Multimedia Systems
DOI 10.1007/s00530-017-0557-6

SPECIAL ISSUE PAPER

User-aware topic modeling of online reviews

Xiaojia Pu[1] · **Gangshan Wu**[1] · **Chunfeng Yuan**[1]

Abstract The online reviews are one type of social media which are opinions generated by the users to comment on some special items. Since the sentiments are dependent on topics, probabilistic topic models have been widely used for sentiment analysis. However, most of existing methods only model the text, but rarely consider the users, who express the opinions, and the items, which the opinions are expressed on. Different users are usually concerned with different topics and use different sentiment expressions, a lenient user might tend to give positive review than a critical user. High-quality items tend to receive positive reviews than low-quality items. To better model the topics and sentiments, we argue that it is essential to explore reviews as well as users and items. To this end, we propose a novel model called User Item Sentiment Topic (UIST) which incorporates users and items for topic modeling and produces topic–word, user–topic, and item–topic distributions simultaneously. Extensive experiments on several datasets demonstrate the advantages and effectiveness of our method. The extracted topics with our method are more coherent and informative; consequently, the performance of sentiment classification is also improved. Furthermore, the user preference obtained with our method could be utilized for many personalized applications.

✉ Xiaojia Pu
puxiaojia@gmail.com

Gangshan Wu
gswu@nju.edu.cn

Chunfeng Yuan
cfyuan@nju.edu.cn

1 State Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing, China

Keywords Topic model · Online reviews · User preference · Sentiment topic

1 Introduction

With the advent of Web 2.0, people are very convenient to express and share their opinions on web regarding products, service, and global issues as well. These reviews are with great value and grow rapidly; therefore, to efficiently analyze them becomes essential [7, 17, 25].

Many research work has been performed to analyze users' sentiments and opinions from online reviews [17]. For a opinion target, the sentiment is often dependent on topics; therefore, it is more suitable to analyze the topic and sentiment simultaneously. To discover the opinion topics in the text, various topic models are proposed [11, 15, 19, 20, 37–39], most of them accomplish the task with only review texts, but rarely consider the user, who expresses the opinions, and the item, which the opinions is expressed on.

For different users, the preferences of both topics and sentiments are different, and they may have different sentiment expression preferences. A lenient user might tend to give higher rating than a critical user, for example, some users choose to use 'good' to describe a just-so–so product, but others may use 'good' to describe an excellent product. Beside the user bias, there is also an item bias. One may use the same opinion word to express different sentiment polarities for different items, for example, the opinion word 'long' can express a 'positive' feeling for battery life, but may have a 'negative' feeling for a camera's focus time. Besides, the items with good quality are easier to get positive reviews than low-quality items.

Therefore, to better model the sentiment topics, we argue that it is essential to explore reviews as well as users

Published online: 19 June 2017

and items. To this end, we propose a novel model called User Item Sentiment Topic (UIST) which incorporates users and items for topic modeling. In our method, the users, items, and reviews are modeled simultaneously in a generative process. Besides the topics, with our model, the user topic and item topic distributions can also be learned. Extensive experiments on several datasets demonstrate the advantages of our method. The extracted topics are more coherent and informative, and the performance of sentiment classification is also improved. Furthermore, the user preference obtained with our method could be utilized for many personalized applications.

The contribution of our work could be summarized as follows:

1. We propose a novel model which efficiently captures user and item information for topic modeling of online reviews. In comparison with existing methods, our method can obtain more coherent topics and improve the performance of document sentiment classification.
2. The user topic distribution obtained with our method implies the user preference at topic level. It is helpful for better understanding the user, and could be utilized for many applications, e.g., recommendation systems.
3. We conduct experiments on three datasets and show that our method is effective for topic modeling of online reviews.

The rest of this paper is organized as follows. Section 2 gives a brief review of some related work. Sections 3 and 4 present the proposed method and formulation in detail. In Sect. 5, we evaluate the performance through extensive experiments, and in Sect. 6 we conclude the paper and present the future work.

2 Related work

Statistical topical model is a well-known method to model the latent topics in an unsupervised approach [1, 10, 28, 33]. The data can be assigned to topics with different weights and these topics provide an intuitive interpretation.

For sentiment analysis, the topic models are often used to discover the opinion aspects as well as the sentiment towards the target [6, 11, 18–20, 38, 39]. Multi-grain LDA (MG-LDA) aims to discover the local aspects as well as the global topics [38]. Titov and McDonald [37] and Lu et al. [18] aim to extract aspect level sentiment summaries with rating for online reviews. The Joint Sentiment/Topic model (JST) [15] detects sentiment and topic simultaneously from text with a unsupervised approach. In JST, each document has a sentiment label distribution. Topics are associated with sentiment labels, and words are associated with both topics and sentiment labels [9]. The CFACTS model [11] extends the JST model to capture facet coherence in a review using Hidden Markov Model.

Though these models have made a great progress in topic modeling of online reviews, most of them ignore the author preference and item characteristics, which have curial effects in maintaining topic coherence as well as the sentiment polarities in reviews. An approach to capture author-specific topic preference is described in [34]; however, it neglects the item information. JMARS [6] aims to model the aspects, ratings, and sentiments for recommendation; however, the sentiment rating is needed, which is not suitable for many unsupervised scenarios.

Meanwhile, the user information has been utilized and explored with many other works [23, 26, 27, 29, 30]. The author-topic model uses the authorship information together with the text to learn topics, with which each author is a multinomial distribution over topics and each topic is a multinomial distribution over words [26]. However, for online reviews, besides the user, the item characteristics and sentiment should also be modeled, thus these methods are not suitable for this task.

There have been some supervised methods including some deep learning algorithms which explore the user and item information for document sentiment classification [3, 4, 8, 12, 13, 32, 36]. The target of these works is sentiment classification rather than the topic modeling of the reviews. For them, the labeled dataset is needed. In our work, we aim to provide an unsupervised topic modeling approach to model the sentiment topics with the incorporation of user and item information.

3 The structure of reviews

Generally, a review contains the opinions of an author (user) about some topics (aspects) of an item (object) e.g. product, service. One user may contribute multiple reviews towards different items, and one item may be commented by multiple users. Thus, the users, reviews, and items form a heterogeneous network as shown in Fig. 1.

