Information Sciences 575 (2021) 762–778

Contents lists available at ScienceDirect

Information Sciences

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

Adapting topic map and social influence to the personalized hybrid recommender system

Wang Hei-Chia [a,*], Jhou Hsu-Tung [a,b], Tsai Yu-Shan [a]

[a] Institute of Information Management, National Cheng Kung University, Taiwan
[b] Department of Information Management, Chinese Naval Academy, Taiwan

ARTICLE INFO

Article history:
Received 2 March 2017
Received in revised form 19 March 2018
Accepted 3 April 2018
Available online 13 April 2018

Keywords:
Hybrid recommender system
Cold start
Social network
Ontology
Sentiment analysis

ABSTRACT

A recommender system utilizes information filtering techniques to help users obtain accurate information effectively and efficiently. The existing recommender systems, however, recommend items based on the overall ratings or the click-through rate, and emotions expressed by users are neglected. Conversely, the cold-start problem and low model scalability are the two main problems with recommender systems. The cold-start problem is encountered when the system lacks initial rating. Low model scalability indicates that a model is incapable of coping with high-dimensional data. These two problems may mislead the recommender system, and thus, users will not be satisfied with the recommended items. A hybrid recommender system is proposed to mitigate the negative effects caused by these problems. Additionally, ontologies are applied to integrate the extracted features into topics to reduce dimensionality. Topics mentioned in the items are displayed in the form of a topic map, and users can refer to these similar items for further information.

1. Introduction

Currently, people rely on electronic word of mouth (eWOM) to keep in touch with the world. With the rapid growth of user-generated content (UGC), it is much more difficult for users to efficiently extract relevant messages from the archives, especially while making decisions. 92% of users read online reviews, which are a form of UGC, when they search for local businesses. There were four major factors considered by users when judging a local business: the overall rating, the number of reviews, the sentiment of reviews, and the recency of reviews [3]. However, without face-to-face communication, users have to rely on other information to judge whether a review is credible. According to Lis et al. [17], expertise, trustworthiness, and aggregate rating were the positive influential factors for eWOM credibility.

To help users obtain more personalized buying suggestions from excessive UGC, a well-designed recommender system is capable of gathering the appropriate information. Furthermore, the ease of accessing relevant messages for consumers will be beneficial to the businesses. User preference is indispensable to a recommender system. Nevertheless, most existing websites such as Amazon [16] and eBay give recommendations based on the rating similarities of users/items or the connections between users/items, regardless of the emotions that users express in their reviews, and those integrating sentiment analysis at the document level to improve the recommendations lose information that is of value to users.

* Corresponding author.
E-mail address: hcwang@mail.ncku.edu.tw (H.-C. Wang).

https://doi.org/10.1016/j.ins.2018.04.015
0020-0255/

Additionally, recommender systems have suffered from cold starts and low scalability of the models due to the rapid growth in data volume. These problems are the main disadvantages and compromise the performance of a recommender system. Therefore, a personalized recommender system that can address these two main problems is necessary. Recently, some recommendation methods have been proposed to resolve the cold-start problem. Chen et al. [5] proposed a cold start recommendation method for the new user that integrates a user model with trust and distrust networks to identify trustworthy users. And Guo et al. [14] proposed a probabilistic method to derive a prediction from the views of both ratings and trust relationships to accommodate (cold) users who cannot be clustered due to insufficient data. Although these recommendation methods considered relationship trust to decrease the side effects of cold-start problems, some factors affecting the social network were not accounted for, such as user preferences and social influences.

Therefore, this paper aims to build a recommender system that considers topic-level user preferences and social influences between users for the purpose of decreasing the side effects of the cold-start problem and increasing the scalability of the model. To summarize, this paper makes the following contributions.

(1) Extract features from user-generated reviews and integrate these features into topics by means of ontologies to reduce review dimensions. We consider the feature integration as a method for stabilizing the scalability of the model.
(2) Utilize the polarity strengths of descriptive terms to weight user preferences toward topics extracted from features, and establish links between those users who have similar tastes for topics.
(3) Build a recommender system that integrates user influences on social networks constructed by multiple relationships between users for the purpose of diminishing the negative effects caused by the new-user cold-start problem.

2. Related works

This section will present concepts and techniques related to the proposed hybrid recommendation method to more easily understand the development of recent studies.

2.1. Personalized recommender system

It is becoming increasingly difficult for users to identify relevant information from the tremendous amount of user-generated data. The personalized recommender system is one effective way to help users filter out irrelevant messages. It discovers the potential preferred items for users based on user profiles or similarities between users and lists the recommended items for target users. Therefore, the personalized recommender system allows users to retrieve the needed information more efficiently.

2.1.1. Information filtering

Information filtering addresses the dispatch of information that is useful or relevant to the users, and it is also a form of recommender system. Four major techniques for information filtering are described below:

(1) *Content-based Filtering (CBF):* CBF builds user profiles by extracting the features of items that users have preferred and recommends items that have similar properties. Since CBF is performed on the basis of user profiles, it is prone to overspecialize, and thus, its recommendation capabilities are limited.
(2) *Collaborative Filtering (CF):* Through similar users or items, CF determines the preferences of target users toward unknown items. It is the most commonly used filtering technique since it can generate recommendations with limited information [2]. However, the sparsity of the rating matrix results in unacceptable recommendation performance, and negative effects caused by the cold-start problem are evident.
(3) *Rule-based Filtering:* Rule-based filtering generates association rules or sequential patterns through the use of data-mining tools. Consider online shopping as an example. Once the user puts a gallon of milk in the shopping cart, he or she may find several brands of cereal in the recommended list. However, as a less-complicated technique, rule-based filtering is not able to excavate the excluded items.
(4) *Hybrid Filtering:* Hybrid filtering combines multiple filtering approaches or explores auxiliary information as a supplementary source to neutralize the defects of other methods. Thuan and Puntheeranurak [30] constructed a hybrid model based on the combination of product-preference features and review-helpfulness features and then predicted the unknown ratings with the help of nearest neighbors of the target users. Choi et al. [6] derived implicit ratings from transaction data and predicted user preferences using the similarities of the implicit ratings of the items. Because of its exceptional performance in multiple domains, the hybrid filtering technique is adopted in this thesis.

2.1.2. The cold-start problem

The cold-start problem results in unreliable recommendations when the system lacks initial ratings. It has two variants: the new-item cold-start problem and the new-user cold-start problem. Son [29] argued that prior ratings of the new items addressed the former problem, the new-user cold-start problem; on the other hand, this solution was difficult to implement due to the privacy policy. Thus, this thesis mainly focuses on the latter problem.

To mitigate the cold-start problem, several approaches have been proposed. Qian et al. [24] combined social factors, including personal interest, interpersonal interest similarity, and interpersonal influence, to recommend items in which users might be interested. Guo et al. [13] proposed a trust-based recommender model called "Merge," which incorporated the ratings of trusted neighbors in providing recommendations. These approaches improve recommendation performance with the support of the social relationships in the network. Thus, we conceive the idea of merging the social influence between users into the hybrid recommender model.

2.2. Social-network analysis

Social-network analysis (SNA) focuses on the structured relationships between individuals or groups. A social network is often expressed in the form of a graph (see Fig. 1). A node represents a user in the system, and an edge represents the relationship between two users. SNA is also useful for user-behavior analysis, network modeling, recommendation-system construction [12], etc. Centrality is one of the core measures in SNA [23]. It estimates the importance of a node in the network, and there are various ways to assess the centrality of a node, such as degree, betweenness, and closeness.

Degree centrality is defined as the number of relationships a node has, and two distinct measurements in the directed graph are the indegree centrality and outdegree centrality. If the indegree centrality of a node is high, it is likely to be an authority site. If the outdegree centrality is high, it tends to be a hub site. However, degree centrality will have a bias on the importance of a node if the node is within a few connections of two clusters.

Betweenness centrality captures the importance of a node in the flow of information. A node with high betweenness is a key intermediary between other nodes. Despite that, its applicability is limited due to the high computational cost, which is a major indicator of centrality [21]. Closeness centrality represents the closeness between a node and others in the graph. A large closeness centrality means that the node is not far from the center.

We assume that if a user is closer to the center of the social network, his or her preferences are more likely to impact the preferences of other users. Additionally, Shafiq et al. [27] indicated that utilizing social-network analysis data produced more relevant information for users based on their interests. Accordingly, we attempt to introduce SNA into our recommender model for the purpose of reducing the side effects of the cold-start problem.

