Information Processing and Management 51 (2015) 444–457

Contents lists available at ScienceDirect

Information Processing and Management

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

Mining affective text to improve social media item recommendation

Jianshan Sun [a,*], Gang Wang [a], Xusen Cheng [b], Yelin Fu [c]

[a] School of Management, Hefei University of Technology, Hefei, Anhui, PR China
[b] School of Information, University of International Business and Economics, Beijing, PR China
[c] School of Management, University of Science and Technology of China, Hefei, Anhui, PR China

ARTICLE INFO

Article history:
Received 15 August 2013
Received in revised form 1 June 2014
Accepted 18 September 2014
Available online 27 October 2014

Keywords:
Social media
Recommender system
Sentiment classification
OCCF

ABSTRACT

Social media websites, such as YouTube and Flicker, are currently gaining in popularity. A large volume of information is generated by online users and how to appropriately provide personalized content is becoming more challenging. Traditional recommendation models are overly dependent on preference ratings and often suffer from the problem of "data sparsity". Recent research has attempted to integrate sentiment analysis results of online affective texts into recommendation models; however, these studies are still limited. The one class collaborative filtering (OCCF) method is more applicable in the social media scenario yet it is insufficient for item recommendation. In this study, we develop a novel sentiment-aware social media recommendation framework, referred to as SA_OCCF, in order to tackle the above challenges. We leverage inferred sentiment feedback information and OCCF models to improve recommendation performance. We conduct comprehensive experiments on a real social media web site to verify the effectiveness of the proposed framework and methods. The results show that the proposed methods are effective in improving the performance of the baseline OCCF methods.

1. Introduction

Lately, social media websites (e.g., YouTube[1] and Flicker[2]) are increasingly receiving attention. A rapid convergence of online content sharing network websites has been observed in recent years. A large volume of content can be generated and diffused by users in these social media websites. For example, it was reported in 2012 that YouTube received 60 h of uploads per minute and 4 billion views per day[3]. In such an environment, there is an urgent requirement for an intelligent tool to effectively recommend social media items. Great challenges on traditional recommendation techniques to provide personalized content to users are seen, due to the dynamic behavior of users in social media websites and the volume of content they generate. Traditional recommendation models assume user preference ratings are available and often suffer from the problem of "data sparsity" (Adomavicius & Tuzhilin, 2005). In the real-world social media websites, it is hard to obtain rating information which means that traditional models are limited in their functionality. To address this issue, some researchers have explored rich user generated content as a supplementary source to support personalized recommendation. Tag-aware recommender systems were

* Corresponding author. Tel.: +86 18956561661; fax: +86 0551 62904991.
E-mail address: Sunjs9413@gmail.com (J. Sun).

[1] http://www.youtube.com.
[2] http://www.flicker.com.
[3] http://www.telegraph.co.uk/technology/news/9033765/YouTube-uploads-hit-60-hours-per-minute.html.

http://dx.doi.org/10.1016/j.ipm.2014.09.002

proposed to incorporate rich tagging information into traditional recommendation models and achieved good performance (Zhang, Zhou, & Zhang, 2011). Furthermore, a user's online activity (searching and browsing) and his/her social connections have also been explored and utilized to improve recommendation accuracy (Li, Zhai, & Chen, 2012; Wu, Chen, Yu, Han, & Wu, 2013).

Recently, affective texts (such as reviews and comments) generated by online users have been given more attention and fruitful sentiment analysis works has begun to emerge rapidly (Balahur, Hermida, & Montoyo, 2011; Feldman, 2013). Some attempts have been made to integrate sentiment analysis results into recommendation generations. The majority of existing work combines sentiment analysis techniques with collaborative filtering techniques to conduct movie rating predication (Jakob, Weber, Muller, & Gurevych, 2009; Leung, Chan, Chung, & Ngai, 2011; Peleja, Dias, Martins, & Magalhães, 2013). However, in relation to the social media context, item recommendation is limited due to a lack of sentiment-awareness. To the best of the author's knowledge, there are only two studies relevant to sentiment-aware item recommendation in social media websites: (Zhang, Ding, Chen, Li, & Zhang, 2013) and (Pappas & Popescu-Belis, 2013b). These studies only applied sentiment analysis results into the basic neighborhood-based collaborative filtering models. Advanced one-class collaborative filtering (OCCF) models were not explored. Therefore, the combination of sentiment analysis with a recommender system is limited and more extensive experimental work is required.

In this paper, we develop a sentiment-aware social media recommendation framework, referred to as SA_OCCF, to tackle the key challenges highlighted above. Firstly, the affective text from user comments is explored and mined by our proposed ensemble learning-based sentiment classification (ELSC) method. Secondly, the derived sentiment feedback information by ELSC is incorporated into the OCCF models and we formalize our sentiment-aware recommendation models (SA_OCCF) to improve social media item recommendation performance. The proposed methods and models are evaluated through comprehensive experiments using the TED dataset. The results show that the proposed SA_OCCF models outperform the baseline methods using a variety of recommendation accuracy metrics.

The main contributions of this paper can be summarized as:

(1) We propose a sentiment-aware social media recommendation framework, referred to as SA_OCCF, to recommend social media items for online users. This type of framework has rarely been addressed in state-of-the-art social media recommendation research.
(2) An ensemble learning-based sentiment classification combining BOW features and lexicon-based features is proposed to mine affective texts from comments. We extend the applications of sentiment analysis studies by incorporating their results into the social media recommendation domain.
(3) To evaluate the performance of the proposed ensemble learning-based sentiment classification method and the SA_OCCF framework, we conduct comprehensive experiments and present results showing the effectiveness of the proposed method and framework.

The rest of the paper is organized as follows. In Section 2, we survey related work on sentiment classification and recommender systems. The details of the sentiment-aware recommendation framework are introduced in Section 3, while Section 4 presents the design and methodology used in the experiments. The results are analyzed in Section 5. We conclude the work in Section 6 and discuss future research applications.

2. Literature review

Our work is related to two main research threads: (1) sentiment classification and (2) recommender systems. We review the pertinent existing literature in these two fields.

2.1. Sentiment classification

Sentiment classification has been studied since the late 1990s and has swiftly become an active research topic. It has progressed extensively since its early onset, and has been successfully applied in the areas of data mining, information retrieval and natural language processing (Boiy & Moens, 2009; Feldman, 2013; Pang & Lee, 2008). A number of researchers have investigated sentiment classification from various perspectives. In general, sentiment analysis is performed at different levels of text units which included word/phrase level, aspect level, sentence level and document level (Abbasi, Chen, & Salem, 2008; Feldman, 2013). The majority of work has been conducted to judge sentiment polarity at the document level (Abbasi, Chen, Salem, 2008; Boiy & Moens, 2009; Dave, Lawrence, & Pennock, 2003; Pang, Lee, & Vaithyanathan, 2002; Wang, Sun, Ma, Xu, & Gu, 2014; Xia, Zong, & Li, 2011). In our study, we also perform sentiment classification of online comments at the document level. The techniques used in sentiment classification are mainly classified into heuristic-based and machine learning methods (Wang et al., 2014). Heuristic-based methods generally employ predefined lexicons and calculation rules to classify text sentiments based on the total number of derived positive or negative sentiment features (Pang & Lee, 2008). For example, Turney (2002) determined the semantic orientation of a phrase using its point-wise mutual information with predefined sentiment words, such as "excellent" and "poor". The overall sentiment information of phrases was then aggregated and the sentiment classification task was achieved. On the other hand, machine learning approaches have received significant

Table 1
Selected previous studies in recommender systems combined with sentiment analysis.