From the perspective of users, they have different preference and styles of commenting. Some users tend to be very strict and give more negative reviews in average, while others may be the opposite. On the other hand, some users may focus on certain aspects of an item, but others may be concerned with the overall performance. Therefore, for the same item, different users may give different opinions.

From the perspective of items, each item has multiple aspects or topics for users to comment on. These aspects may have different qualities. If the quality of an aspect or

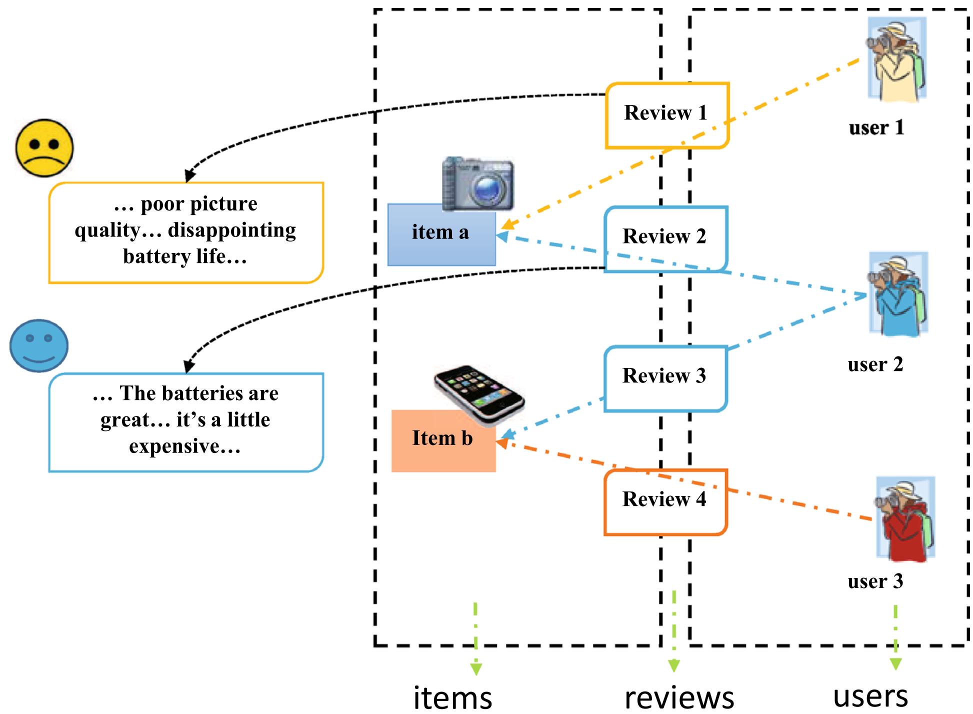


Fig. 1 The illustration of the structure of the reviews, an author may write multiple reviews towards different items and an item may be commented by multiple authors

item is good, most users tend to give positive comments, and vice versa.

Besides the user preference and item quality, there should be some common experience or judgements, for example, no one likes a bad but expensive product. This kind of experience is applicable to all the users and items.

Generally, the user preference can be mined from reviews commented by him, and the item quality can be mined from reviews commenting it. At the same time, the user preference and item quality both influence the analysis of the reviews associated with them.

4 The proposed model

In this section, we introduce the UIST model, which integrates the user preference and item quality to model the topics in a unified framework. With our model, the topic structure of the authors, reviews, and items can be learned. It can also help the estimation of sentiment at topic level.

4.1 Model formulation

The graphical illustration of our model is shown in Fig. 2. The model is designed and inspired based on the user's commenting process, in which a term may be generated from the user, the item or the experience. We adopt a switch variable x to control the influence of the user, the item (object), and the experience.

As shown in Fig. 2, assume that we have a corpus of D reviews, N_d denotes the number of terms in review d, K denotes the number of topics, V denotes the size of vocabulary, S denotes the number of sentiment labels, U denotes the number of users and O denotes the number of objects in the dataset. The α^e, α^u, α^o, γ^e, γ^u, γ^o, η and β are hyper-parameters and priors of Dirichlet distributions. For convenience, we also illustrate the notations in Table 1.

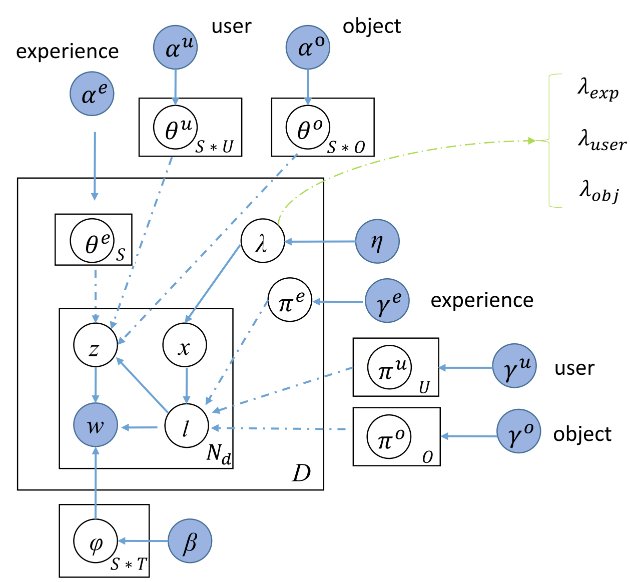


Fig. 2 The illustration of our proposed model

Table 1 Basic notations used in our paper

Variable	Description
d	A review or document
D	The number of reviews
N_d	The number of terms or words in review d
K	The number of topics
V	The size of vocabulary
s	The sentiment labels of reviews (positive, negative, etc.)
S	The number of sentiment labels
u	User
U	The number of users
K	The number of topics
w	Term or word
V	Vocabulary
o	The objects or items
O	The number of objects or items
l	The sentiment categories of reviews
z	Topic
x	Switch variable
l_{-i}	The sentiment categories of the words except the ith word
z_{-i}	The topic distribution of the words except the ith word
x_{-i}	The switch variable of the words except the ith word
$n^{k,s}$	The number of words which are denoted as sentiment s and topic k
$n^{k,s}_{w_i}$	The number of word w_i which is denoted as sentiment s and topic k
$\varphi_{i,k,s}$	The sentiment/topic-word distribution