2.3. Natural language processing

Natural language processing (NLP) is a field of computer science that explores the comprehensible interaction between computer language and human language. Most applications in NLP require semantic and grammatical analyses to solve the artificial intelligence problem. Recently, some techniques based on computational intelligence (artificial neural networks, evolutionary programming, etc.) have been used to solve the artificial intelligence problem. Evolutionary computation is a promising approach because it does not require any derivative information, contrary to the conventional gradient-based methods. The linear dynamic systems based on artificial bee colony (ABC) algorithm is flexible and applicable in evolutionary computation [22]. In addition, other techniques relevant to NLP are discussed in the rest of this section.

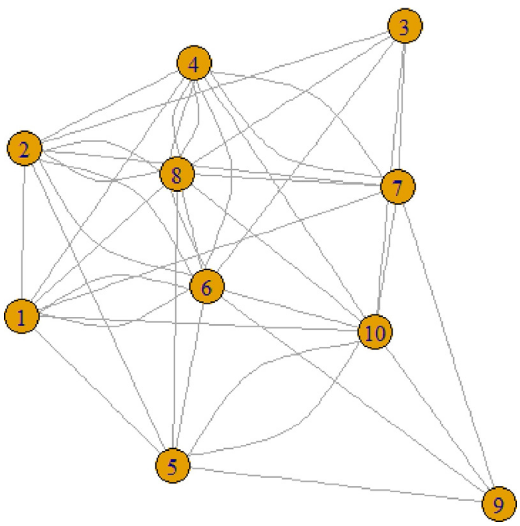


Fig. 1. Concept of SNA.

2.3.1. Part-of-Speech tagging

Part-of-Speech (POS) tagging refers to the syntactic classification of terms in a sentence on the basis of their contexts. There are two major approaches for automatically labeling POS tags, rule-based tagging and statistical tagging [11].

Rule-based taggers assign possible POS tags based on dictionaries and establish language-dependent rules to determine a proper POS tag for each term. Statistical taggers use a tagged corpus to build probabilistic models for tag sequences.

A few studies have revealed that statistical approaches perform better than rule-based approaches [28]. Therefore, we apply a well-known probabilistic POS tagger, the Stanford POS tagger [31], in Section 3 to extract features and their corresponding opinions from the obtained review set.

2.3.2. A vector-space model

A vector-space model (VSM) is a model for representing documents as vectors of term weights [15]. Let the document set be D= $\{d_1, \ldots, d_j, \ldots, d_{|D|}\}$, and the weight vector of document d_j be $\vec{d_j} = (w_{1j}, \ldots, w_{ij}, \ldots, w_{|T|j})$, where $|D|$ is the size of document set D, $|T|$ is the total number of terms that occur in D, and w_{ij} is the weight of term t_i in document d_j. As shown in Fig. 2, the smaller the deviation between the angles $\vec{d_1}$ and $\vec{d_2}$, the more similar they are semantically.

The common term-weighting schemes are listed below [9,10]:

(1) Term frequency (TF): TF considers the frequency of a term in a given document to be the term weight.

$$TF(t_i, d_j) = \frac{n_{ij}}{\sum_k n_{kj}} \tag{1}$$

$TF(t_i, d_j)$ represents the TF of term t_i in document d_j, n_{ij} is the number of occurrences of the given term t_i in document d_j, and $\sum_k n_{kj}$ represents the total number of terms in document d_j.

(2) Term frequency-inverse document frequency (TF-IDF): TF-IDF represents the TF scheme that penalizes the frequency of a given term across the document set.

$$TFIDF(t_i, d_j) = TF(t_i, d_j) \times log \frac{|\boldsymbol{D}|}{|\{j|t_i \in d_j\}|} \tag{2}$$

$TFIDF(t_i, d_j)$ is the TF-IDF of term t_i in document d_j, $|\boldsymbol{D}|$ is the size of document set $\boldsymbol{D}$, and $|\{j|t_i \in d_j\}|$ represents the number of documents that include the term t_i.

(3) Mutual Information (MI): MI represents the mutual dependence of a given term and the document set in a certain class.

$$MI(t_i, \boldsymbol{D}_c) = \log \frac{p(t_i, \boldsymbol{D}_c)}{p(t_i) \times p(\boldsymbol{D}_c)} \tag{3}$$

$MI(t_i, \boldsymbol{D}_c)$ is the MI of term t_i and document set $\boldsymbol{D}_c$, $\boldsymbol{D}_c$ is the document set in the specific class, $p(t_i, \boldsymbol{D}_c)$ is the joint probability of term t_i and document set $\boldsymbol{D}_c$, and $p(t_i)$ and $p(\boldsymbol{D}_c)$ represent the probabilities of t_i and $\boldsymbol{D}_c$, respectively.

In addition to these weighting schemes, there are other approaches to assess term weights. Most of them focus on quantitative measurements and disregard the emotions a user has expressed toward a given feature. Therefore, we attempt to adopt a modified weighting scheme in which sentiment strength is considered.

2.4. Sentiment analysis

Opinions in the reviews reflect user preferences. Therefore, it is necessary to analyze the emotions expressed in the user-generated documents. Sentiment analysis, also known as opinion mining, is used for recognizing and extracting the emotions

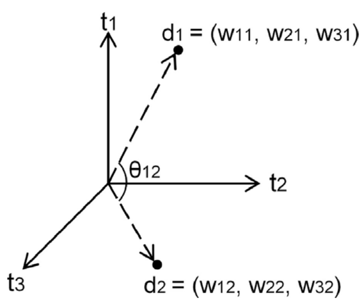


Fig. 2. Vector space model diagram.

a user expresses from the partial or whole document, i.e., determining the sentiment polarity from opinion terms. Accordingly, an opinion term is divided into three types of polarity: positive, objective, and negative. Furthermore, the polarity strength of an opinion term is assigned based on its sentiment intensity.

Sentiment analysis can be implemented at the document level, sentence level, and feature level. Lu et al. [18] argued that the overall rating might fail to properly represent the aspects that are preferred and therefore introduced the LDA topic model for sentiment analysis. Since analyzing sentiment at the coarser level was less precise due to the conflicting sentiment polarities in the collected data, this thesis will conduct sentiment analysis at the feature level.

2.4.1. Keyword spotting methods

The existing approaches in sentiment analysis are grouped into four categories: keyword spotting, lexical affinity, statistical methods, and concept-based techniques [4]. Keyword spotting classifies the polarity of a text based on the appearance of affective terms such as "good" and "awful," but it cannot address negation words or ambiguous terms. Lexical affinity assigns probabilities of a particular sentiment to various terms by means of training on a given corpus. Since it estimates the polarity at the term level, the probabilities of negations and ambiguities may be biased, and the results may be influenced by the chosen corpus.

Seol et al. [26] presented a hybrid system that decomposes to keyword-based and machine learning methods. If the input sentence has affective keywords, keyword-based approach is applied. In other cases, the system uses the machine learning approach – Knowledge-Based Artificial Neural Network (KBANN) – to identify the emotion of sentences. Recently, the learning ability of artificial neural networks makes them useful for dynamic identification and prediction. Coban [7] proposed the training procedure of a recurrent neural network by adding a context layer, which is easily trainable using a learning algorithm.

2.4.2. Statistical methods

A statistical method is a type of supervised machine-learning algorithm that feeds on a considerable amount of annotated text. Therefore, it is suitable for documents with substantial content, such as news and academic reports, but not for UGC such as tweets and comments. Concept-based methods obtain the conceptual and affective information of opinions through the use of ontologies and detect sentiments at a fine-grained level. However, the ability to distinguish between opinions is constrained by the integrity of its knowledge base.

Owing to their extensive coverage of semantic information, concept-based methods can be considered a solution to complement the drawbacks of lexical affinity. Accordingly, we perform sentiment analysis for user-generated reviews with the help of sentiment lexicons and semantic networks.

2.5. Topic map

The construction of a topic map is a standard approach for knowledge integration, and it can be used to sketch out the semantic relations between topics [34]. The core elements of a topic map are listed below:

(1) Topic: A topic represents an entity or a concept.
(2) Association: Topics are connected to one another by associations.
(3) Occurrence: An occurrence points the topic to the relevant resources, such as articles and films.

The framework of a topic map is shown in Fig. 3. A given resource, "Review 3," is linked to topics such as "Wine" and "Parking"; and "Wine" is related to "Tea" since they share the same association "is a kind of drink." A topic map provides matches for several resources, and as an information provider, it helps users locate the corresponding resources for topics of interest.

2.6. Ontology

Since the number of terms extracted from documents is massive, dimension reduction is necessary. Furthermore, it is crucial to maintaining the integrity of the information. Therefore, we intend to extract the representative topics from numerous features with the use of ontologies.

An ontology defines properties and interrelations of instances in a particular domain. Since the meanings of an instance vary between domains, it is crucial to identify the real sentiment that an opinion implies. For instance, a user commented on the price of Apple products. In this scenario, the term "apple" represents a company instead of an edible fruit of a temperate-zone deciduous tree.