Study	Task	Affective text	SA techniques	Combined features	RS techniques	Domain
Aciar et al. (2007)	Item recommendation	Product reviews	Heuristic-based (Ontology)	Product quality + opinion quality	Scoring	E-commerce
Jakob et al. (2009)	Rating prediction	Movie reviews	Heuristic-based (lexicon)	Explicit ratings + sentiment features	MRMF	Movie
Moshfeghi, Piwowarski, and Jose (2011)	Rating Prediction	Movie reviews	Emotion classifier	Semantic feature + Emotion feature	LDA	Movie
Leung et al. (2011)	Rating prediction	Movie reviews	Heuristic-based (initialization)	Opinion words	Probabilistic model	Movie
Zhu et al. (2012)	Rating prediction	Movie reviews	Heuristic-based (lexicon)	Explicit ratings + inferred sentiment rating	UCF, ICF, OCF	Movie
Yuanhong, Yang, and Xiaohui (2012)	Rating prediction	Movie reviews	Heuristic-based (lexicon)	Explicit ratings + inferred aspect rating	Tensor factorization	Movie
Chen and Wang (2013)	Item recommendation	Product reviews	Heuristic-based (lexicon)	Explicit ratings + feature-opinion pairs	Clustering-based	E-commerce
Liu, He, Wang, Song, and Du (2013)	Item recommendation	Online reviews	Heuristic-based (lexicon)	Explicit ratings + implicit opinions	Scoring	Restaurant
Zhang et al. (2013)	Item recommendation	Video comments	Semi-supervised classification	Inferred virtual rating	UCF, ICF	Social media website
Pappas and Popescu-Belis (2013b)	Item recommendation	Talk comments	Rule-based classifier	Implicit feedback + inferred sentiment feedback	SA_ICF	Social media website
Peleja et al. (2013)	Rating prediction	Movie reviews	VM classifier (SentiWordNet)	Explicit ratings + inferred unrated rating	MF (SVD)	Social TV (Movie)
Our proposed approach	Item recommendation	Talk comments	Ensemble learning (SentiWordNet)	Implicit feedback + inferred sentiment feedback	SA_OCCF	Social media website

attention to sentiment classification due to their predominant classification performance (Abbasi, Chen, Thoms, & Fu, 2008; Pang & Lee, 2008). Through the construction of predictive models from labeled training datasets, these methods can model more features and adapt to changing inputs more robustly, given them an advantage in comparison to heuristic-based methods. In a similar vein, a number of ensemble learning-based methods have been proposed to construct multiple classifiers and then yield the integrated classifier with overall performance. Such ensemble based methods have outperformed single machine learning techniques for sentiment classification (Abbasi, Chen, Thoms, et al., 2008; Wang et al., 2014; Wilson, Wiebe, & Hwa, 2006; Xia et al., 2011).

Although most sentiment analysis addresses tasks in commercial settings, such as extracting opinions from product reviews, there is increasing interest in analysis of affective text from social media websites (Thelwall, Buckley, & Paltoglou, 2012). The "SentiStrength" tool was proposed in order to analyze the level of sentiments in short informal text and it was proven useful to classify emotions in Myspace (Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010) and Twitter (Thelwall, Buckley, & Paltoglou, 2011). The SentiWordNet thesaurus was employed to detect sentiment for YouTube comments and predict the usefulness of those comments (Siersdorfer, Chelaru, Nejdl, & Pedro, 2010). In (Qiu, Zhang, Hu, & Zhao, 2009), the SELC model (self-supervised, lexicon-based, and corpus-based) was proposed for sentiment classification in Youku,[4] a Chinese video sharing website. However, ensemble-based approaches for sentiment classification within video sharing contexts were not investigated. In this study, we will leverage the ensemble learning-based methods for sentiment classification of online video comments and generate sentiment feedbacks to improve social media recommendation performance.

2.2. Recommender systems

Personalization techniques provide the ability to tailor content and services to individuals based on their preferences and tastes. A recommender system, as the primary tool of personalization, is required to match potential interesting content with a user's expectations (Adomavicius & Tuzhilin, 2005). The preference ratings are assumed to be available and many techniques have been developed in recommender systems in order to derive improved performance (Liu, Guo, & Zhang, 2011; Takacs, Pilaszy, Nemeth, & Tikk, 2009). However, most recommenders cannot perform effectively when the user preference data (i.e., ratings) are limited, which is often the case in the real world (e.g., social media sites). To address this limitation, rich user generated content has been mined and utilized as a supplementary source to support personalized recommendation. Tag-aware recommender systems have been attracted increasing attention since fruitful tagging information can be leveraged and modeled into traditional collaborative filtering models (Zhang et al., 2011). In (Tan et al., 2011), they proposed a novel music recommendation algorithm by using multiple social media data (e.g., groups, tags, etc.) and music acoustic-based content in the framework of a hypergraph model. Recently, affective text (such as reviews and comments) generated by online users has been considered, with some researchers integrating sentiment analysis results into the generation of

[4] www.youku.com.

recommendations. However, these studies are limited in scope and a more comprehensive understanding of the impacts of the sentiment information on recommendations is required. Table 1 presents selected previous studies in recommender systems combined with sentiment analysis.

The majority of existing research has focused on movie rating prediction tasks by mining sentiment information from review text. In (Jakob et al., 2009), the sentiment review information is considered as additional features and then incorporated into a Multi-Relational Matrix Factorization (MRMF) algorithm. The results also demonstrated the advantages of the new recommendation method using sentiment features. Leung et al. (2011) inferred ratings from reviews and integrated them with a probabilistic model for movie recommendation. In (Zhu, Xing, & Liang, 2012), the sentiment rating matrix was inferred through review opinion analysis and an aggregated model combining user-based collaborative filtering (UCF), item-based collaborative filtering (ICF) and opinion-based collaborative filtering (OCF) was proposed to improve the predication accuracy. Heuristic-based methods were employed by all of these researchers to extract sentiment information from the review text. Recently, Peleja et al. (2013) used the Vowpal Wabbit (VM) linear classifier to extract sentiment ratings from reviews and combined explicit ratings with inferred unrated rating to predict users' preferences. They conducted several experiments to verify the effectiveness of the sentiment analysis method and their proposed recommendation method.

In the case of E-commerce websites, many commercial systems often show 'best bet' recommendations rather than the predicted rating values. Currently, recommender system researchers are giving attention to the task of item recommendation, which is also called top-*N* recommendation, where the goal of the recommender system is to find a few specific items which are the most appealing to users. Chen and Wang (2013) analyzed product reviews to extract feature-opinion pairs and then employed a novel clustering-based method to identify reviewer preference homogeneity. Finally, a recommendation method was also proposed based on inter-relevance preference between active buyers and reviewers. The performance was also demonstrated and seen to be excellent. In a different study, Aciar, Zhang, Simoff, and Debenham (2007) analyzed a user's reviews by developing an ontology to mine the review text. The ontology measures the quality of several features within a product to create user recommendations.

In the case of social media websites, it is typical that only positive implicit feedback is available. For a user preference matrix, the portion of unlabeled examples is large and the matrix becomes extremely sparse. Due to this limitation, only a few papers have been published using sentiment analysis results from reviews in the generation of recommendations. In (Zhang et al., 2013), the sentiment information from reviews and a user's facial expressions were extracted and the virtual rating matrix was generated for sentiment-based recommendation algorithms. In their systems, only traditional UCF and ICF methods were employed. In our work, we will place a heavier reliance on OCCF models. The method proposed by (Pappas & Popescu-Belis, 2013b) is more relevant to our work. They claimed that they employed sentiment analysis of user comments for OCCF recommendations, yet only a simple modified item-based CF was proposed and some recent advanced OCCF models were not explored. Furthermore, we propose an ensemble learning-based method for sentiment classification which also differs from Pappas & Popescu-Belis's work.