The generative process of review in the model can be described as follows:

- For each document d, sample $\pi^e{}_d \sim Dir(\gamma^e)$.
- For each sentiment label s under each document d, sample $\theta^e{}_{d,s} \sim Dir(\alpha^e)$.
- For each user u, sample $\pi^u{}_u \sim Dir(\gamma^u)$.
- For each sentiment label s under each user u, sample $\theta^u{}_{u,s} \sim Dir(\alpha^u)$.
- For each object o, sample $\pi^o_o \sim Dir(\alpha^o)$.
- For each sentiment label s under each object o, sample $\theta^o{}_{o,s} \sim Dir(\alpha^o)$.
- For each document d, sample $\lambda_d \sim Dir(\eta)$.
- For each of the K topics k and sentiment labels s, sample $\varphi^s{}_k \sim Dir(\beta)$.
- For each of the N_d word tokens w_i in document d: choose $x_i \sim Multinomial(\lambda_d)$.
  - If $x_i = exp$, choose a sentiment label $l = s_i \sim \pi^e{}_d$, choose a topic $z_i \sim \theta^e{}_{d,s}$.
  - If $x_i = user$, choose a sentiment label $l = s_i \sim \pi^u{}_u$, choose a topic $z_i \sim \theta^u{}_{u,s}$.
  - If $x_i = obj$, choose a sentiment label $l = s_i \sim \pi^o{}_o$, choose a topic $z_i \sim \theta^o{}_{o,s}$.
  - Choose a word w_i from the distribution over words defined by topic z_i and sentiment label s_i, $w_i \sim \varphi^{s_i}_{z_i}$.

In comparison with JST model, our model adds a user and object layer which integrates the user preference and object quality to the estimation of review sentiment at topic level. Our model is an extension of both Author-Topic Model and JST. Through the latent variable x, we try to distinguish which term is associated with the author, which term is associated with the object and which term is associated with the experience.

4.2 Learning algorithm

There have been several typical methods for estimating the latent parameters in LDA model, such as the variational expectation maximization [1] and Gibbs sampling [33]. Gibbs sampling often yields relatively simple algorithms for approximate inference in high-dimensional models. Therefore, we select this approach for parameter estimation. The sampling equations for our model are listed below.

There are three sets of latent variable: l, z, x. The joint probability of the l, z, x and the words w can be factored into the following terms:

$$\begin{aligned} p(w,z,l,x) &= p(w|z,l)p(z,l|x) \\ &= p(w|z,l)p(z|l,\theta^e)p(l|x=exp)p(x=exp) \\ &\quad + p(z|l,\theta^u)p(l|x=user)p(x=user) \\ &\quad + p(z|l,\theta^o)p(l|x=obj)p(x=obj) \end{aligned} \tag{1}$$

We draw each $(l_i; z_i; x_i)$ pair as a block, conditioned on all other variables.

When $x = exp$, the sampling equation is:

$$\begin{aligned} & p(z_i=k, l_i=s, x=exp|w, z_{-i}, l_{-i}, x_{-i}) \\ & \propto \frac{n^{k,s}_{w_i}+\beta}{n^{k,s}+V\beta} \cdot \frac{n^e_{s,k}+\alpha^e}{n^e_s+K\alpha^e} \cdot \frac{n^e_s+\gamma^e}{n^e+S\gamma^e} \cdot \frac{n^e_d+\eta_{exp}}{N_d+\Sigma\eta_x} \end{aligned} \tag{2}$$

where n^e_d is the number of times that words are generated from experience in review d; n^e_s is the number of times that words assigned sentiment label s are generated from experience in the review d; $n^e_{s,k}$ is the number of times that words assigned sentiment label s and topic k are from experience in the review d; $n^{k,s}$ is the number of times that words are assigned sentiment label s and topic k. $n^{k,s}_{w_i}$ is the number of times that word in the position i are assigned sentiment label s and topic k.

When $x = user$, the conditional posterior for z_i, l_i is:

$$p(z_i = k, l_i = s, x = user|w, z_{-i}, l_{-i}, x_{-i}) \propto \frac{n_{w_i}^{k,s} + \beta}{n^{k,s} + V\beta} \cdot \frac{n_{s,k}^u + \alpha^u}{n_s^u + K\alpha^u} \cdot \frac{n_s^u + \gamma^u}{n^u + S\gamma^u} \cdot \frac{n_d^u + \eta_{user}}{N_d + \Sigma\eta_x} \quad (3)$$

where n^u is the number of times that words are generated from user u; u_d^u is the number of times that words in review d are generated from user u; u_s^u is the number of times that words assigned sentiment label s are generated from user u; $u_{s,k}^u$ is the number of times that words assigned sentiment label s and topic k are generated from user u.

When $x = obj$, the conditional posterior for z_i, l_i is:

$$p(z_i = k, l_i = s, x = obj|w, z_{-i}, l_{-i}, x_{-i}) \propto \frac{n_{w_i}^{k,s} + \beta}{n^{k,s} + V\beta} \cdot \frac{n_{s,k}^o + \alpha^o}{n_s^o + K\alpha^o} \cdot \frac{n_s^o + \gamma^o}{n^o + S\gamma^o} \cdot \frac{n_d^o + \eta_{obj}}{N_d + \Sigma\eta_x} \quad (4)$$

where n^o is the number of times that words are generated from object o; n_d^o is the number of times that words are generated from object o in review d; n_s^o is the number of times that words assigned sentiment label s are generated from object o; $n_{s,k}^o$ is the number of times that words assigned sentiment label s and topic k are generated from object o.