For the sake of automatic recognition of synonyms and polysemous terms in the documents, it is necessary to identify the relations between the extracted features. The well-known lexical resource WordNet groups terms into synsets, and synsets are interlinked on the basis of their relations with one another [32]. Nevertheless, due to the spread of new words and special terms, WordNet is no longer an adequate solution to provide a suitable semantic network.

Therefore, another lexical database, named BabelNet, is considered. BabelNet integrates WordNet, Wikipedia, and other resources to construct the semantic network [19]. As the wisdom of the crowd, Wikipedia is a satisfactory complement to

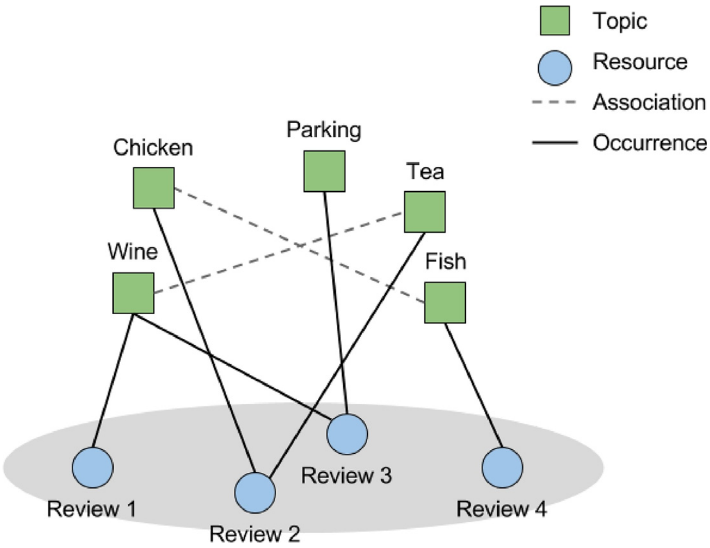


Fig. 3. Topic map.

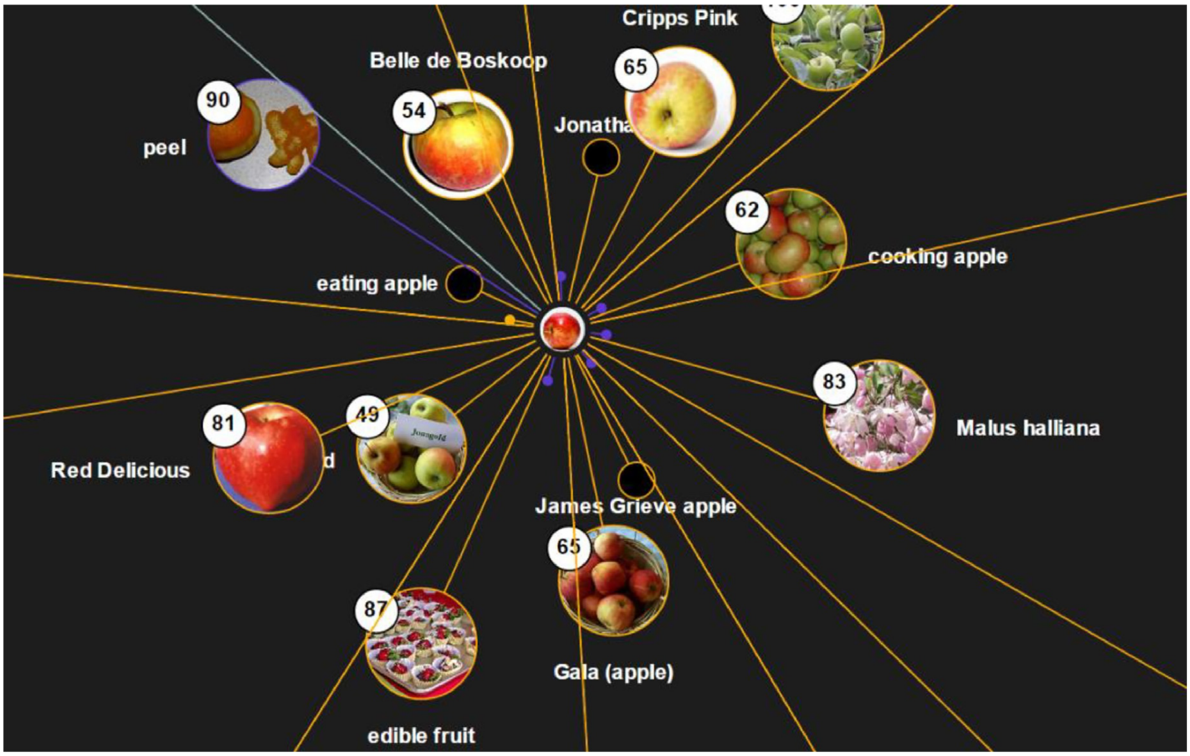


Fig. 4. Ontology map [20].

WordNet. With a wider coverage of terms, BabelNet addresses rarely used words. It is on such grounds that BabelNet is put to use. Fig. 4 shows the ontology of the term "apple" provided by BabelNet.

2.8. Summary of the related works

In summary, the hybrid filtering technique is adopted in the thesis for the purpose of mitigating the cold-start problem and maintaining the scalability of the model. For the cold-start problem, utilizing social-network analysis allows the model to gain more information about users. The social influence of a user in the network is measured by the closeness centrality, which is capable of determining the importance of a node in the network.

Additionally, UGC is used to extract user preferences with the help of the semantics-based sentiment lexicon to build one of the edges in the social network. To maintain the number of dimensions in the reviews, features are integrated into topics on the basis of ontologies provided by BabelNet, which was chosen for its rich semantic relations and wide coverage of terms.

3. Research methods

The proposed model consists of four parts: the data preprocessing module, topic mapping module, sentiment analysis module, and hybrid recommender module. The collected data are preprocessed before further analyses to reduce the noise in the dataset, and features are extracted along with opinions and multipliers from the obtained reviews. The extracted features are then expanded with the help of ontologies, and a representative topic is assigned to each cluster of similar features.

Next, the sentiment weight of each feature is measured by computing the polarity strength of its processed opinions and multipliers, and the sentiment weight of a topic is aggregated according to the features represented by the topic. The similarities between users are then calculated to determine whether or not users are alike with respect to their topic sentiments. Finally, the social-influence network is constructed based on the relationships between users, and the predicted ratings of the unknown items are estimated accordingly. Fig. 5 shows the framework of the proposed recommender system.

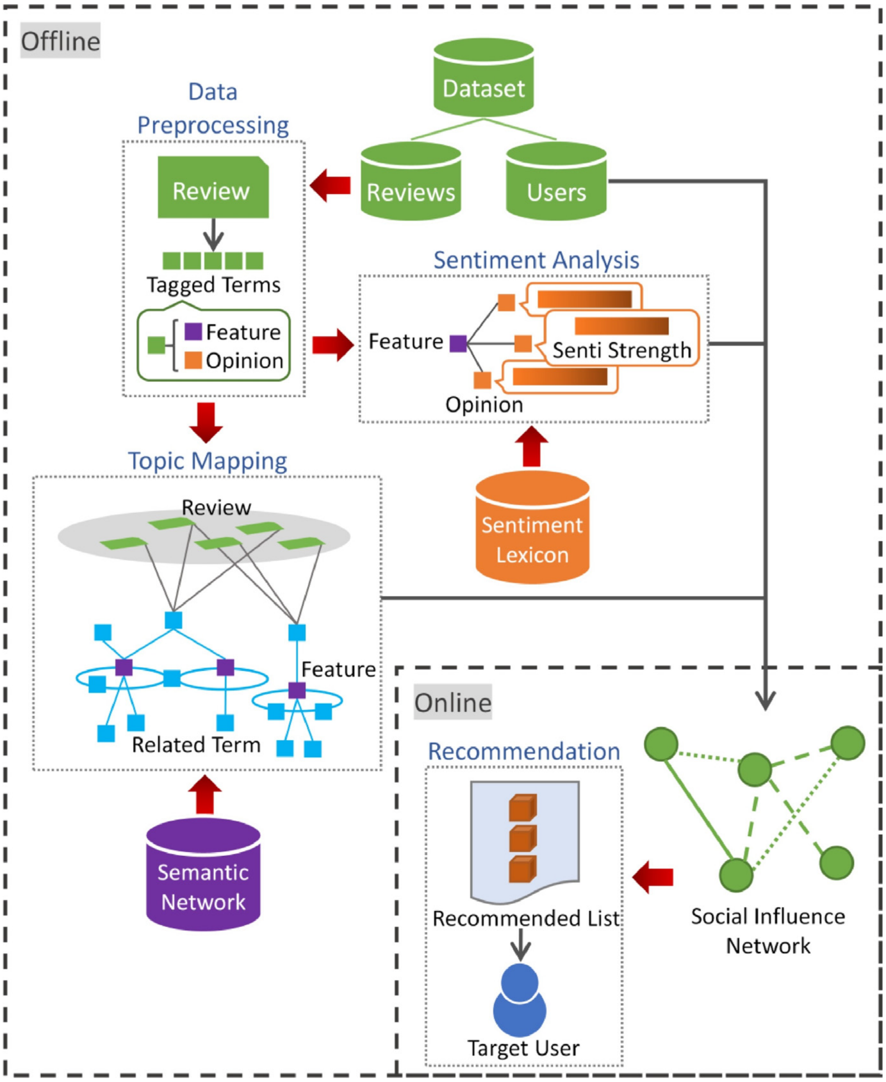


Fig. 5. Overview of the proposed framework.