3. Sentiment-aware social media item recommendation

Computational applications are pushing the boundaries of personal computing and beginning to facilitate social interactions (Tavakolifard & Almeroth, 2012). Social media websites are now prevalent and researchers are engaging in the research of information technology concerned with the intersection of human and social studies connected through computer

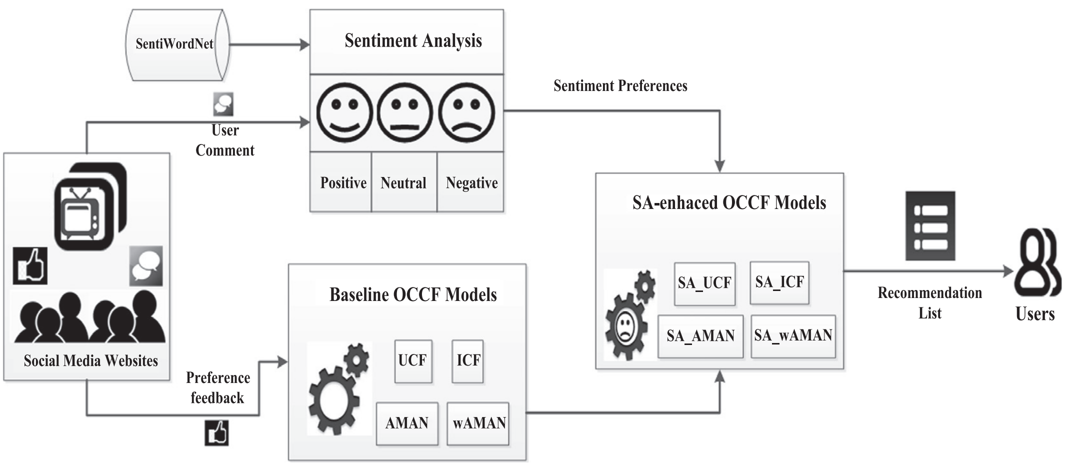

Fig. 1. Overview of the proposed sentiment-aware recommender system.

networks. Online users in these websites have generated a large volume of content which contains fruitful opinion data. In this study, we propose a sentiment-aware recommender system which employs sentiment analysis techniques to mine affective texts and leverage OCCF models to provide personalized media item recommendations. In Fig. 1 the recommendation process is shown. The framework of the proposed recommendation method contains two primary modules: sentiment analysis for comments and OCCF recommendation models. The sentiment analysis component is employed to mine emotional comments and generate sentiment feedbacks to improve the recommendation performance. The OCCF models component will first user the preference feedbacks models and baseline OCCF models to conduct recommendations and then leverage sentiment feedbacks and SA-enhanced OCCF models before providing their recommendations. These two components are covered in detail in the following sections.

3.1. Sentiment analysis for comments

The goal of sentiment analysis for comments is to mine the affective texts and infer a user's sentiment feedback for social media items. This task can be formally defined as follows: given a set of comments, a sentiment classifier model classifies each comment into one of two classes, positive or negative. In this study, we conduct sentiment classification using an ensemble learning-based method, which achieved best performance among many baseline methods in our previous study (Wang et al., 2014). The ensemble learning-based sentiment classification (ELSC) method consists of three steps: (1) extracting features and converting them into feature vectors, (2) leaning the appropriate classifier model using the vectors, and (3) selecting the proper ensemble strategy to enhance the classifier. Accordingly, each step is introduced in the following subsections.

3.1.1. Feature construction

The construction of features is an important part of machine learning methods for sentiment classification. From the machine learning perspective, it is useful to set a rule that restricts features to include only relevant information while also being independent of each other (Forman, 2003). To extract effective features, the comment text is first segmented using natural language processing (NLP) tools before various annotation methods are used. The Bag-of-Words (BOW) framework is the primary annotation method used in sentiment classification literatures (Pang & Lee, 2008). In this framework, the text is considered as a bag of words and represented by a vector containing all the words appearing in the corpus. Furthermore, the Term Frequency-Inverse Document Frequency (TF-IDF) weighting method is used to represent words in this framework while unigram features are often generated to learn the classifiers. In comparison to BOW features, lexicon-based features have been proposed for sentiment analysis. SentiWordNet is a popular lexical resource used for opinion mining (Esuli & Sebastiani, 2006). Ohana and Tierney (2009) classified the sentiment of reviews using five types of lexicon-based features which included overall document scores, score ratio to total terms, positive to negative score ratios, scores per document segment, and negation. We use BOW features and lexicon-based features for sentiment classification of comments, due to their unique properties.

3.1.2. Model learning

After the feature vectors (unigram features combined with lexicon-based features) are constructed, they should be normalized and fed into the classifiers for learning. Sentiment classification tasks have been investigated in a number of machine learning methods: Naïve Bayes classifier (NB), Decision Tree classifier (DT), mutual entropy classifier (ME), support vector machines (SVM), etc. We choose all of these classifiers as the base learners in this study. Among them, SVM is a state-of-the-art data mining technique which has proven its performance in many applications (Vapnik, 2000). It has a sound theoretical foundation (statistical learning theory and structural risk minimization principle) and also has excellent learning performance and generalization capabilities. The strength of this technique lies with its ability to model non-linearity, resulting in complex mathematical models. In comparison to ANN, SVM can capture the inherent characteristics of the data more concisely.

3.1.3. Ensemble strategy selection

Prior researchers have shown that ensemble based methods outperformed single machine learning technique for sentiment classification (Abbasi, Chen, Salem, 2008; Li, Wang, & Chen, 2012; Xia et al., 2011). Ensemble learning is a machine learning paradigm where multiple learners are trained to solve the same problem (Polikar, 2006; Zhou, 2012). In contrast to traditional machine learning approaches that attempt to learn one hypothesis from the training data, ensemble methods attempt to construct a set of hypotheses and combine them (Liu & Zsu, 2009). Taking into account our previous sentiment classification comparative study (Wang et al., 2014), we choose the random subspace method to leverage both types of features for sentiment classification. The Random Subspace method is an ensemble construction technique proposed by Ho (1998). In the random subspace, the training dataset is modified in the feature space rather than the instance space. The pseudo-code for the random subspace algorithm is given in Table 2.

The random subspace method may benefit from using both random subspaces for construction and aggregation of the base classifiers. When the dataset has many redundant or irrelevant features, one may obtain a better base classifier in the random subspaces than in the original feature space (Ho, 1998). The combined decision of such base classifiers may

Table 2
The random subspace algorithm.

Input: Data set $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$;
Base classifier algorithm, L;
Number of random subspace rate, k;
Number of learning rounds, T.
Process: For $t = 1, 2, \ldots, T$;
$D_t = RS(D, k)$; % Randomly generate a subspace sample from D
$h_t = L(D_t)$; % Train a base classifier h_t from the subspace sample
End.
Output: $H(x) = Argmax_{y \in Y} \sum_{t=1}^{T} \tau(y = h_t(x))$;% the value of $\tau(\alpha)$ is 1 if α is true and 0 otherwise

Table 3
Key notations for social media recommendation.

