IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 23, 2018, accepted June 27, 2018, date of publication July 13, 2018, date of current version August 20, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2855690

Using Heterogeneous Social Media as Auxiliary Information to Improve Hotel Recommendation Performance

JUI-HUNG CHANG[1], CHEN-EN TSAI[2], AND JUNG-HSIEN CHIANG[2]
[1]Computer and Network Center, Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan
[2]Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan

Corresponding author: Jung-Hsien Chiang (jchiang@mail.ncku.edu.tw)

ABSTRACT Research conducted on social media is currently increasing. Information obtained by users of social media has resulted in the development of many recommendation systems that analyze user preferences in an attempt to locate the most suitable products for recommendation to users. A great number of people have used social media platforms, such as Twitter, to develop a hotel recommendation system. Here, we propose a Twitter-based recommendation system via the aid of heterogeneous social media. First, a model is designed to predict user preferences by improving matrix factorization based on user preferences and users' personal data, where basic information about hotels collected from Yelp is used as auxiliary information. On the other hand, an analytical user posting behavior algorithm is created for establishing users' posting behavior vectors based on earlier posts in Twitter and Yelp. This results of the experiments show that the proposed method can improve accuracy by 30% in terms of RECALL compared with the Twitter-based recommendation system without the use of heterogeneous social media. Furthermore, it can improve the accuracy of the mean reciprocal rank by 80% and can increase precision by as much as 100%.

INDEX TERMS Analytical user posting behavior algorithm, hotel recommendation system, mean reciprocal rank.

I. INTRODUCTION

The rapid development of social media has accelerated user exchanges in terms of messages and comments [1]. The goal of this paper is to establish a system for analyzing user opinion data in order to recommend target items [2]. Such systems are called recommendation systems.

Social media consists of many heterogeneous community networks, such as Facebook, community friend networks, and community networks. A great amount of research on this topic has used heterogeneous community networks to improve efficiency. This paper puts social media in the spotlight since it is at a higher scale than social community networks. The drawbacks of recommendation systems come from the characteristics of social media itself. They include the length of the posts, the accuracy of the information provided by users, and the detailed media classification methods. These superposed problems result in inaccurate recommendations. Social media is divided into two main categories. The first system, like Twitter and Facebook, omits preferences for the target objects. The second system, like Yelp, shows target objects with user preferences [3]. These two categories demonstrate the heterogeneous aspects of social media. Heterogeneous social media refers to social media with different characteristics. Consider the length of posts, for example. Users can use long sentences in Yelp according to the rules. In contrast, the tweets on Twitter are short due to the word count limitation. As a result, Twitter provides limited, somewhat trivial information only, leading to its inferiority to Yelp as a recommendation system [4], [5].

This paper focuses on all users who need recommendations to build a system. A great number of people use the first category of social media by using platforms such as Twitter. Thus, a hotel recommendation system is developed based on studying Twitter as the study object. However, the heavy noise on Twitter has led to the inefficient performance of the Twitter recommendation system. Hence, we add Yelp, which is categorized as heterogeneous social media, as auxiliary information to improve the system. It is proposed that the inaccuracy of Twitter mentioned above can thus be resolved by adding appropriate heterogeneous social media.

First, a model is designed to predict user preferences. By improving the matrix factorization, an algorithm is designed for supplementary information transformation (SITA). This algorithm establishes a matrix based on user preferences, users' personal data, and basic information on the hotels collected from Yelp and used as auxiliary information. Then, this auxiliary information is used to improve the efficiency of the recommendation system. An analytical user posting behavior algorithm (APBA) is also developed. The function of the APBA is to establish users' posting behavior vectors (Vpb) based on earlier posts in Twitter and Yelp. Finally, the users' Vpbs and the Twitter community network are applied to construct an algorithm to obtain a Twitter recommendation list based on supplementary information (TRSIA). It is expected to predict user preferences for the target hotels. In this recommendation model, the user reference relationship (URR) is modelled by allocative techniques and Vpb. Every URR represents a relative similar habit between the Twitter user and the Yelp user. Through URR, we enhance the efficiency of Yelp by adding auxiliary information into the system.

This paper is an attempt to prove that a Twitter-based recommendation system, via the aid of heterogeneous social media, can boost accuracy by 30% in terms of RECALL compared to that of the heterogeneous social media used in the Twitter-based recommendation system. The MRR can improve be improved by 80%. The precision can be increased by as much as 100%. From the discussion above, it is confirmed that recommendation systems such as Twitter can be boosted greatly by adding auxiliary information as established by social media.

This dissertation is organized as follows: Chapter I introduces the motivation and the purpose of our study. Chapter II focuses on related studies and methods for our recommendation system. Chapter III includes a discussion of our methods and the data collection. Chapter IV provides a case study intended to evaluate the performance of our system. Chapter V offers conclusions and suggestions for future work.