3.1. Scope and limitations

In this paper, the sentiment score is computed with the help of a POS tagger and the sentiment lexicon, and topics in the reviews are extracted based on the semantic resource. Accordingly, the proposed methods are not suitable for languages lacking the foregoing techniques and resources. Additionally, information on the social relationships between users is required. Based on privacy considerations, the personal information of users in social networks is protected, and we only observe part of the social relationships. If the dataset does not include the social information in detail, the proposed model is not applicable. Since the review recency is one of the four major factors of concern in judging a local business, we only consider the latest review that the user posted for the business.

3.2. Data preprocessing module

In this paper, data preprocessing includes four steps. After tokenization, words are labeled with a part-of-speech (POS) tag using the Stanford Part-Of-Speech Tagger. Next, different types of terms are extracted based on their POS tags. Nouns and consecutive nouns are selected because the features are generally nouns or noun phrases [25].

The third step is the extraction of descriptive terms. The surrounding verbs and adjectives of a given feature, whose path lengths are shorter than the threshold δ_d, are chosen as opinions of the feature because descriptive terms usually lie close to the subject, i.e., a noun/noun phrase. In the case of adverbs, Dragut et al. [8] indicated that adverbs were equivalent to sentiment intensifiers. Consequently, those adverbs occurring with opinions are annotated as multipliers for further polarity-strength measurement.

After extracting the necessary terms, these words are lemmatized. In the end, a review is regarded as a set of features, opinions, and multipliers. Fig. 6 exemplifies the result of data preprocessing.

3.3. The topic mapping module

A lexical resource, named BabelNet, is introduced into the model for the purpose of feature expansion after data preprocessing. In this paper, a relation between words belongs to one of three types in the relation set $\boldsymbol{L} = \{hypernym,\ synonym,\ hyponym\}$, and words related to feature f are retrieved and assigned a weight w_l based on their relations with feature f. In the structure of semantic fields, the higher the level a term reaches, the more general the meaning it represents. Under the concept of taxonomy, the weight w_l is defined as

$$w_l = \begin{cases} 5, & l = hypernym \\ 3, & l = synonym \\ 1, & l = hyponym \end{cases} \quad (4)$$

Next, the modified Jaccard similarity coefficient is used for measuring the similarity between features f and f'. X_{fl} and $X_{f'l}$ in the Jaccard function represent the related-term set of feature f and feature f' based on relation l, and $\boldsymbol{F}$ is the feature set extracted from the reviews.

$$sim(f,f') = \frac{\sum_{l\in L} w_l \times |X_{fl} \cap X_{f'l}|}{\sum_{l\in L} w_l \times |X_{fl} \cup X_{f'l}|},\ \forall f \in \mathrm{F} \Lambda \forall f' \in \mathrm{F} \quad (5)$$

Once the similarity crosses the threshold δ_f, features f and f' are regarded as sharing common characteristics, and therefore these similar features are grouped. Algorithm 1 demonstrates the process of feature integration. A topic term is chosen to represent the words in the same group, based on the most frequently occurring hypernym in the group.

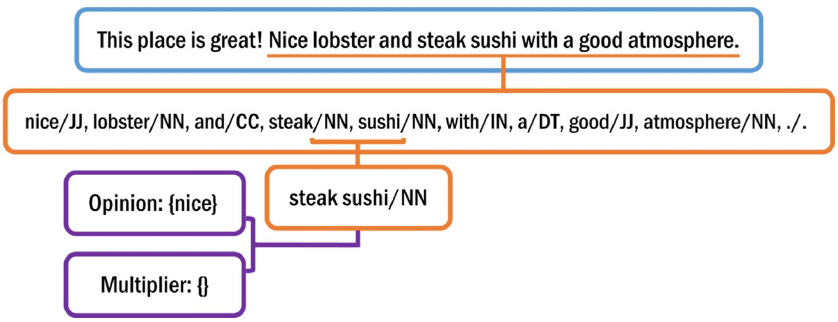


Fig. 6. Example of data preprocessing.

```
for feature in featureSet do
    if feature not belong to any group in groupSet then
        add feature to the new group
    for feature' in featureSet do
        if sim(feature, feature') _ _f then
            if feature' not excluded from group then
                if feature' not belong to any group in groupSet then
                    add feature' to the group
                else
                    if maxSim(feature', group') < sim(feature, feature') then
                        transfer feature' from group' to group
        else
            exclude feature' from group
for group in groupSet do
    set the most frequently-occurring hypernym of group to be topic of group
```

Fig. 7. Algorithm 1: integration of features into topics.

3.4. Sentiment analysis module

The extracted descriptive terms are each assigned a sentiment score using SentiWordNet to account for user's interests [1]. Each word has multiple senses, i.e., meanings, and each sense of the word has three types of polarity strength: positive, objective, and negative. The *k*th sense of opinion *o* is denoted by o^k, and the *k*th sense of multiplier *m* is denoted by m^k. Based on the three types of polarity strength, the sentiment score of o^k is restricted to the range from −2 to 2, and it is defined as

$$S_{o^k} = 2 \times \left[ps^{+(o^k)} - ps^{-(o^k)}\right] \tag{6}$$

The sentiment score s_{m^k} is in the range between −1 and 1, and it is defined as

$$s_{m^k} = ps^{\,+\left(m^k\right)} + ps^{*\left(m^k\right)} - ps^{-\left(m^k\right)} \tag{7}$$

The functions ps^+, ps^*, and ps^- map the descriptive term into the degrees of positivity, objectivity, and negativity, respectively. After that, the sentiment scores of o^k and m^k, i.e., s_{o^k} and s_{m^k}, are integrated into the sentiment scores of opinion *o* and multiplier *m*, i.e., s_o and s_m, respectively. For the purpose of normalization, s_{o^k} and s_{m^k} are normalized based on the sense ranks.

Next, the sentiment weight of feature *f* is determined using the sentiment scores of its opinions and multipliers. The weight of feature *f* is defined as

$$w_f = \frac{\sum_{o \in \boldsymbol{o}_f} s_o}{|\boldsymbol{o}_f|} \times \left(I_{\sum\limits_{m \in M_f} s_m \geq 0} + I_{\sum\limits_{m \in M_f} s_m \neq 0} \times \frac{\sum_{m \in M_f} s_m}{|M_f|} \right), \ \forall f \ \in \ F_{ui} \tag{8}$$

where $\boldsymbol{O}_f$ and $\boldsymbol{M}_f$ are the opinion set and the multiplier set of feature *f*, $\boldsymbol{F}_{ui}$ is the feature set of item *i* of user *u*, and I is the indicator function that determines whether the condition is satisfied.

For those features without opinions, the feature weight w_f is assigned a default value, which is equal to 1 in this paper, because these mentioned features are supposed to be more important than the unmentioned ones. Next, we consolidate feature weights to form topic weights. Only the strongest emotion a user expresses toward the topic is retained to be the weight of the topic for the purpose of eliminating opinions in which revealed emotions are too obscure to determine the user's interest in the topic. The weight of the topic is defined as

$$w_t \ = \ w_f, if \ |w_f| \ \geq \ |w_{f'}| \forall f \ \in \ \boldsymbol{F}_t \ \Lambda \ \forall f' \ \in \ \boldsymbol{F}_t \ \setminus \ \{f\} \tag{9}$$

where $\boldsymbol{F}_t$ is the feature set classified to topic *t*. After the integration of sentiment scores, each review is represented as a vector of topic weights and is defined as

$$\vec{r_{ui}} \ = \ \left(w^{ui}_{t1}, \ldots, w^{ui}_{tj}, \ldots, w^{ui}_{t|T|} \in \mathrm{R}^{|\boldsymbol{T}|}\right) \tag{10}$$

where w^{ui}_{tj} is the *j*th topic weight given by user *u* to item *i* and $\boldsymbol{T}$ is the extracted topic set.