Notation	Meaning
U	A set of users
I	A set of items
u_i	Index of user, $u_i \in \{u_1, i_2, \ldots u_{\lvert U \rvert}\}$
i_j	Index of item, $i_j \in \{i_1, i_2, \ldots i_{\lvert I \rvert}\}$
$(R_{ij})_{\lvert U \rvert \times \lvert I \rvert}$	User-item preference matrix, $R_{ij} \in \{0, 1\}$
$(S_{ij})_{\lvert U \rvert \times \lvert I \rvert}$	User-item comment sentiment matrix $S_{ij} \in \{-1, 0, 1\}$
L	Top k recommended items, $L = (L_1, L_2 \ldots L_k)$

be superior to a single classifier constructed on the original training dataset in the complete feature sets. In our experiment, we also choose the AdaBoost ensemble-based method as the baseline.

3.2. OCCF recommendation models

In this section, we define the problem of item recommendation and then proceed to introduce two representative models, which we will extend by incorporating sentiment information. Recent studies (Li, Zhai, et al., 2012; Pappas & Popescu-Belis, 2013b; Rong et al., 2008) have been concerned with Top N Item recommendation, and we will also consider this task in the framework of one class collaborative filtering. In this framework, OCCF is intended to recommend items of interest to users based on their available favorite items given that there are only positive implicit feedback examples (i.e. purchases, bookmarks, views) in the available datasets. In this paper, we are interested in the recommendation of social media items to users by leveraging affective text from comments. For the sake of illustration, we list the key notations in Table 3. For a given user, u, the task of item recommendation is to provide a sorted list of top N items, L, based on the available user-item preference matrix **R** and user-item comment sentiment matrix **S**.

3.2.1. Baseline OCCF models

In memory-based CF methods, the recommendation score is computed based on the entire (or a sample of) the user-item preference matrix (Su & Khoshgoftaar, 2009). The similarity computation between users or items is the primary task of these algorithms. Neighborhood-based CF is one of memory-based CF models for recommender systems. The method typically includes the following steps: calculation of the similarities between two users or two items, finding the k most similar users or items and aggregating the neighbors' ratings to generate top N most frequent items as suitable recommendations. Neighborhood-based CF models can also be classified into user-based CF (UCF) models and item-based CF (ICF) models. A UCF model provides recommendations by determining the most similar neighbors for a given user and then aggregating neighbors' rating. This model can be formally defined as:

$$P(u, i) = \overline{r_u} + \frac{\sum_{v \in Nr_u} sw(u, v)(r_{vi} - \overline{r_v})}{\sum_{v \in Nr_u} sw(u, v)} \tag{1}$$

where $P(u, i)$ is the predicted rating of the given user u to an unrated item i, r_{vi} is the rating of user v for item i, $\overline{r_u}$ and $\overline{r_v}$ are the mean ratings of u and v respectively, $Nr_u \in U$ is the set of nearest neighbors of the given user u and $sw(u, v)$ is the similarity weight between user u and user v. The similarity weight $sw(u, v)$ can be calculated using a function such as cosine similarity (Eq. (2)) based on the preference matrix R.

$$sw(u, v)_{\cos} = \frac{\sum_{i=1}^{\lvert I \rvert} r_{ui} * r_{vi}}{\sqrt{\sum_{i=1}^{\lvert I \rvert} r_{ui}^2} * \sqrt{\sum_{i=1}^{\lvert I \rvert} r_{vi}^2}} \tag{2}$$

Furthermore, the ICF model provides recommendations by finding similar items to the given user's rated items and then calculating the weighted combination of their ratings. The model can be formally defined as:

$$P(u,i) = \overline{r_i} + \frac{\sum_{j\in Nr_i} sw(i,j)(r_{uj} - \overline{r_j})}{\sum_{j\in Nr_i} sw(i,j)} \tag{3}$$

where $\overline{r_i}$ is the mean rating of item $sw(u, v)$ i, Nr_i is the set of nearest item neighbors from the given user's rated items, and $sw(i,j)$ is the similarity weight between item i and item j. The computation of $sw(i,j)$ is similar to, which is shown in Eq. (2). Therefore, we have chosen the UCF and ICF models as our baseline models in the memory-based methods. We will introduce the SA-enhanced memory-based models by incorporating a user's sentiment comment information in Section 3.2.2.

Model-based CF methods attempt to learn complex patterns based on the training data (for example, preference matrix) and provide recommendations based on the learned models (Su & Khoshgoftaar, 2009). Matrix factorization (MF) techniques are typical model-based CF methods and have been successfully applied in movie recommendation context (Bennett & Lanning, 2007). Many studies have empirically suggested that MF methods perform better than neighborhood-based methods in rating predication tasks where the main goal is to reduce a root-mean-square-error (RMSE) (Adomavicius & Tuzhilin, 2005; Bell, Koren, & Volinsky, 2007; Bennett & Lanning, 2007). The idea behind a MF based method is to predict the rating $\breve{r}_{ui}$ by the learned latent user factors $x_{u\in} R^f$ and latent item factors $y_{i\in} R^f$. In order to learn these latent factors, recent studies have suggested direct modeling on the observed ratings and use of the Tikhonov regularization to avoid over-fitting issues. Therefore, the learning process is equivalent to solving the unconstrained optimization problem:

$$\text{Argmin}_{x^*,y^*} \sum_{u,i} (\|r_{ui} - x_u^T y_i\|^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)) \tag{4}$$

where $\|\cdot\|$ is the Euclidean norm of a matrix and λ is used to regularize the model. This model has already demonstrated its prediction excellence for the Netflix datasets (Bennett & Lanning, 2007). Therefore, we will use it as a baseline model. In the OCCF setting, the above model treat all missing values in R as negative examples (AMAN) where the response variables only take the value 0 or 1. Hereafter, we call this model AMAN_MF. However, this strategy is obviously limited since it does not cover the real world case that missing examples are likely to be positive ones. A better approach proposed by Rong et al. (2008) treat all missing values as weighted negative examples (wAMAN). The weights are used to control their relative contribution to the loss function:

$$L(x_u, y_i) = \sum_{u,i} w_{ui}(\|r_{ui} - x_u^T y_i\|^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)) \tag{5}$$

Minimization of the function $L(x_u, y_i)$ allows the parameters (x_u and y_i) to be learnt using Alternating Least Squares (ALS) as suggested in (Rong et al., 2008; Yifan, Koren, & Volinsky, 2008). The ALS algorithm is an effective and efficient iterative method to solve the optimization problem. In the case of the wMAN model, a weighted ALS (wALS) was proposed in (Rong et al., 2008) and used to solve the above low rank approximation problem mentioned above. The pseudo code of the wALS algorithm is given in Table 4.

In this paper, we decide to use the AMAN_MF and wAMAN_MF methods as our baselines in the model-based methods. In the case of the wAMAN_MF method, we use the global weighting method to assign the weights. In Section 3.2.2, we will incorporate user sentiment information into the wAMAN method to enhance the basic model.

3.2.2. SA-enhanced OCCF models

Social media websites provide online users with efficient and convenient methods of consuming media items while simultaneously recording large volumes of user generated content during their interactions. Users express their opinions and emotions of the consumed media using affective text in comments. Such affective information, as another kind of implicit feedback, can be exploited and combined with consumption feedback information to improve recommendation performance. In this paper, we leverage affective texts to utilize sentiment information in order to improve the OCCF

Table 4
The wALS algorithm.

```
Input: Preference matrix, R;
       Weight matrix, W;
       Number of factors, f;
       Number of iterations, Iter;
Process:
     Define the latent factor matrices X and Y with f factors;
     Initialize X and Y with
     Gaussian Randomness (zero mean and small standard deviation);
       Do
         Update x_u ∈ X, using Equation x_u = (Y^T W_u Y + λI)^{-1} Y^T W_u r_u;
         Update y_i ∈ Y, using Equation y_i = (X^T W_i X + λI)^{-1} X^T W_i r_i;
       While Iter > ε
     Return X, Y
Output: X and Y
```

recommendation performance. We incorporate sentiment information into the OCCF framework and generate two corresponding enhanced models: SA-enhanced neighborhood and SA-enhanced MF.

