Lang. Process. Asian Federation Natural Language Process.*, 2015, pp. 1353–1364.

[19] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*. Boston, MA, USA: Springer, 2012, pp. 415–463.

[20] T. H. Nguyen and K. Shirai, "Aspect-based sentiment analysis using tree kernel based relation extraction," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*, 2015, pp. 114–125.

[21] P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*, 2002, pp. 1–8.

[22] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 168–177.

[23] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1601–1608.

[24] S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 17–24.

[25] Y. Gao, M. Wang, Z. J. Zha, J. Shen, X. Li, and X. Wu, "Visual-textual joint relevance learning for tag-based social image search," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 363–376, Jan. 2013.

[26] Y. Huang, Q. Liu, S. Zhang, and D. N. Metaxas, "Image retrieval via probabilistic hypergraph ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 3376–3383.

[27] Y. Gao, M. Wang, D. Tao, R. Ji, and Q. Dai, "3-D object retrieval and recognition with hypergraph analysis," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4290–4303, Sep. 2012.

[28] F. Chen, Y. Gao, D. Cao, and R. Ji, "Multimodal hypergraph learning for microblog sentiment prediction," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jun./Jul. 2015, pp. 1–6.

[29] Y. Huang, Q. Liu, and D. Metaxas, "Video object segmentation by hypergraph cut," in *Proc. IEEE Conf. Computer Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1738–1745.

[30] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, nos. 1–2, pp. 177–196, Jan. 2001.

[31] D. Zhou, J. Huang, and B. Schölkopf, "Learning from labeled and unlabeled data on a directed graph," in *Proc. 22nd Int. Conf. Mach. Learn.*, 2005, pp. 1036–1043.

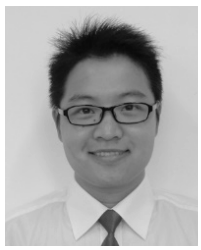
[32] A. Andreevskaia and S. Bergler, "When specialists and generalists work together: Overcoming domain dependence in sentiment tagging," in *Proc. ACL*, 2008, pp. 290–298.

[33] Q. Zhang, L. T. Yang, X. Liu, Z. Chen, and P. Li, "A tucker deep computation model for mobile multimedia feature learning," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 13, no. 3s, pp. 39:1–39:18, 2017.
[34] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proc. ACL*, vol. 7. 2007, pp. 1–8.
[35] C. C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 389–396, Apr. 2011.
[36] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.
[37] M. Eirinaki, S. Pisal, and J. Singh, "Feature-based opinion mining and ranking," *J. Comput. Syst. Sci.*, vol. 78, no. 4, pp. 1175–1184, Jul. 2012.
[38] Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang, "Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS)," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 193–202.
[39] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 1, pp. 5228–5235, 2004.
[40] A. Klašnja-Milićević, M. Ivanović, and A. Nanopoulos, "Recommender systems in e-learning environments: A survey of the state-of-the-art and possible extensions," *Artif. Intell. Rev.*, vol. 44, no. 4, pp. 571–604, Dec. 2015.
[41] Q. Zhang, T. T. Yang, Z. Chen, P. Li, and M. J. Deen, "Privacy-preserving double-projection deep computation model with crowdsourcing on cloud for big data feature learning," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2017.2732735.
[42] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

ZHIKUI CHEN received the B.E. degree in mathematics from Chongqing Normal University, Chongqing, China, in 1990, and the Ph.D. degree in solid mechanics from Chongqing University, Chongqing, in 1998. He is currently a Professor with the School of Software Technology, Dalian University of Technology, China. His research interests include Internet of Things and big data.

FEI LU received the B.S. degree in computer software from Nanchang University, Nanchang, China, in 2015. He is currently pursuing the M.S. degree with the School of Software Technology, Dalian University of Technology, Dalian, China. His research interests include sentiment analysis and text mining.

XU YUAN received the B.S. degree in information management from Carnegie Mellon University. He is currently an Associate Professor with the Dalian University of Technology, Dalian, China. His research interests are big data visualization analysis, mobile Internet, financial data mining, and electronic commerce.



FANGMING ZHONG received the B.S. and M.S degrees in software engineering from the Dalian University of Technology, Dalian, China, in 2012 and 2014, respectively, where he is currently pursuing the Ph.D. degree with the School of Software Technology. His research interests include multimodal learning, cross-modal retrieval, and subspace learning.