3.5. Hybrid recommender module

In this paper, if two users show similar tastes in topics, they are regarded as neighbors. To find the nearest neighbors, the cosine similarity between users is computed. First, we average the review vectors to obtain the user vector. Those users with similarities above the threshold δ_n with the target user are considered the neighbors of the target user. Similarity between users is defined as

$$sim(u,u') = \frac{\vec{r_u} \cdot \vec{r_{u'}}}{\|\vec{r_u}\| \times \|\vec{r_{u'}}\|} \tag{11}$$

After identifying neighbors with similar tastes in topics, these users are related because of their similarities with one another. However, Guo et al. [13] indicated that recommendation performance was improved with the help of trusted users. Therefore, people are likely to identify themselves with the trusted users, and their opinions are likely to be consistent. Friendship is also considered since people are prone to be affected by their friends, and they may consult their friends while making decisions [35]. Accordingly, a social network is built on the basis of these three relationships. Fig. 8 displays the network between users.

In the social influence network, edges are undirected since the efficacies of the three relationships are considered equivalent, and the social influence of a user is measured by the closeness centrality because the reviews written by a well-connected user are prone to shift the opinions of other users. The closeness centrality of user u, i.e., cc_u, is defined as

$$cc_u = \frac{|\boldsymbol{U}| - 1}{\sum_{u' \in U \setminus \{u\}} sd(u,u')} \tag{12}$$

where cc_u is the user set, and sd(u, u') is the shortest distance between users u and u′. To calculate the shortest distance, we utilize a breadth-first search (BFS) algorithm for traversing the social influence network, due to the unweighted graph.

After that, we account for the distinct rating habits of users to predict the user preferences. Accordingly, the predicted rating is estimated by

$$p_{ui} = \begin{cases} \dfrac{\sum_{u' \in U \setminus \{u\}} cc_{u'} \times a_{u'i}}{\sum_{u' \in U \setminus \{u\}} cc_{u'}}, |I_u| \le 1 \\ \bar{a}_u + \dfrac{\sum_{u' \in U \setminus \{u\}} cc_{u'} \times \left(a_{u'} - \bar{a}_{u'}\right)}{\sum_{u' \in U \setminus \{u\}} cc_{u'}}, |I_u| > 1 \end{cases} \tag{13}$$

where p_{ui} is the possible rating that user u may give to item i, $a_{u'i}$ is the actual rating given to item i by user u′, $\bar{a}_u$ and $\bar{a}_{u'}$ are the average ratings given by user u and user u′, and I_u is the set of items on which user u commented.

Among the predictive outcomes, the recommender system adjusts the predicted rating to the normal rating range of the target user. Additionally, since social relationships complement the deficiency of a cold-start user, it is feasible for the recommender system to estimate ratings that a user is likely to give. Finally, the recommender system gives a recommended item list based on the predicted ratings of the target user, sorted in descending order.

4. Experiments and analyses

4.1. Experimental settings

Figs. 9 and 10 demonstrate how the system presents the recommended businesses to the target user. A list of businesses sorted according to the predicted ratings is recommended to the user, and users can check the information about each business, including the geographical location, the degree of user preference, and the most frequently mentioned topics. Accordingly, it gives users a brief overview of many aspects of a business and identifies the preferred business efficiently since it is impossible for users to experience every single business personally.

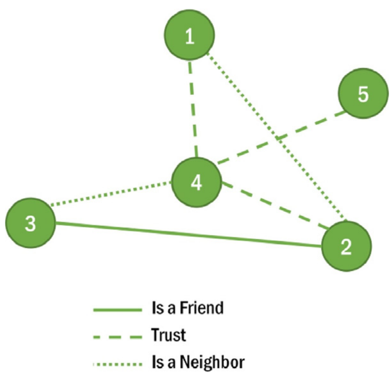


Fig. 8. Social influence network.

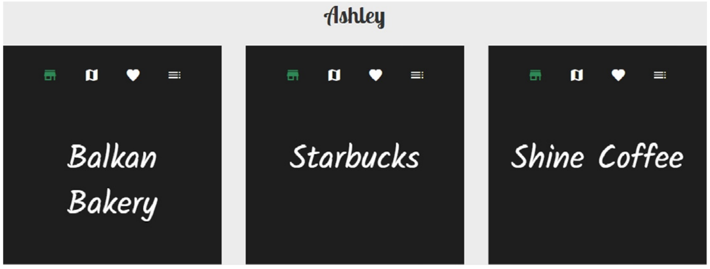


Fig. 9. Example of recommended list interface.

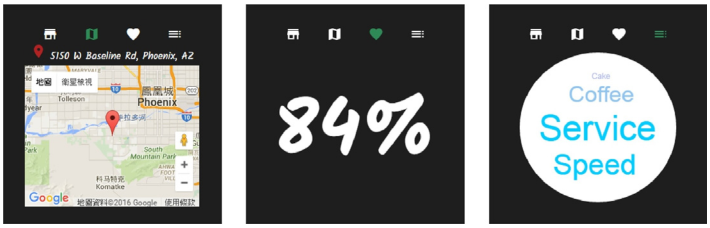


Fig. 10. Example of the recommended business.

As for the dataset in the experiments, the Yelp Dataset Challenge [33] was chosen to evaluate the proposed method because of its popularity and trustworthiness. The dataset consisted of 3276 users, 1493 businesses commented on by at least two users, and 14,003 reviews posted from 2005 to 2015.

The dataset was sparse: the average number of reviews per person (NRU) was approximately four, and there were 1311 users who wrote only one review. The sparsity of the dataset was approximately 99.714%. Since nearly 40% of the users posted only one review, we defined cold-start users as those users with an NRU equal to one.

To avoid overfitting, 10-fold cross validation was utilized as the evaluation strategy. Two types of evaluation metrics were selected for further comparisons. The first type was the metrics of rating prediction, which measured the gap between the predicted values and the actual values. Root mean square error (RMSE) and mean absolute error (MAE) were used in the following experiments:

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} \left(r_i - \hat{r}_i \right)^2} \tag{14}$$

$$\boldsymbol{MAE} = \frac{1}{n} \times \sum_{i=1}^{n} |r_i - \hat{r}_i| \tag{15}$$

where r_i is the observed value, $\hat{r}_i$ is the predicted value, and n is the number of observations. The other type is the metrics of preference classification. Precision, recall, F1, and accuracy (ACC) were used for evaluation. According to the confusion matrix listed in Table 1, these metrics are defined as:

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \tag{16}$$

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} \tag{17}$$

$$F1 = \frac{2 \times FP}{2 \times TP + FP + FN} \tag{18}$$

$$\boldsymbol{ACC} = \frac{TP + TN}{TP + FP + FN + TN} \tag{19}$$

4.2. Determination of thresholds

Before implementing algorithms of the proposed model, we needed to define parameters and determine their values. First, the performance of feature extraction was evaluated. Three-hundred seventy-four reviews were randomly selected

Table 1
Confusion matrix.

		Actual condition	
		Positve	Negative
Predicted Condition	Positve	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

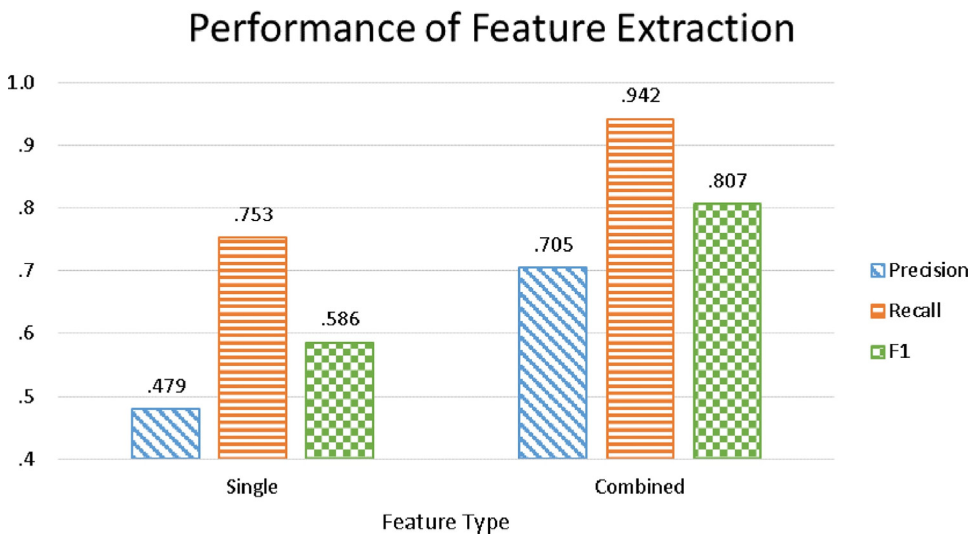


Fig. 11. Performance of feature extraction.

and manually marked. Fig. 11 shows that combinations of the consecutive nouns were more illustrative of features than single nouns due to "noun as adjective." Accordingly, it was reasonable to combine the consecutive nouns.