In the case of the neighborhood framework, the preference matrix R can be merged with the sentiment matrix S to learn the neighborhood-based CF models. The sentiment values of the comments can be considered as inferred preference ratings and are incorporated into the preference matrix R to generate a denser matrix R'. Using the matrix R', previous neighborhood-based CF models must be modified: firstly, inferred ratings from the sentiment analysis performed over the comments must be considered in the model; secondly, the neighbors must be re-calculated with the additional inferred rating data. Therefore, the previous user-based CF method is modified as:

$$P(u,i) = \overline{r'_u} + \frac{\sum_{v \in Nr'_u} sw'(u,v)(r'_{vi} - \overline{r'_v})}{\sum_{v \in Nr'_u} sw'(u,v)} \tag{6}$$

where Nr'_u denotes the user u's nearest neighbors calculated by the denser matrix R', $sw'(u,v)$ is the similarity weight between u and v based on the matrix R' and r'_{vi} accounts for both feedback and inferred ratings from the comments. r'_{vi} can be defined as:

$$r'_{vi} = \begin{cases} 1, & \text{if } r_{vi} = 1 \\ s_{vi}, & \text{if } r_{vi} = 0 \end{cases}, \tag{7}$$

Similarly, we can introduce the modified item-based CF method which is defined as:

$$P(u,i) = \overline{r'_i} + \frac{\sum_{j \in Nr'_i} sw'(i,j)(r'_{uj} - \overline{r'_j})}{\sum_{j \in Nr'_i} sw'(i,j)}, \tag{8}$$

where Nr'_i is the set of nearest item neighbors from the given user's rated items or commented items based on R', and $sw'(u,v)$ is the similarity weight between item i and item j based on the matrix R'. Therefore, we regard these methods as SA_NN models.

For the MF framework, AMAN_MF and wAMAN_MF are two state-of-the-art baseline methods for item recommendation. We leverage sentiment information in order to enhance the AMAN_MF method using the generated denser matrix R'. Therefore, the optimization problem in Eq. (4) should be learned and solved on the matrix R' (as defined in Eq. (9)) as:

$$\text{Argmin}_{x^*,y^*} \sum_{u,i} (\|r'_{ui} - x_u^T y_i\|^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)) \tag{9}$$

To enhance the wAMAN_MF method, we also consider missing examples as negative examples. Instead of using a global weighting scheme, we leverage sentiment information extracted from user comments to assign the weight for each negative example. We employ a new weight matrix W' to substitute the loss function of Eq. (5) and generate the following equation:

$$L_s(x_u, y_i) = \sum_{u,i} w'_{ui}(\|r_{ui} - x_u^T y_i\|^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)) \tag{10}$$

where w'_{ui} can be defined as:

$$w'_{ui} = \begin{cases} 1 + s_{ui}, & \text{if } r_{ui} = 1 \\ 1, & \text{if } r_{ui} = 0 \wedge S_{ui} = -1 \\ 0, & \text{else} \end{cases} \tag{11}$$

Similarly, the loss function $L_s(x_u, y_i)$ can also be optimized by the wALS algorithm in Table 4. Therefore, we regard these methods as SA_MF methods.

4. Experimental design

4.1. Datasets

In order to evaluate the proposed sentiment-aware recommendation framework, we investigated the TED talk dataset. TED stands for Technology, Entertainment, and Design, which is a non-profit media organization that develops and curates global conferences, and represents an alternative to traditional channels for communicating ideas (Heffernan, 2009). TED.com[5] broadcasts TED Talks as video streams on their website. The talks are given in English and are usually transcribed and then translated into several other languages by volunteers. TED has become one of the most popular online lecture repositories due to its high quality talks, and was named as "The Conference of Cool" in the Financial Times by Peter Aspden (2010), which reflected growing acceptance and prestige of TED as an alternative medium. As reported in (Banker & Gournelos, 2013), TED talks were viewed a total of one billion times in November, 2012. In recent studies, TED was used to educate teachers (DaVia Rubenstein, 2012) and involve students in lifelong learning (Banker & Gournelos, 2013).

[5] http://www.ted.com.

The TED website also supplies an online repository of audiovisual recordings of popular scientific lectures. Under a Creative Commons noncommercial license, the recordings and the metadata accompanying speakers are made available (Pappas & Popescu-Belis, 2013a). The website provides extended metadata as well as user-contributed material such as comments relating to the talks. In (Sugimoto & Thelwall, 2013), a range of bibliometric (citation) and webometric (views and comments) indicators were used to examine TED videos in order to provide insights into the type and scope of their impact. It was suggested that TED's primary dissemination channel (website) is interactive through its comments, making it possible to identify engagement with the videos and the extent to which they generate discussion. In (Pappas & Popescu-Belis, 2013a), TED content metadata (such as title, description and tags) were combined with user preferences to provide personalized lecture recommendations. Since talk comments constituted a more implicit form of preference expression, they were exploited to augment the rating information for recommendation in (Pappas & Popescu-Belis, 2013b). In our study, we propose a sentiment classification method to extract sentiment information from user comments and leverage them in the OCCF recommendation settings.

In our work, we have used the second version of the published TED dataset which was crawled on the TED website on September 10th, 2012.[6] The original dataset contained 12,403 users, 1203 talks, 134,533 favorites and 117,516 comments. In order to validate the sentiment analysis component, we prepare the labeling dataset which contain positive and negative comments. Since we did not obtain the sentiment labels of the TED comments, we employed a similar strategy as was used in (Pappas & Popescu-Belis, 2013b) to label the subset of TED comments by human annotations. In the task of polarity annotation, we distinguish three functional evaluative components according to Wiebe's work (2004): the source (i.e. the person or entity that expresses or experiences the private state), the evaluation (i.e. words or phrases inherently bearing a positive or negative value), the evaluated target (i.e. the talk video in this dataset). We restricted our analysis to evaluative opinions/states only. Three native speakers (two master students and one PhD student) were selected as annotators to judge the overall sentiment of 1000 comments, which were randomly selected from the TED dataset. We employed Fleiss' kappa (k) metric to measure the inter-annotator agreement and the metric was found to be $k = 0.69$. As agreement was substantial, we used the entire set as ground truth to evaluate sentiment classification methods (cases of disagreements were reconciled by a majority vote). As a result, we obtained 584 positive comments, 325 negative comments and 91 neutral comments (including undecided comments).

In order to train and validate our proposed sentiment classification method, 300 positive comments and 300 negative comments were randomly selected to form the ground truth dataset. Finally, the learned sentiment classifier was employed across entire comment dataset and the sentiment feedback matrix was generated for item recommendation.

We chose users with at least 5 favorites in order to do 5-fold cross validation for item recommendation. The user comments were analyzed by the previous sentiment classification method to generate the sentiment feedback matrix. The resulting data set had 5657 users, 1203 talks, 121,621 favorites and 21,076 comments.

4.2. Evaluation metrics

For sentiment analysis, the established standard measure, the average accuracy, is adopted to evaluate the performance of the proposed method. The definition of the average accuracy can be explained with a confusion matrix as shown in Table 5.