With the posterior distributions calculated by above equations, the sampling processes could be performed for parameter estimation. It is then used to approximate the per-corpus sentiment topic word distribution:

$$\varphi_{i,k,s} = \frac{n_{w_i}^{k,s} + \beta}{n^{k,s} + V\beta}. \quad (5)$$

4.3 Sentiment prediction of the document

The overall approximated per-document sentiment distribution, e.g., for review d, can be estimated as follows:

$$p(l|d) \sim \frac{n_s^e + \gamma^e}{n^e + S\gamma^e} \cdot \frac{n_d^e + \eta_{exp}}{N_d + \Sigma\eta_x} + \frac{n_s^u + \gamma^u}{n^u + S\gamma^u} \cdot \frac{n_d^u + \eta_{user}}{N_d + \Sigma\eta_x} + \frac{n_s^o + \gamma^o}{n^o + S\gamma^o} \cdot \frac{n_d^o + \eta_{obj}}{N_d + \Sigma\eta_x} \quad (6)$$

As shown in equation 6, the sentiment distribution for a review is estimated based on three factors: the experience, the user preference, and the object quality.

We only consider the probability of positive and negative labels for a given document, with the neutral label probability being ignored. There are two reasons for this. First, sentiment classification is effectively a binary classification problem, i.e., documents are being classified either as positive or negative, without the alternative of neutral. Second, the prior information we incorporated merely contributes to the positive and negative words, and consequently there will be much more influence on the probability distribution of positive and negative labels for a given document, rather than the distribution of neutral labels in the given document. Therefore, we define that a document d is classified as a positive-sentiment document if the probability of a positive-sentiment label $p(l_{pos}|d)$ is greater than its probability of negative sentiment label $p(l_{neg}|d)$, and vice versa.

4.4 Sub-models

In our model, an additional multinomial distribution λ is used to indicate the probability of the word generated from each factor, i.e., the experience, the author and the object.

$$\lambda = \{\lambda_{exp}, \lambda_{user}, \lambda_{obj}\}$$

$$\lambda_{exp} + \lambda_{user} + \lambda_{obj} = 1$$

Especially, if λ_{user} is set to 0, the words will be generated from experience or object, not from user. Then, our model is degraded to another model called Item Sentiment Model (IST), which only explores the item-review graph. Similarly, if only λ_{obj} is set to 0, our model is transformed to a simpler model called User Sentiment Model (UST) which is based on the user-review graph.

5 Experiments

In the experiments, we evaluate our proposed model through the performance of document sentiment classification and the extracted topics. Our main target is to evaluate the performance for topic modeling, and sentiment classification is an indirect approach for evaluation.

5.1 Experimental setup

5.1.1 Datasets

We conduct extensive experiments in both English and Chinese data and evaluate the performance of our method. For English data, the movie review polarity dataset 2.0[1] is utilized [24]. This dataset contains 2000 movie reviews from a unprocessed pool of 27,886 html files from IMDB. We extract the authors and movie ids of the 2000 reviews from the pool. For Chinese data, both the restaurant and movie reviews are utilized.

For the English data, the reviews are processed using stemmed lowercased unigram words. The punctuation, numbers, stop words, and other non-alphabet characters are removed. For the Chinese data, word segmentation is performed with the Stanford Word Segmenter.[2] Table 2 summarizes the dataset statistics after preprocessing.

5.1.2 Implementation details

The UIST has eight Dirichlet prior parameters, which influence the convergence of Gibbs sampling but not much the output results [44]. Generally we set $\alpha^e = 0.1$, $\alpha_u = 0.1$, $\alpha^o = 0.1$, $\beta = 0.05$, $S = 2$, and $\eta = 0.5$ for all experiments. For the rest hyper-parameters γ^e, γ^u and γ^o, we set to 0.1 for positive-sentiment label and 2 for negative sentiment label.

Sentiment prior We choose MPQA subjectivity lexicon for English data, which is also used in JST [15]. The sentiment prior for Chinese data is the HowNet sentiment lexicon.[3]

5.1.3 Methods for comparison

For topic modeling of the reviews, we compare our model with JST [15] as well as our sub models. The methods for comparison are listed below:

- **JST** The joint sentiment topic model [15] is extend from LDA, and is a classic topic model for sentiment analysis. The online code[4] is utilized in our experiments.
- **UST** It is the sub-model of our model which only explores the user-review graph without the item information.
- **IST** It is also a sub model of our model which only explores the item-review graph without the user information.

For sentiment classification, we also compare with some other methods as shown in Table 4, including unsupervised, semisupervised, and supervised methods. Here, we need to point out that the topic models are unsupervised; therefore, in the classification task, they are usually not very competitive with some supervised methods.

[1] http://www.cs.cornell.edu/people/pabo/movie-review-data/.

[2] http://www-nlp.stanford.edu/software/segmenter.html.

[3] http://www.keenage.com/html/c_index.html.

[4] https://github.com/linron84/JST.

Table 2 Some statistics of the dataset

Dataset (reviews)	# Docs (pos/neg)	Avg. length	# Users	# Objects
Movie (MR) (English)	2000 (1000/1000)	313	312	1107
Restaurant (Chinese)	8000 (4000/4000)	413	365	3428
Movie (Chinese)	8000 (4000/4000)	304	4570	435

5.2 Evaluation metric

Finding an objective metric to evaluate the quality of topics is hard. Some commonly used metrics, e.g. the perplexity or the likelihood of the held-out data, cannot directly measure the semantic coherence of the learned topics [2]. Presented quantitative methods to measure the semantic coherence of the learned topics, and found that the likelihood of the held-out data is not always a good indicator. Measuring the semantic coherence of the learned topics has received increasing attention, in [22], PMI is used to measure the semantic coherence of topics. We adopt the same topic coherence metric for the comparison in our experiments.

For a given topic T, the top N most probable words $w_1, w_2, \ldots, word_N$ are chosen, and the PMI score is calculate as the average relatedness of each pair of these words:

$$PMI - Score(T) = \frac{2}{N(N-1)} \sum_{i \le i < j \le N} log \frac{p(w_i, w_j)}{p(w_i)p(w_j)}, \quad (7)$$

where $p(w_i, w_j)$ is the joint probability of words w_i and w_j co-occurrence in the same document, while $p(w_i)$ is the marginal probability of word w_i appearing in a document. These probabilities are computed from a much larger corpus. We set $N = 15$ in our analysis.