Next, since the word distance determines which terms should be the opinions or the multipliers of a feature, a distance threshold δ_d was required to abate the noise while covering more descriptive terms.

Figs. 12 and 13 show the results yielded by different word distances. It is important to retrieve the relevant terms, i.e., to have high recall. However, recall is easily biased if more descriptive terms are included, and this results in low precision due

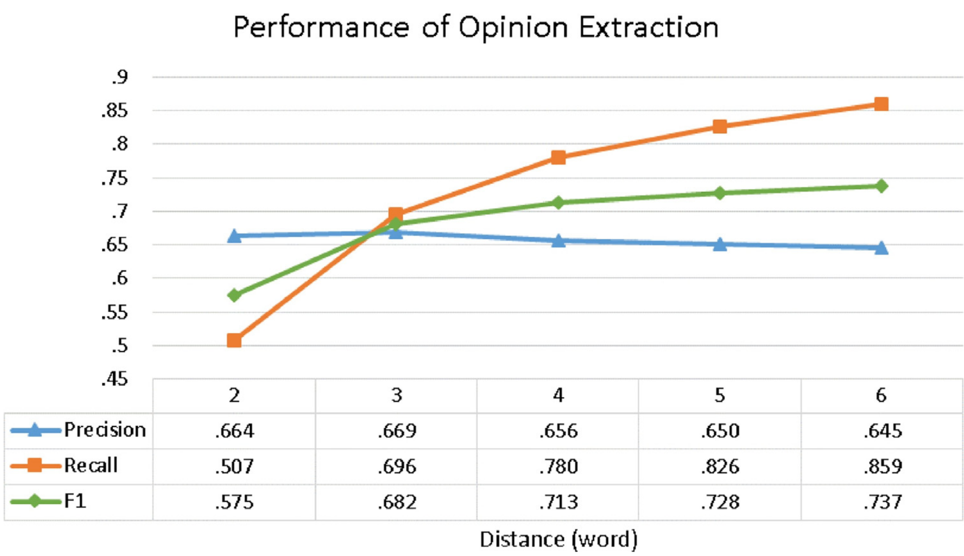
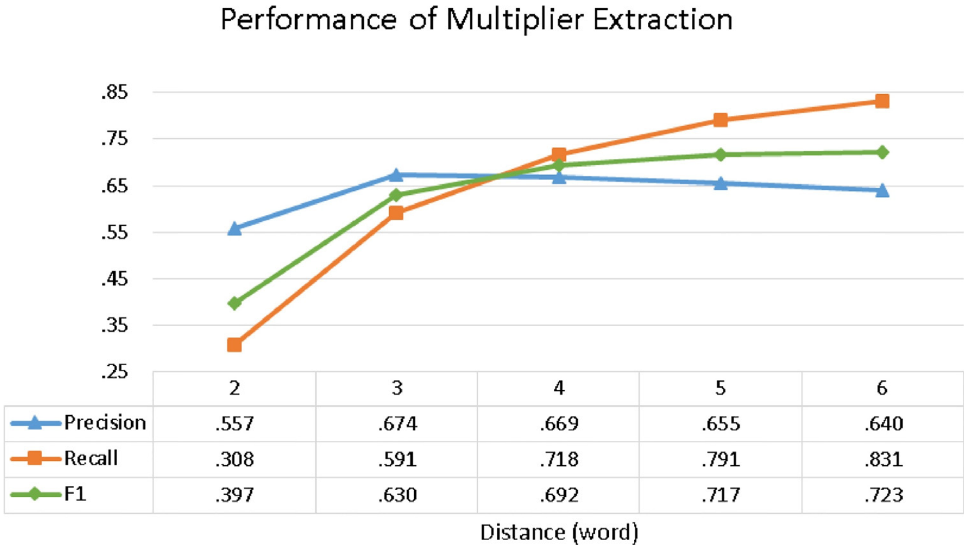


Fig. 12. Performance of opinion extraction.



	2	3	4	5	6
Precision	.557	.674	.669	.655	.640
Recall	.308	.591	.718	.791	.831
F1	.397	.630	.692	.717	.723

Fig. 13. Performance of multiplier extraction.

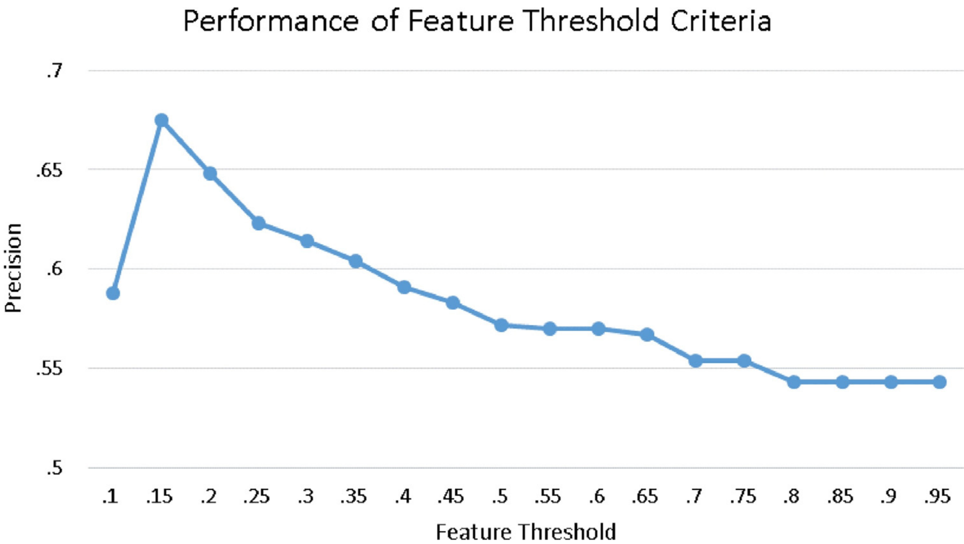


Fig. 14. Performance of feature integration.

to too much noise. As a result, the F1 measure was used to evaluate different word distances. Both distance thresholds stabilized gradually when the word distance was larger than 4. Accordingly, the distance threshold δ_d was set to 4.

The feature threshold δ_f determines the number of dimensions of a review. We randomly sampled and marked 381 features from the feature set to evaluate the integration performance.

The less similar the features within groups are, the less likely it is for them to belong to the same topic. However, the choice of topics may be limited when the threshold is too high because the topic is randomly selected if the hypernym frequencies are the same. The feature threshold δ_f was fixed to 0.15 due to its better performance.

For the purpose of retrieving users with similar preferences, different thresholds were tested to find the one that results in better recommendation performance. Fig. 15 demonstrates positive correlations between two evaluation metrics and the neighbor threshold δ_n. The higher δ_n is, the more likely it is to diminish the possibility of predicting user preference accurately since users who have been to the businesses may not have similar tastes with the target user, and this results in the failure of the prediction.

Fig. 16 displays the performance of the preference classification with different thresholds. High precision means that if the system can predict the rating, this rating will approximate the actual rating that the target user would give. However,

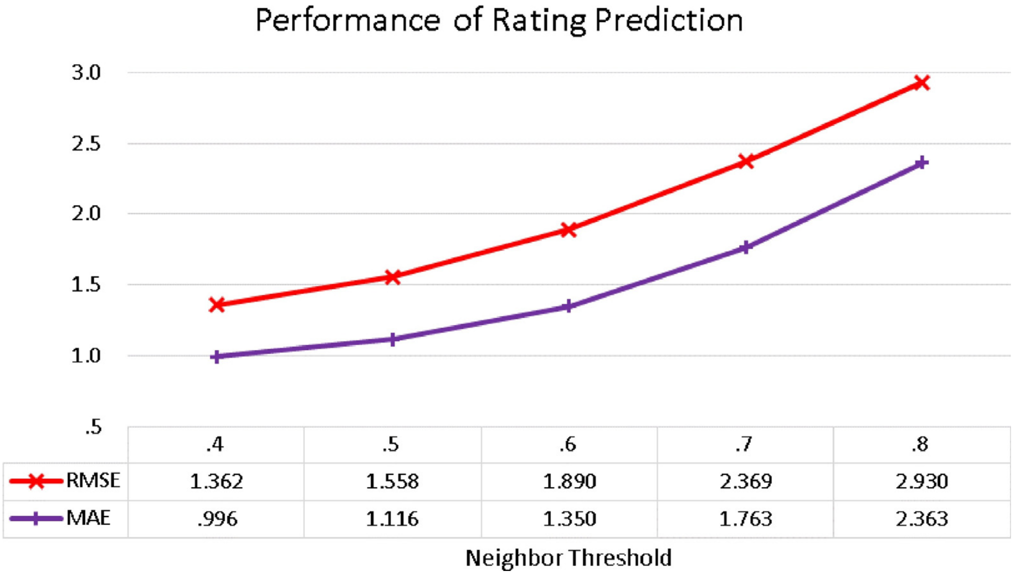


Fig. 15. Rating gap under different neighbor thresholds.