Formally, the average accuracy is defined as:

$$\text{Average accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{12}$$

In the case of item recommendation, we treat it as a content retrieval system that recommends social media items to online users. The evaluation metrics, Precision@K (Prec@K), Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR) (Croft, Metzler, & Strohman, 2010) are employed to evaluate the recommendation accuracy of different methods. The Prec@K measure only evaluates the ability to return overall relevant items. However, the MAP and MRR measures consider the rank information of relevant items in the recommendation list. They are defined as

$$P@K = \frac{N_{\text{relevant}}}{K}, \tag{13}$$

$$\text{MAP} = \frac{1}{|U|}\sum_{j=1}^{|U|}\frac{1}{m_j}\sum_{k=1}^{N}P(R_{jk}), \tag{14}$$

$$\text{MRR} = \frac{1}{|U|}\sum_{j=1}^{|U|}\frac{1}{\text{rank}F_j}, \tag{15}$$

where K is the number of recommended items and (5 and 10 in this setting); N_{relevant} is the number of relevant items in the ranking list, $|U|$ denotes the number of users, m_j is the number of relevant items to the user j, $P(R_{jk})$ represent the precision of

[6] www.idiap.ch/dataset/ted.

Table 5
Confusion matrix for sentiment classification.

Predicted	Actual	
	Positive sentiment	Negative sentiment
Positive sentiment	True Positive (TP)	False Positive (FP)
Negative sentiment	False Negative (FN)	True Negative (TN)

recommended results from the top result until reaching item k, and Rank F_j is the rank (position) of the first relevant item to user j.

4.3. Experimental procedure

To minimize the influence of variations in the training set, n-fold cross validation is performed. For sentiment analysis, the labeled comments are partitioned into ten subsets with similar sizes and distributions. The union of nine subsets is then used as the training set while the remaining subset is used as the test set, which is repeated ten times such that every subset has been used as the test set once. The average test result is regarded as the result of the 10-fold cross validation. The above process is repeated 10 times with random partitions of the ten subsets, and the average result of these different partitions is recorded. We measured seven approaches in the comparison: SentiWordNet,[7] NB, DT, ME, SVM, SVM-Boosting, and our proposed ELSC method.

In the item recommendation, we implement our SA-enhanced method and the baseline OCCF models in the literature for the performance comparison. They are listed as follows.

(1) UCF: This method employs user similarities calculated from the favorite feedback matrix to provide recommendations. Refer to Eq. (1).
(2) ICF: This method employs item similarities calculated from the favorite feedback matrix to provide recommendations. Refer to Eq. (4).
(3) AWAN_MF: This method employs the matrix factorization technique for the favorite feedback matrix and uses the AWAN strategy to provide recommendations. Refer to Eq. (5).
(4) wAWAN_MF: This method employs the weighted matrix factorization technique for the favorite feedback matrix and uses the wAWAN strategy to provide recommendations. Refer to Eq. (6).
(5) SA_UCF: This method employs user similarities calculated from the combined feedback matrix to provide recommendations. Refer to Eq. (7).
(6) SA_ICF: This method employs item similarities calculated from the combined feedback matrix to provide recommendations. Refer to Eq. (8).
(7) SA_AWAN_MF: This method employs the matrix factorization technique for the combined feedback matrix and uses the AWAN strategy to provide recommendations. Refer to Eq. (9).
(8) SA_wAWAN_MF: This method employs the proposed weighted matrix factorization technique for the favorite feedback matrix and uses the wAWAN strategy to provide recommendations. Refer to Eq. (10).

Of these eight methods, the first four represent baseline OCCF methods while the remainder represents SA-enhanced OCCF methods. Furthermore, the first two of the baseline OCCF methods are neighborhood-based CF methods while the next two of are MF-based CF methods.

5. Results and discussions

We used the data mining toolkit WEKA (Waikato Environment for Knowledge Analysis) version 3.7.0 (with default parameter settings) for learning-based sentiment classification (Hall et al., 2009). SentiWordNet 3.0 was employed as the baseline sentiment classification method (Baccianella, Esuli, & Sebastiani, 2010). The item recommendation techniques were implemented and the results were compared in the MyMediaLite recommender system library (Gantner, Rendle, Freudenthaler, & Schmidt-Thieme, 2011).

We determined the sensitivity of different algorithms before beginning the full experimental evaluation, and used the sensitivity plots to fix the optimum values of various parameters. These values remained fixed throughout the experiments. To determine the parameter sensitivity, we used only the training data and further subdivided it into training and validation portions to perform our experiments. We argue that our experimental design is fair because: (1) we do not user any privileged information from the test set in which results of all methods are reported; (2) all parameters are discovered in the same validation set. As a result, the neighborhood sizes of 40 and 70 were set for the UCF and ICF models, respectively, while the neighborhood sizes of 30 and 50 were set for the SA_UCF and SA_ICF models, respectively. We set the number of factors

[7] http://sentiwordnet.isti.cnr.it/.

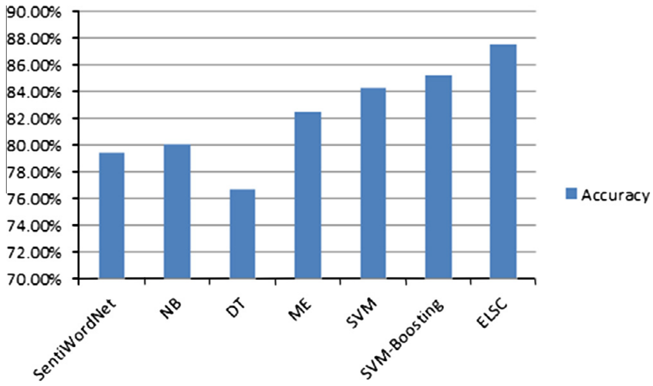

Fig. 2. Sentiment classification results of the seven methods.

Table 6
Sample size of the TED dataset.

Set	Users	Favorites	Comments
Training set	5657	121,621	21,076
Testing set	2099	69,164	6372

as 10 and total iterations as 30 for AWAN_MF and SA_AWAN_MF. The factors value, iteration value, and negative weight value were set as 20, 15, and 2 respectively for wAWAN_MF. The number of factors and iterations were learned to be 15 and 20 for SA_wAWAN_MF. Thereafter, the performance of the three sentiment classification methods and eight item recommendation methods is reported in the following sub-sections.

5.1. Sentiment classification accuracy

The accuracy results of sentiment classification performed on talk comments are shown in Fig. 2. It can be observed that the machine learning methods have superior performance to the Heuristic-based method, SentiWordNet, with the exception of DT. This may be due to the availability of the training data in addition to the BOW features and lexicon features are leveraged by the learning-based methods. The SVM method achieves the best performance among all machine leaning sentiment classifiers. An ensemble-based method, SVM-Boosting, further improve the classification accuracy over SVM. This is consistent with our previous work investigating the performance contributions of ensemble methods (Wang et al., 2014). The best performance is achieved by our proposed ELSC method with a classification accuracy of 86.7%. This method can be employed to effectively extract sentiment information from talk comments to improve social media item recommendation performance.

Therefore, the above sentiment classification results indicate that our ELSC method is capable of overcoming the challenges associated with the special features of affective comments while providing reliable results to generate a sentiment feedback matrix in order to enhance item recommendation performance.

5.2. Item recommendation accuracy

For item recommendation, we split the user favorite matrix into training (80%) and testing (20%) sets. More precisely, for each user we retain 80% of their positive feedback (value of ‘1’) for training and hide 20% for testing. We keep all test users with at least 10 favorite talks and one comment. The details of the sample size are shown in Table 6.