5.3 Results in document sentiment classification

With our model, as shown in Eq. 6, the document sentiment is classified based on the probability of a sentiment label-given document. For sentiment classification, accuracy is adopted as the evaluation metric, which is commonly used for text classification.

5.3.1 Sentiment classification results in English data

For English data, following the approach in [24], we also perform subjective detection in MR dataset for sentiment classification. For convenience, the filtered subjective dataset is called Sub. MR. We first compare our method with IST (Item Sentiment Topic), UST (User Sentiment

Table 3 Sentiment classification results (accuracy) of different methods

Lexicon/method	IST	UST	UIST	JST
MR	0.827	0.832	**0.845**	0.828
sub. MR	0.855	0.863	**0.871**	0.848

The bold values indicate better results than other methods

Table 4 Sentiment classification results (accuracy) in comparison with some state-of-the-art methods

Method	Accuracy
Eigen vector clustering [5]	0.709
LSM: unsupervised with prior info [16]	0.741
SO-CAL: full lexicon [35]	0.7637
NMTF: semisupervised (40% doc. label) [14]	0.735
RAE (semisupervised recursive auto-encoders) [31]	0.768
RAE (supervised recursive auto-encoders) [31]	0.777
Supervised tree-CRF [21]	0.773
SVM [24]	0.8545
SVM + sub. MR [24]	0.8715
UIST (sub. MR)	**0.871**

The bold value indicates the result of our method

Table 5 Sentiment classification results of different methods in Chinese data

Method	IST	UST	UIST	JST
Restaurant	0.711	0.724	**0.736**	0.702
Movie	0.746	0.684	**0.745**	0.684

The bold values indicate better results than other methods

Topic) and JST. IST and UST are sub-models of UIST. For JST, both the user and item information is not incorporated.

From Table 3, we can observe that under all the data setups, the results of our UIST method are the best. Besides, the performances in Sub. MR is better than original MR dataset. The results of IST and UST are comparable to the result of JST. In the Sub. MR dataset, they are slightly better than JST. The results validate our hypothesis that user preference and item quality are both beneficial for topic modeling, and based on the extracted sentiment topics, the sentiment classification results are also improved.

The performance of UST is slightly better than IST. This result indicates that in MR dataset, the users may play a more important role than objects, i.e., the movies. Another possible reason is that in the MR dataset, the number of movies is larger than users; therefore, the item-review graph is sparser than user-review graph which leads to the insufficient model parameter estimation.

A large number of works have been reported on the *MR* dataset, we compare the performance of our approach to them on this dataset. Table 4 shows the accuracy of all the methods. SO-CAL is a rule-based method, Eigen vector and LSM are both unsupervised methods; with these methods, the sentiment lexicon is usually needed. NMTF is semisupervised method which takes advantage of negative matrix factorization. The RAE method is a representation learning approach. We also compare our method with several supervised methods, i.e., SVM and Tree-CRF.

From Table 4, we can find that our method is better than all the unsupervised and semisupervised methods; however, in comparison with SVM+Sub. MR, the accuracy our method is a little worse. Since the topic modeling approach is unsupervised with no labeled data, it is reasonable that the performance is less competitive with some supervised methods.

5.3.2 *Sentiment classification results in Chinese data*

In Chinese data, we also compare our method with JST. The results in Table 5 show that our UIST model performs better than JST.

There is an interesting result in Table 5 that IST is better than UST in Chinese movie data, contrary to the result in MR data. From Table 2, we can find that the number of users in MR is smaller than the number of movies; however, in Chinese movie data, the number of movies is smaller than the number of users. Since the data is too sparse, it becomes difficult to learn the user preference. Based on this observation, we can conclude that for accurate parameter estimation, sufficient data is needed. The results in restaurant data also validates the conclusion.

5.3.3 *Results with different topic numbers in English data*

We also analyze the influence of topic numbers in English data. The MPQA subjectivity lexicon is adopted as the prior sentiment knowledge on the subjective MR dataset. Figure 3 shows the sentiment classification accuracy of our model with different number of topics. The number of topics, i.e., K, is set as 50, 100 and 150, respectively.

From Fig. 3, we can see that when the number of topics increases, the performance of sentiment classification also grows, which indicates that the learned topic information indeed helps in sentiment classification. We can also find that whatever K is set, the performance of UIST is better than UST and IST. In comparison with UST and IST, UIST incorporates both the user preference and item quality which leads to better extracted sentiment

Fig. 3 Sentiment classification performance (accuracy) with different topic numbers