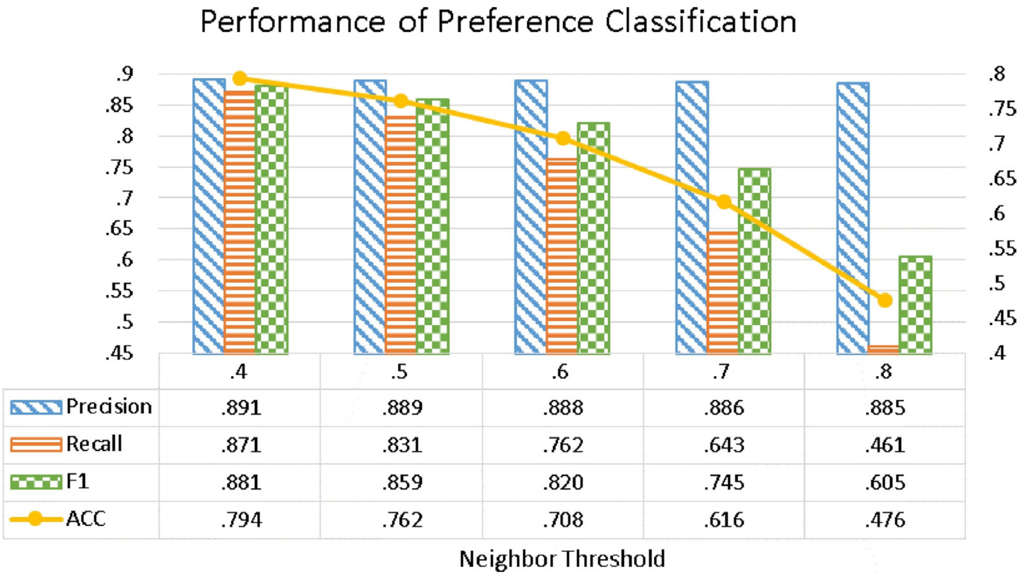


Fig. 16. Preference classification under different neighbor thresholds.

insufficient links between users lead to the failure of the rating estimation and therefore have a negative effect on recall and ACC. Additionally, this results in a decrease in the F1 measure.

From Figs. 15 and 16, we observe that when the similarity between users is greater than 0.4, the recommendation performance was optimal. As a result, the neighbor threshold δ_n was set to 0.4.

4.3. Comparison of relationships

The impacts of different types of relationships on user preference varied. To identify which of them was the most influential, we experimented with all of the combinations to evaluate their recommendation performances. In Figs. 17 and 18, the letter “T” represents trust, the letter "F" represents friendship, and the letter "N" stands for neighbor, i.e., the similar taste between users.

From both figures, we observe that the performances of these relationships are all good, and these values imply that the proposed method addresses the data sparsity and reduces the side effects caused by the cold-start problem. However, friend-

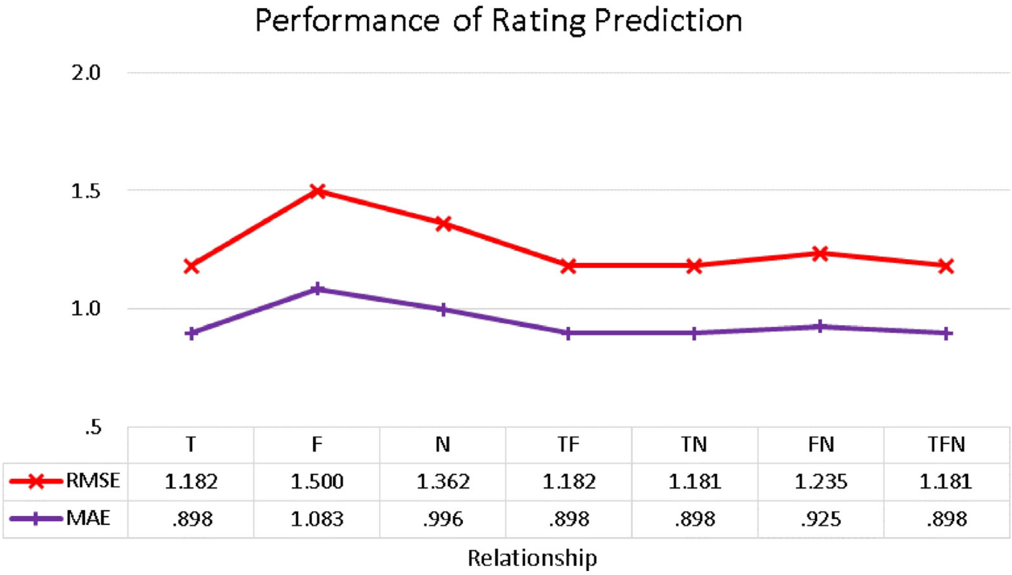


Fig. 17. Rating gap of seven relationships.

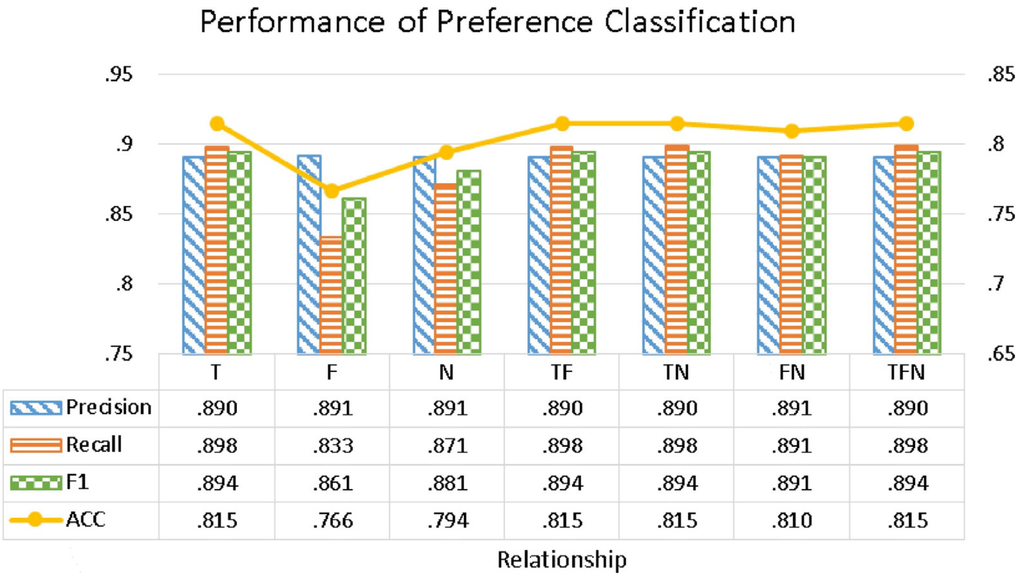


Fig. 18. Preference classification performance of seven relationships.

ship does not perform well among all relationships since the businesses that friends recommend are not always consistent with the ones that the target user prefers. Additionally, the number of comments on a business is also an influential factor of the recommender system. As a result, the impact of friendship is not as strong as the impacts of the other factors due to insufficient information about businesses.

Regarding the combination of relationships, it was difficult to determine which relationship is the most helpful to estimate user preference. Therefore, hypothesis testing was conducted to determine statistically whether the rating gaps of these seven relationships were equal. We ran Brown-Forsythe and Welch tests instead of analysis of variance (ANOVA) due to the rejection of Levene's test, i.e., the test for the homogeneity of variances, which is the basic assumption of ANOVA.

In both tests, p-values were less than 0.05, which is the significance level of the test. Since the null hypothesis was rejected, the test indicated that there were significant differences between the performances of these seven relationships. After running the Brown-Forsythe and Welch tests, the Games-Howell test was used for follow-up post hoc analysis. According to the p-values in Table 2, the performance of the "neighbor" relationship was better than "friendship," but it did not perform as well as the other five relationships. From the statistical result, there was no significant difference between these

Table 2
Significance between two relationships.

	T	F	N	TF	TN	FN	TFN
T	–	0.000***	0.000***	1.000	1.000	0.506	1.000
F		–	0.000***	0.000***	0.000***	0.000***	0.000***
N			–	0.000***	0.000***	0.000***	0.000***
TF				–	1.000	0.506	1.000
TN					–	0.498	1.000
FN						–	0.498
TFN							–

***p < 0.001.

five relationships. Accordingly, we conclude that despite the lack of the "trust" relationship, the recommender system predicts user preference with the help of friendship and the users' personal interests.