In this sub-section, we present the results of the eight recommendation models utilized on the TED dataset (Table 7). We will report results on the Neighborhood-based and MF-based models respectively, and explore the performance effects of incorporating comment sentiment information.

In the case of the Neighborhood-based methods, ICF generally performs better than UCF on the favorite feedback matrix as shown in Table 7 This is consistent with previous item recommendation research where it was attributed to the existence of more users than items in the database (Bogers & Van Den Bosch, 2008). It can also be observed that sentiment feedbacks as additional information improve the results significantly, where more than 15% MAP performance improvements have been achieved by the SA-enhanced neighborhood-based models (SA_UCF and SA_ICF). In particular, further improvements are obtained by the SA_UCF model since incorporating sentiment feedback information increases the matrix density and the calculated similarities between users become more trustful.

Table 7
Performance results of the eight recommendation methods.

Methods		Prec@5	Prec@10	MAP	MRR	MAP improvements over baseline OCCF
Neighborhood-based method	UCF	0.0639	0.0490	0.0847	0.1081	–
	ICF	0.0678	0.0532	0.0900	0.1397	–
	SA_UCF	0.0837	0.0605	0.1180	0.1420	39.3%
	SA_ICF	0.0815	0.0623	0.1037	0.1531	15.2%
MF method	AWAN_MF	0.0416	0.0403	0.0738	0.0926	–
	wAWAN_MF	0.0698	0.0621	0.1012	0.1345	–
	SA_AWAN_MF	0.0673	0.0569	0.0905	0.1067	22.3%
	SA_wAWAN_MF	0.1007	0.0798	0.1275	0.1647	26.0%

Table 8
Performance comparison of the fused model and SA-enhanced OCCF models.

	Prec@5	Prec@10	MAP	MRR
SA_UCF	0.0837	0.0605	0.1180	0.1420
SA_wAWAN_MF	0.1007	0.0798	0.1275	0.1647
Fused model	0.1142	0.0932	0.1514	0.1789
Performance improvements	13.4%	16.8%	18.7%	8.6%

For MF-based methods, AWAN_MF performs worst since treating all missing examples as negative examples without weighting does not work well in the case of the highly sparse OCCF problem. Previous studies have shown that the weighted MF methods perform well for OCCF while results in this study show that the wAMAN_MF model with a global weighting scheme just slightly outperform the ICF model. This result suggests that the MF model is not guaranteed to capture the relevant information between users and items by simply assigning a uniform weight in the negative examples. In comparison the neighborhood-based models performed more robustly across datasets since the relevant information from neighborhoods is easier to explain. Furthermore, SA-enhanced MF-based models have achieved better performance in terms of MAP improvements (22.3% for SA_AWAN_MF and 26.0% for SA_wAWAN_MF). The sentiment feedback information demonstrates its effectiveness as additive evidence to the baseline model. Using sentiment feedbacks to determine the weights of the wAWAN_MF model, our proposed SA_wAWAN_MF model is able to get the best results.

Motivated by recent item recommendation research (Bogers & Van Den Bosch, 2011; Li, Zhai, et al., 2012), we propose a fused model by combining the rankings produced by the Neighborhood-based and MF-based models in order to overcome the limitations of the individual models. Data fusion has also been widely investigated in the information retrieval community. The CombMNZ aggregation method (Fox & Shaw, 1994) is one of the most popular fusion-based methods and is selected to combine the rankings produced by the Neighborhood and MF based models. The CombMNZ method uses not only the scores in each of the ranked lists in addition to number of the supporting evidence. We test this fused model on the TED dataset. The SA_UCF and SA_wAWAN_MF models are used as the two basic models for fusion since they are seen to have superior performance.

In Table 8, the fused model significantly outperforms other models in terms of all performance metrics. Specifically, the MAP improvement (18.7%) of the fused model over SA_wAWAN_MF is encouraging and promising. Therefore, these results demonstrate that combining Neighborhood-based and MF-based models is very effective to improve OCCF accuracy.

6. Conclusion and future work

In this paper, we proposed a sentiment-aware recommender system for use in the social media websites. The proposed method leveraged sentiment information from user generated affective texts to improve OCCF performance. Comprehensive experiments were conducted on a real social media website (TED dataset) in order to evaluate the effectiveness of the proposed models. After careful analysis of the results, we have arrived at the following conclusions: (1) Sentiment information is very effective to overcome the sparsity of OCCF. (2) Neighborhood-based models work robustly when the data matrix is very sparse, while MF-based models perform better when the data matrix is less sparse. (3) An ensemble model combining UCF and MF based models can further improve the prediction accuracy.

Several future research directions also exist for this study. Firstly, in this paper, only one type of social media websites (TED dataset) was employed to verify the proposed method. We will extend our method to other websites (such as Flicker) so as to improve its applicability and generality. Secondly, as a user's online comments often have imbalanced sentiment class distributions, some more advanced classification methods will be explored to address these issues. Thirdly, we will classify the comments in a finer granularity, providing additional classifications in addition to "positive" and "negative" and explore their applications on state-of-the-art recommendation algorithms. Finally, this paper adopted the CombMNZ technique as the rank aggregation method. More complex data fusion techniques can also be considered such as Condorcet fusion (Montague & Aslam, 2002) and other techniques that model score distribution (Nandakumar, Chen, Dass, & Jain, 2008).

Acknowledgments

This work was supported in part by 973 Project (Grant No. 2013CB329600), National Natural Science Foundation of China (71371062, 71302064, 71101042, 71471054), the Specialized Research Fund for the Doctoral Program of Higher Education (20110111120014), the China Postdoctoral Science Foundation (2011M501041, 2013T60611) and Program for Young Excellent Talents of UIBE (Grant No. 13YQ08).