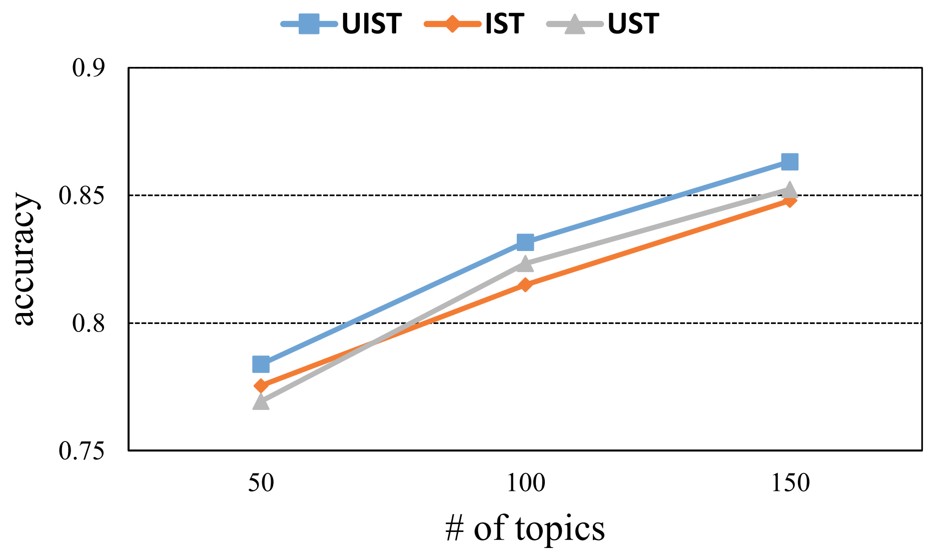


Table 7 Examples of the extracted topics with our method

T_1 (pos.)	T_2 (pos.)	T_3 (pos.)	T_4 (neg.)	T_5 (neg.)	T_6 (neg.)
Like	Dinosaur	Funni	Hollow	Polic	Poorly
Good	Park	Comedy	Sleepi	Action	Violence
Love	Jurass	Laugh	Hard	Cop	Comic
Star	Island	Joke	Bore	Fight	Early
Great	Adventur	Fun	Waste	Chasecr	Someth
Effect	Rocket	Eye	Crane	Thriller	Not
Real	Laura	Talk	Horseman	Crime	Long
Funni	Fund	Hour	Particularly	Explos	Every
Origin	Uplift	Act	Headless	Plot	Support
Right	Grant	Moment	Dull	Villain	Type
Moment	Skillful	Close	Murder	Gun	Somewhat
Friend	Sky	Scene	Town	Partner	Question
Live	Excellent	Picture	Wood	Laugh	Fall

topics. Therefore, the sentiment classification results are also better.

5.4 Results of the extracted topics

We conduct both quantitative and qualitative analysis of the extracted topics. For the quantitative analysis, the topic coherence is calculated. The number of topics extracted here is 150.

5.4.1 Topic coherence

For the topic coherence, the PMI scores of all methods are presented in Table 6.

Table 6 Topic coherence with different models

Method	IST	UST	UIST	JST
MR (English)	1.621	1.757	1.939	1.515
Restaurant	0.922	1.051	1.192	0.871
Movie (Chinese)	1.121	0.823	1.336	0.781

We can observe that our method yields the highest PMI score, this is reasonable since our method incorporates both the user preference and item quality. In comparison with JST, all our models including UIST, IST, and UST are better. In MR and restaurant data, UST is slightly better than IST, while in Chinese movie data, IST is slightly better than UST. This is consistent with results in sentiment classification.

The PMI scores in Table 6 suggest that our method can produce more coherent topics. With the incorporation of users and items, the performance of topic modeling of online reviews is improved. This is a very encouraging result.

5.4.2 Qualitative analysis

To further evaluate the quality of the topics, we also select some examples of the results to have an intuitive understanding. The top words (most probable word) for each topic (with sentiment) distribution could approximately reflect the meaning of the topic. Table 7 shows the selected examples of the extracted topics. The left three topics of Table 7 are generated under the positive-sentiment label, while the right three topics are generated under the negative sentiment label, each of which is represented by the top 13 words.

As shown in Table 7, the extracted topics are quite informative and coherent. For example, topic T_1 is very positive review comments for a movie; topic T_2 is apparently the opinion about the movie "Jurassic Park"; topic T_3 is likely to be the positive comments for comedy. Topic T_4 is a negative review comment; topic T_5 is probably about gangster movies; topic T_6 is likely to be the negative opinion about actor.

5.5 User sentiment/topic preference

Figures 4, 5, and 6 show the topic distributions of several users. The horizontal axis indicates the topic and the vertical axis indicates the probability or strength of the sentiment in each topic.

Figure 4 shows the topic distribution in movie data under the positive label, while Fig. 5 shows the distribution in movie data under the negative label. Towards the movies, author 8 likes the 'actor', 'comedy'; however, author

Fig. 4 User preference of sentiment/topic in movie reviews (positive)

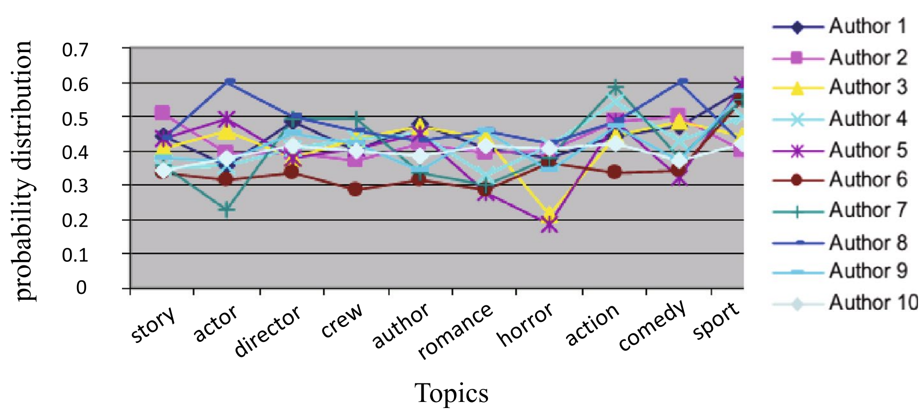


Fig. 5 User preference of sentiment/topic in movie reviews (negative)

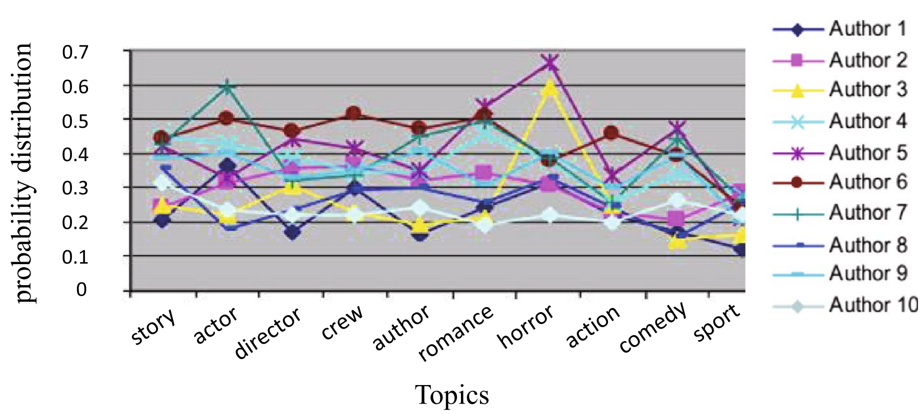
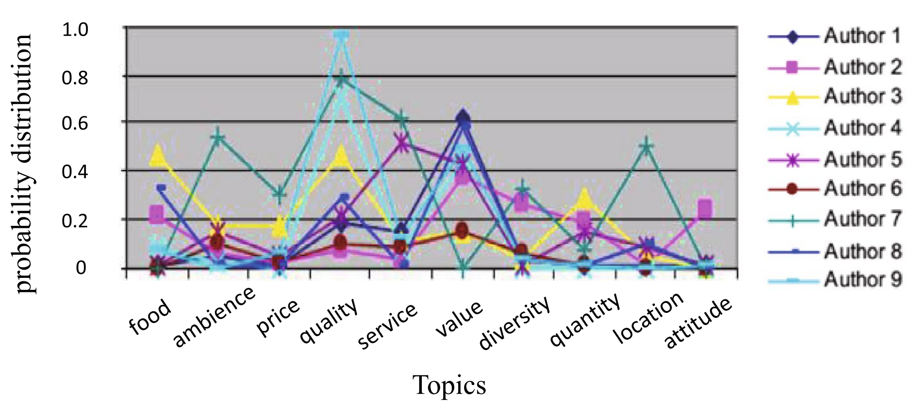