5. Conclusions

Due to the rapid growth of UGC, it is becoming increasingly difficult for people to extract relevant information from massive data. A personalized recommender system, as an effective information filtering technique, is capable of recommending items that users are likely to prefer and reducing the time cost to users. Therefore, we proposed a recommendation method that combined topic mapping and social influence to decrease the negative effects resulting from the cold-start problem and to stabilize the data dimensions.

This paper utilized user information, including user-generated reviews and social information, to build a personalized recommender system. Latent user interest was extracted using SentiWordNet and BabelNet. After measuring user similarities, neighbors with similar tastes were extracted, and users were linked together based on their similar interests, along with friendship and trustworthiness, to build a social network. Afterwards, the recommender system gave recommendations based on the social influence that users gain in the network. Thus, it managed the data sparsity caused by the cold-start problem.

Experiments were conducted to validate the performance of the proposed method. The results indicated that the use of social influence in the recommender system solved the problems caused by cold-start users. Utilizing friendship alone, however, was less helpful since its choices for recommendation were limited. However, it was difficult to determine the differences between the relationships, except for "friendship" and "neighbor." Therefore, despite the lack of information about trust, the recommender system accurately predicted user preference by "friendship" and "neighbor."

Acknowledgments

The research is based on work supported by Taiwan Ministry of Science and Technology under Grant No. MOST 103-2410-H-006-055-MY3.

References

[1] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining, in: In 7th Language Resources and Evaluation Conference, 2010, pp. 2200–2204.
[2] J. Bobadilla, F. Ortega, A. Hernando, J. Bernal, A collaborative filtering approach to mitigate the new user cold start problem, Knowl.-Based Syst. 26 (2012) 225–238. http://www.sciencedirect.com/science/article/pii/S0950705111001882.
[3] BrightLocal, Local consumer review survey 2015 https://www.brightlocal.com/learn/local-consumer-review-survey-2015/, 2015.
[4] E. Cambria, B. Schuller, Y. Xia, C. Havasi, New avenues in opinion mining and sentiment analysis, Intell. Syst. IEEE 28 (2) (2013) 15–21.
[5] C. Chen, Y. Wan, M. Chung, Y. Sun, An effective recommendation method for cold start new users using trust and distrust networks, Inf. Sci. 224 (2013) 19–36. https://www.sciencedirect.com/science/article/pii/S0020025512007074.
[6] K. Choi, D. Yoo, G. Kim, Y. Suh, A hybrid online-product recommendation system: combining implicit rating-based collaborative filtering and sequential pattern analysis, Electron. Commerce Res. Appl. 11 (4) (2012) 309–317. http://www.sciencedirect.com/science/article/pii/S156742231200018X.
[7] R. Coban, A context layered locally recurrent neural network for dynamic system identification, Eng. Appl. Artif. Intell. 26 (1) (2012) 241–250. https://www.sciencedirect.com/science/article/pii/S0952197612002618.
[8] E. Dragut, C. Fellbaum, The role of adverbs in sentiment analysis, in: *Proceedings of* Frame Semantics *in nlp: A Workshop in* Honor *of* Chuck Fillmore *(1929–2014)*, Baltimore, MD, USA, 2014, pp. 38–41. http://www.aclweb.org/anthology/W14-3010.
[9] H.J. Escalante, M.A. García-Limón, A. Morales-Reyes, M. Graff, M.M. y Gómez, E.F. Morales, J. Martínez-Carranza, Term-weighting learning via genetic programming for text classification, Knowl.-Based Syst. 83 (2015) 176–189. http://www.sciencedirect.com/science/article/pii/S0950705115001197.
[10] M.A. Fattah, New term weighting schemes with combination of multiple classifiers for sentiment analysis, Neurocomputing 167 (2015). http://www.sciencedirect.com/science/article/pii/S0925231215005202.
[11] R. Forsati, M. Shamsfard, Hybrid pos-tagging: a cooperation of evolutionary and statistical approaches, Appl. Math. Modell. 38 (13) (2014), http://www.sciencedirect.com/science/article/pii/S0307904×1300783X.
[12] J. Golbeck, Chapter 6 – trust, in: J. Golbeck (Ed.), Analyzing the Social Web, Morgan Kaufmann, Boston, 2013, pp. 75–89. https://www.sciencedirect.com/science/article/pii/B9780124055315000067.
[13] G. Guo, J. Zhang, D. Thalmann, Merging trust in collaborative filtering to alleviate data sparsity and cold start, Knowl.-Based Syst. 57 (2014) 57–68. http://www.sciencedirect.com/science/article/pii/S0950705113003870.

[14] G. Guo, J. Zhang, Y. Neil, Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems, Knowl.-Based Syst. 74 (2015) 14–27. https://www.sciencedirect.com/science/article/pii/S0950705114003918.
[15] S. Kara, Özgür Alan, O. Sabuncu, S. Akpınar, N.K. Cicekli, F.N. Alpaslan, An ontology-based retrieval system using semantic indexing, Inf. Syst. 37 (4) (2012) 294–305. http://www.sciencedirect.com/science/article/pii/S030643791100113X.
[16] G. Linden, B. Smith, J. York, Amazon.com recommendations: item-to-item collaborative filtering, IEEE Internet Comput. 7 (1) (2003) 76–80.
[17] B. Lis, In ewom we trust, Bus. Inf. Syst. Eng. 5 (3) (2013) 129–140.
[18] B. Lu, M. Ott, C. Cardie, B.K. Tsou, Multi-aspect sentiment analysis with topic models, in: In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, Washington, DC, USA, 2011, pp. 81–88.
[19] R. Navigli, S.P. Ponzetto, BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, Artif. Intell. 193 (2012). https://www.sciencedirect.com/science/article/pii/S0004370212000793.
[20] R. Navigli, BabelNet. Available at: http://babelnet.org/
[21] A.G. Nikolaev, R. Razib, A. Kucheriya, On efficient use of entropy centrality for social network analysis and community detection, Social Networks 40 (2015). http://www.sciencedirect.com/science/article/pii/S0378873314000550.
[22] R.Ç. Özden Erçin, Identification of linear dynamic systems using the artificial bee colony algorithm, Turkish J. Electr. Eng. Comput. Sci. 20 (Supl. 1) (2012) 1175–1188.
[23] T. Özyer, J. Rokne, G. Wagner, A.H. Reuser, The Influence of Technology on Social Network Analysis and Mining, Springer-Verlag Wien, 2013.
[24] X. Qian, H. Feng, G. Zhao, T. Mei, Personalized recommendation combining user interest and social circle, IEEE Trans. Knowl. Data Eng. 26 (7) (2014) 1763–1777.
[25] C. Quan, F. Ren, Unsupervised product feature extraction for feature-oriented opinion determination, Inf. Sci. 272 (2014) 16–28. http://www.sciencedirect.com/science/article/pii/S0020025514001698.
[26] Y.-S. Seol, D.-J. Kim, H.-W. Kim, Emotion recognition from text using knowledge based ANN, in: ITC-CSCC: 2008, 2008, pp. 1569–1572.
[27] O. Shafiq, R. Alhajj, J.G. Rokne, On personalizing web search using social network analysis, Inf. Sci. 314 (2015) 55–76. http://www.sciencedirect.com/science/article/pii/S0020025515001292.
[28] P. Sinha, N.M. Veyie, B.S. Purkayastha, Enhancing the performance of part of speech tagging of nepali language through hybrid approach, Int. J. Emerg. Technol. Adv. Eng. 5 (5) (2015) 354–359. http://www.ijetae.com/files/Volume5Issue5/IJETAE_0515_61.
[29] L.H. Son, Dealing with the new user cold-start problem in recommended systems: a comparative review, Inf. Syst. 58 (2016) 87–104. http://www.sciencedirect.com/science/article/pii/S0306437914001525.
[30] T.T. Thuan, S. Puntheeranurak, Hybrid recommended system with review helpfulness features, in: In Tencon 2014–2014 IEEE Region 10 Conference, 2014, pp. 1–5.
[31] K. Toutanova, D. Klein, C.D. Manning, Y. Singer, Feature-richpart-of-speech tagging with a cyclic dependency network, in: In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Stroudsburg, PA, USA, 2003, pp. 173–180.
[32] Princeton University, About wordnet available at: http://wordnet.princeton.edu, 2010.
[33] Yelp, Yelp dataset challenge available at: http://www.yelp.com/dataset_challenge, 2015.
[34] T. Yuan, J. Cheng, X. Zhang, Q. Liu, H. Lu, How friends affect user behaviors? An exploration of social relation analysis for recommendation, Knowl.-Based Syst. 88 (2015) 70–84. http://www.sciencedirect.com/science/article/pii/S0950705115003093.
[35] J. Zhang, Q. Peng, S. Sun, C. Liu, Collaborative filtering recommendation algorithm based on user preference derived from item domain features, Physica A 396 (2014) 66–76. http://www.sciencedirect.com/science/article/pii/S0378437113010558.