References

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems, 26*, 1–34.
Abbasi, A., Chen, H., Thoms, S., & Fu, T. (2008). Affect analysis of web forums and blogs using correlation ensembles. *IEEE Transactions on Knowledge and Data Engineering, 20*, 1168–1180.
Aciar, S., Zhang, D., Simoff, S., & Debenham, J. (2007). Informed recommender: Basing recommendations on consumer product reviews. *IEEE Intelligent Systems, 22*, 39–47.
Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering, 17*, 734–749.
Aspden, P. (2010). *Conference of cool*. The Financial Times.
Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the seventh international conference on language resources and evaluation* (pp. 2200–2204). European Language Resources Association.
Balahur, A., Hermida, J. M., & Montoyo, A. (2011). Detecting implicit expressions of sentiment in text based on commonsense knowledge. In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis* (pp. 53–60). Association for Computational Linguistics.
Banker, C., & Gournelos, T. (2013). TED talks and the need for a changing education system. *Florida Communication Journal, 41*.
Bell, R., Koren, Y., & Volinsky, C. (2007). Modeling relationships at multiple scales to improve accuracy of large recommender systems. In *Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 95–104). San Jose, California, USA: ACM.
Bennett, J., & Lanning, S. (2007). The netflix prize. In *Proceedings of KDD Cup and workshop* (pp. 35). ACM.
Bogers, T., & Van Den Bosch, A. (2008). Recommending scientific articles using citeulike. In *Proceedings of the 2nd ACM conference on recommender system* (pp. 287–290). ACM.
Bogers, T., & Van Den Bosch, A. (2011). Fusing recommendations for social bookmarking web sites. *International Journal of Electronic Commerce, 15*, 31–72.
Boiy, E., & Moens, M.-F. (2009). A machine learning approach to sentiment analysis in multilingual Web texts. *Information Retrieval, 12*, 526–558.
Chen, L., & Wang, F. (2013). Preference-based clustering reviews for augmenting e-commerce recommendation. *Knowledge-Based Systems, 50*, 44–59.
Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search engines: Information retrieval in practice*. Addison-Wesley.
Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519–528). ACM.
DaVia Rubenstein, L. (2012). Using TED talks to inspire thoughtful practice. *The Teacher Educator, 47*, 261–267.
Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC* (Vol. 6, pp. 417–422).
Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of ACM, 56*, 82–89.
Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *The Journal of Machine Learning Research, 3*, 1289–1305.
Fox, E., & Shaw, J. (1994). Combination of multiple searches. *NIST Special Publication SP*. 243–243.
Gantner, Z., Rendle, S., Freudenthaler, C., & Schmidt-Thieme, L. (2011). MyMediaLite: A free recommender system library. In *Proceedings of the fifth ACM conference on Recommender systems* (pp. 305–308). Chicago, Illinois, USA: ACM.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Exploration Newsletter, 11*, 10–18.
Heffernan, V. (2009). *Confessions of a TED addict*. The New York Times.
Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 832–844.
Jakob, N., Weber, S. H., Muller, M. C., & Gurevych, I. (2009). Beyond the stars: Exploiting free-text user reviews to improve the accuracy of movie recommendations. In *Proceedings of the 1st international CIKM workshop on topic-sentiment analysis for mass opinion* (pp. 57–64). Hong Kong, China: ACM.
Leung, C.-K., Chan, S.-F., Chung, F.-L., & Ngai, G. (2011). A probabilistic rating inference framework for mining user preferences from reviews. *World Wide Web, 14*, 187–215.
Li, W., Wang, W., & Chen, Y. (2012). Heterogeneous ensemble learning for Chinese sentiment classification. *Journal of Information and Computational Science, 9*, 4551–4558.
Li, Y., Zhai, C., & Chen, Y. (2012). Exploiting rich user information for one-class collaborative filtering. *Knowledge and Information Systems*, 1–25.
Liu, J. G., Guo, Q., & Zhang, Y. C. (2011). Information filtering via weighted heat conduction algorithm. *Physica a-Statistical Mechanics and Its Applications, 390*, 2414–2420.
Liu, H., He, J., Wang, T., Song, W., & Du, X. (2013). Combining user preferences and user opinions for accurate recommendation. *Electronic Commerce Research and Applications, 12*, 14–23.
Liu, L., & Zsu, M. T. (2009). *Encyclopedia of database systems*. Springer Publishing Company, Incorporated.
Montague, M., & Aslam, J. A. (2002). Condorcet fusion for improved retrieval. In *Proceedings of the eleventh international conference on Information and knowledge management* (pp. 538–548). ACM.
Moshfeghi, Y., Piwowarski, B., & Jose, J. M. (2011). Handling data sparsity in collaborative filtering using emotion and semantic based features. In *Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval* (pp. 625–634). Beijing, China: ACM.
Nandakumar, K., Chen, Y., Dass, S. C., & Jain, A. K. (2008). Likelihood ratio-based biometric score fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*, 342–347.
Ohana, B., & Tierney, B. (2009). Sentiment classification of reviews using SentiWordNet. In *Proceedings of the 9th. IT & T Conference* (pp. 13).
Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*, 1–135.
Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing* (Vol. 10, pp. 79–86). Association for Computational Linguistics.
Pappas, N., & Popescu-Belis, A. (2013a). Combining content with user preferences for TED lecture recommendation. In *Proceedings of the 11th international workshop on content-based multimedia indexing (CBMI)* (pp. 47–52). IEEE.
Pappas, N., & Popescu-Belis, A. (2013b). Sentiment analysis of user comments for one-class collaborative filtering over ted talks. In *Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval* (pp. 773–776). Dublin, Ireland: ACM.

Peleja, F., Dias, P., Martins, F., & Magalhães, J. (2013). A recommender system for the TV on the web: Integrating unrated reviews and movie ratings. *Multimedia Systems*, 1–16.

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine, 6*, 21–45.

Qiu, L., Zhang, W., Hu, C., & Zhao, K. (2009). Selc: A self-supervised model for sentiment classification. In *Proceedings of the 18th ACM conference on information and knowledge management* (pp. 929–936). ACM.

Rong, P., Yunhong, Z., Bin, C., Liu, N. N., Lukose, R., Scholz, M., et al. (2008). One-class collaborative filtering. In *Proceedings of the 8th IEEE international conference on data mining* (pp. 502–511).

Siersdorfer, S., Chelaru, S., Nejdl, W., & Pedro, J. S. (2010). How useful are your comments?: Analyzing and predicting youtube comments and comment ratings. In *Proceedings of the 19th international conference on World Wide Web* (pp. 891–900). Raleigh, North Carolina, USA: ACM.

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*. 2–2.

Sugimoto, C. R., & Thelwall, M. (2013). Scholars on soap boxes: Science communication and dissemination in TED videos. *Journal of the American Society for Information Science and Technology, 64*, 663–674.

Takacs, G., Pilaszy, I., Nemeth, B., & Tikk, D. (2009). Scalable collaborative filtering approaches for large recommender systems. *Journal of Machine Learning Research, 10*, 623–656.

Tan, S. L., Bu, J. J., Chen, C., Xu, B., Wang, C., & He, X. F. (2011). Using rich social media information for music recommendation via hypergraph model. In *ACM Transactions on Multimedia Computing Communications and Applications* (Vol. 7).

Tavakolifard, M., & Almeroth, K. C. (2012). Social computing: An intersection of recommender systems, trust/reputation systems, and social networks. *IEEE Network, 26*, 53–58.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology, 62*, 406–418.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology, 63*, 163–173.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology, 61*, 2544–2558.

Turney, P. D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417–424). Association for Computational Linguistics.

Vapnik, V. N. (2000). *The nature of statistical learning theory*. Springer.

Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. *Decision Support Systems, 57*, 77–93.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational Linguistics, 30*, 3.

Wilson, T., Wiebe, J., & Hwa, R. (2006). Recognizing strong and weak opinion clauses. *Computational Intelligence, 22*, 73–99.

Wu, J., Chen, L., Yu, Q., Han, P., & Wu, Z. (2013). Trust-aware media recommendation in heterogeneous social networks. *World Wide Web*, 1–19.

Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences, 181*, 1138–1152.

Yifan, H., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In *Proceedings of the 8th international conference on data mining* (pp. 263–272). IEEE.

Yuanhong, W., Yang, L., & Xiaohui, Y. (2012). Collaborative filtering with aspect-based opinion mining: A tensor factorization approach. In *Proceedings of the 12th international conference on data mining* (pp. 1152–1157). IEEE.

Zhang, W., Ding, G., Chen, L., Li, C., & Zhang, C. (2013). Generating virtual ratings from Chinese reviews to augment online recommendations. *ACM Transactions on Intelligent System Technology, 4*, 1–17.

Zhang, Z. K., Zhou, T., & Zhang, Y. C. (2011). Tag-aware recommender systems: A state-of-the-art survey. *Journal of Computer Science and Technology, 26*, 767–777.

Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms*. Chapman & Hall.

Zhu, Q., Xing, Z., & Liang, J. (2012). Improve top-*k* recommendation by extending review analysis. In Q. Sheng, G. Wang, C. Jensen, & G. Xu (Eds.). *Web technologies and applications* (Vol. 7235, pp. 435–447). Berlin, Heidelberg: Springer.