Fig. 6 User preference of sentiment/topic in restaurant reviews (positive)



7 likes more about ‘action’, which validates our claim that the topic and sentiment preference of the users differ a lot.

Figure 6 shows the topic distribution of several users in restaurant reviews under the positive label. We can observe that most users give the positive rating due to satisfaction with ‘quality’; however, author 5 is mainly due to ‘service’. This also validates that for different users, towards the same kind of object, the topics or aspects who care mostly are different.

The user topic distribution suggests the user preference, which is useful for many personalized applications, e.g., recommendations systems. Take movie for example, for different movies, the themes and stories are usually different, as well as the actors and directors. With our method, which factor and to which extent the users mostly like can be estimated accurately. That not only helps to clearly understand the underlying factors of users to generate such opinions, but also provides the opportunity to further recommend the possible interested movies.

Summary We have presented comprehensive experiments using three different datasets. The datasets cover two domains and two languages. We evaluate the effectiveness both in sentiment classification and extracted topics. The model results in an effective sentiment distribution of the documents and a coherent word representation of topics. The performance outperforms classical topic models.

6 Conclusion

In this paper, for topic modeling of online reviews, we propose a novel topic model which incorporates user and item information. The proposed model can not only produce sentiment topic distribution of each review, but also generate the sentiment topic distribution of each user and each item simultaneously in an unsupervised manner. Experiments demonstrate the effectiveness of our method. With our model, the extracted topics are more coherent. The sentiment polarity can also be predicted more accurately. The user topic distribution implies the user preference at topic level and is helpful for many applications, e.g., recommendation systems. With the introduction of more variables, the computational complexity increases. In the future, similar with many other media processing methodologies [40–43], a possible solution is to design a parallel framework for our model.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003). http://dl.acm.org/citation.cfm?id=944919.944937
2. Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., Blei, D.M.: Reading tea leaves: how humans interpret topic models. In: Proceedings of the 2009 Advances in Neural Information Processing Systems, pp. 288–296 (2009)
3. Chen, H., Sun, M., Tu, C., Lin, Y., Liu, Z.: Neural sentiment classification with user and product attention. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 1650–1659. Association for Computational Linguistics (2016). http://aclweb.org/anthology/D16-1171
4. Chen, T., Xu, R., He, Y., Xia, Y., Wang, X.: Learning user and product distributed representations using a sequence model for sentiment analysis. IEEE Comput. Intel. Mag. **11**(3), 34–44 (2016). doi:10.1109/MCI.2016.2572539
5. Dasgupta, S., Ng, V.: Topic-wise, sentiment-wise, or otherwise? identifying the hidden dimension for unsupervised text classification. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, 6-7 August 2009, Singapore, A meeting of SIGDAT, a Special Interest Group of the ACL, pp. 580–589 (2009). http://www.aclweb.org/anthology/D09-1061
6. Diao, Q., Qiu, M., Wu, C.Y., Smola, A.J., Jiang, J., Wang, C.: Jointly modeling aspects, ratings and sentiments for movie recommendation (jmars). In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’14, pp. 193–202. ACM, New York, NY, USA (2014). doi:10.1145/2623330.2623758
7. Fang, Q., Xu, C., Sang, J., Hossain, M.S., Ghulam, M.: Word-of-mouth understanding: Entity-centric multimodal aspect-opinion mining in social media. IEEE Trans. Multimedia **17**(12), 2281–2296 (2015). doi:10.1109/TMM.2015.2491019
8. Gao, W., Yoshinaga, N., Kaji, N., Kitsuregawa, M.: Modeling user leniency and product popularity for sentiment classification. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 1107–1111. Asian Federation of Natural Language Processing, Nagoya, Japan (2013). http://www.aclweb.org/anthology/I13-1156
9. He, Y., Lin, C., Alani, H.: Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pp. 123–131. Association for Computational Linguistics (2011)
10. Hofmann, T.: Probabilistic latent semantic indexing. In: Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’99, pp. 50–57. ACM, New York, NY, USA (1999). doi:10.1145/312624.312649
11. Lakkaraju, H., Bhattacharyya, C., Bhattacharya, I., Merugu, S.: Exploiting coherence for the simultaneous discovery of latent facets and associated sentiments. In: Proceedings of the 2011 SIAM International Conference on Data Mining (2011)
12. Li, F., Liu, N., Jin, H., Zhao, K., Yang, Q., Zhu, X.: Incorporating reviewer and product information for review rating prediction. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three, IJCAI’11, pp. 1820–1825. AAAI Press (2011). doi:10.5591/978-1-57735-516-8/IJCAI11-305, 10.5591/978-1-57735-516-8/IJCAI11-305
13. Li, F., Wang, S., Liu, S., Zhang, M.: Suit: A supervised user-item based topic model for sentiment analysis. In: AAAI, pp. 1636–1642 (2014)
14. Li, T., Zhang, Y., Sindhwani, V.: A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In: ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing

of the AFNLP, 2-7 August 2009, Singapore, pp. 244–252 (2009). http://www.aclweb.org/anthology/P09-1028
15. Lin, C., He, Y.: Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th ACM conference on Information and knowledge management, pp. 375–384. ACM (2009)
16. Lin, C., He, Y., Everson, R.: A comparative study of bayesian models for unsupervised sentiment detection. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning, CoNLL 2010, Uppsala, Sweden, July 15-16, 2010, pp. 144–152 (2010). http://aclweb.org/anthology/W/W10/W10-2918.pdf
17. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. **5**(1), 1–167 (2012). doi:10.2200/S00416ED1V01Y201204HLT016
18. Lu, Y., Zhai, C., Sundaresan, N.: Rated aspect summarization of short comments. In: Proceedings of the 18th international conference on World wide web, pp. 131–140. ACM (2009)
19. McAuley, J., Leskovec, J.: Hidden factors and hidden topics: understanding rating dimensions with review text. In: Proceedings of the 7th ACM conference on Recommender systems, pp. 165–172. ACM (2013)
20. Moghaddam, S., Ester, M.: On the design of lda models for aspect-based opinion mining. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM '12, pp. 803–812. ACM, New York, NY, USA (2012). doi:10.1145/2396761.2396863, 10.1145/2396761.2396863
21. Nakagawa, T., Inui, K., Kurohashi, S.: Dependency tree-based sentiment classification using crfs with hidden variables. In: Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA, pp. 786–794 (2010). http://www.aclweb.org/anthology/N10-1120
22. Newman, D., Lau, J.H., Grieser, K., Baldwin, T.: Automatic evaluation of topic coherence. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 100–108. Association for Computational Linguistics (2010)
23. Ovsjanikov, M., Chen, Y.: Topic modeling for personalized recommendation of volatile items. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 483–498. Springer (2010)
24. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42th Annual Meeting of the Association for Computational Linguistics (ACL) (2004)
25. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (2008)
26. Rosen-Zvi, M., Chemudugunta, C., Griffiths, T., Smyth, P., Steyvers, M.: Learning author-topic models from text corpora. ACM Trans. Inf. Syst. (TOIS) **28**(1), 4 (2010)
27. Sang, J.: User-centric cross-osn multimedia computing. In: Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, MM '15, Brisbane, Australia, October 26 - 30, 2015, pp. 1333–1334 (2015). doi:10.1145/2733373.2807423
28. Sang, J., Xu, C.: Browse by chunks: topic mining and organizing on web-scale social media. ACM Trans. Multimed. Comput. Commun. Appl. **7**(1), 30 (2011)
29. Sang, J., Xu, C.: Right buddy makes the difference: an early exploration of social relation analysis in multimedia applications. In: ACM International Conference on Multimedia, pp. 19–28. ACM (2012)
30. Sang, J., Xu, C., Liu, J.: User-aware image tag refinement via ternary semantic analysis. IEEE Trans. Multimed. **14**(3), 883–895 (2012)
31. Socher, R., Pennington, J., Huang, E.H., Ng, A.Y., Manning, C.D.: Semi-supervised recursive autoencoders for predicting sentiment distributions. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27-31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL, pp. 151–161 (2011). http://www.aclweb.org/anthology/D11-1014
32. Song, K., Feng, S., Gao, W., Wang, D., Yu, G., Wong, K.F.: Personalized sentiment classification based on latent individuality of microblog users. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI'15, pp. 2277–2283. AAAI Press (2015)
33. Steyvers, M., Griffiths, T.: Probabilistic topic models. Handb. Latent Semant. Anal. **427**(7), 424–440 (2007)
34. Subhabrata Mukherjee, G.B., Joshi, S.: Joint author sentiment topic model. In: Proceedings of the 2014 SIAM International Conference on Data Mining (SDM'14) (2014)
35. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Comput. Linguist. **37**(2), 267–307 (2011)
36. Tang, D., Qin, B., Liu, T.: Learning semantic representations of users and products for document level sentiment classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 1014–1023. Association for Computational Linguistics (2015). doi:10.3115/v1/P15-1098. http://aclweb.org/anthology/P15-1098
37. Titov, I., McDonald, R.: A joint model of text and aspect ratings for sentiment summarization. In: ACL-08: HLT, pp. 308–316. Association for Computational Linguistics (2008)
38. Titov, I., McDonald, R.: Modeling online reviews with multi-grain topic models. In: Proceedings of the 17th International Conference on World Wide Web, WWW '08, pp. 111–120. ACM, New York, NY, USA (2008). doi:10.1145/1367497.1367513., 10.1145/1367497.1367513
39. Wang, S., Chen, Z., Liu, B.: Mining aspect-specific opinion using a holistic lifelong topic model. In: Proceedings of the 25th International Conference on World Wide Web, WWW '16, pp. 167–176. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2016). doi:10.1145/2872427.2883086, 10.1145/2872427.2883086.
40. Yan, C., Zhang, Y., Dai, F., Wang, X., Li, L., Dai, Q.: Parallel deblocking filter for hevc on many-core processor. Electron. Lett. **50**(5), 367–368 (2014). doi:10.1049/el.2013.3235
41. Yan, C., Zhang, Y., Dai, F., Zhang, J., Li, L., Dai, Q.: Efficient parallel hevc intra-prediction on many-core processor. Electron. Lett. **50**(11), 805–806 (2014). doi:10.1049/el.2014.0611
42. Yan, C., Zhang, Y., Xu, J., Dai, F., Li, L., Dai, Q., Wu, F.: A highly parallel framework for hevc coding unit partitioning tree decision on many-core processors. IEEE Signal Process. Lett. **21**(5), 573–576 (2014). doi:10.1109/LSP.2014.2310494
43. Yan, C., Zhang, Y., Xu, J., Dai, F., Zhang, J., Dai, Q., Wu, F.: Efficient parallel framework for hevc motion estimation on many-core processors. IEEE Trans. Circuits Syst. Video Technol. **24**(12), 2077–2089 (2014). doi:10.1109/TCSVT.2014.2335852
44. Zhou, D., Bian, J., Zheng, S., Zha, H., Giles, C.L.: Exploring social annotations for information retrieval. In: Proceedings of the 17th international conference on World Wide Web, pp. 715–724. ACM (2008)