II. RELATED WORK

In this Chapter, we review current recommendation systems and illustrate the role of heterogeneous networks.

A. RECOMMENDATION SYSTEMS BASED ON SOCIAL MEDIA

Social Media offers information diffusion platforms for sending and receiving messages. There have been a lot of attempts to extract information from social media; thus most recommendation systems are built on social media [6], [7]. Recently, collaborative filtering has been widely utilized in the field of computer science. Recommendation systems mainly use collaborative filtering methods to build user rating matrices. The applicable methods include: (1) the social network-based method, (2) the content-based method, and (3) the hybrid method.

1) SOCIAL NETWORK-BASED METHOD

People and information are two core dimensions of social networks. Recommendation systems for social media usually regard the people dimension as more important. For example, they value the understanding of friendship and followership. The number of available/applicable social networks is a critical factor for building a recommendation system. Bhattacharya proposed a mechanism to extract topics of interest from individual users in the Twitter social network. Hannon *et al.* [8] built a followee recommender system for Twitter using relationships in Twitter social graphs.

2) CONTENT-BASED METHOD

Content based methods recommend items or products to users upon analyzing user content. User content reflects a user's sentiments and ideas. Thus, content-based methods are expected to acquire user performance by analyzing such content. Kim and Shim [9] proposed a recommendation system for Twitter using probabilistic modeling based on a Latent Dirichlet allocation model. It recommends top-K users and top-K tweets to a user [9]. There are hashtag-based, entity-based, and topic-based user profiles built on user content. These include analyzing special hashtags and knowing the content, entity, and topic. These methods can benefit from semantic text enrichment using text mining. For example, Abel *et al.* [10] proposed an analysis of Twitter activities for user modeling by evaluating the quality of user models in the context of new articles recommended, which is an example of how semantic enrichment can enhance the quality of user profiles. Abel *et al.* [11] investigated user modeling strategies for inferring/extracting personal interest profiles from user interactions and also compared different strategies for creating user profiles based on Twitter messages.

3) HYBRID METHOD

The hybrid method incorporates elements of both the social network- and content-based methods. For example, Yang *et al.* [12] proposed a user preference model for hotel selection with extra information, such as a social network that was produced based on users' check-in information and text-based tips which were processed using sentiment analysis techniques. Bostandjiev *et al.* [5] proposed an interactive hybrid recommendation system that generates music predictions from multiple social and semantic web resources, such as multiple social media including Facebook and Twitter and semantic web resources including Wikipedia. Hannon et al. proposed a follow recommendation hybrid-based method that focuses on the creation of relationships between users and the social network of a user. They also evaluated a range of different user profiling and content-based and social recommendation strategies to determine the recommendation efficiency of the system. [13]. Chen et al. proposed a method of making tweet recommendations based on collaborative ranking to capture personal interests. Our recommendation system in this work is hybrid. It uses user content to build the URR. Then, it uses the Twitter friend relationship social network and the URR to build a Twitter rating matrix.

B. COLLABORATIVE FILTERING IN RECOMMENDATION SYSTEM

Recently, collaborative filtering has been regularly used in recommendation systems. We improve on one of the collaborative filtering frameworks, matrix factorization, to build our model. The original matrix factorization framework is used to predict unknown values in the rating matrix according to known values and mathematic formulae. The matrix factorization that we provide refers to information on social influence used as a reference to predict unknown values in the rating matrix. Chen *et al.* [14] proposed a matrix factorization method for item-level social influence modeling and devised a projected gradient method to solve the weakness related to the fact that traditional matrix factorization cannot refer to social influence.

Collaborative filtering may be user-based, item-based, or hybrid-based. User-based CF recommends items to a user based on the preferences of its ''neighbor'' users. The user-based CF first determines a group of neighbors for each user to see if they have the same behavior. Then, it predicts the chosen user's ratings by the ratings of the neighbors. Herlocker *et al.* [15] reviewed key methods for evaluating user-based CF recommendation systems. The method includes a comparison of using one kind of dataset with different CF algorithms to that of using a user-based CF algorithm for the experiment [15]. The item-based CF methods determine a set of items for each user that the user likes. The item-based CF methods then recommend new items to a user, which are similar to those for which the user showed a preference. Sarwar *et al.* [16] evaluated item-based algorithms and found dramatically better performance than that for user-based algorithms. Bostandjiev *et al.* [5] proposed an interactive hybrid recommendation system that generates item predictions from multiple social and semantic web resources, such as Facebook, Twitter, and Wikipedia . The disadvantage of the user-based, item-based, and hybrid-based methods is that their performance suffer from social influence. Too much noise related to social influence leads to the drawbacks during use. Thus, we use another type of social media as auxiliary information to boost the efficiency. This type of social media can minimize the noise issue. However, when setting Yelp as our auxiliary information, we designed our own hybrid matrix factorization. The hybrid factorization can be referred to as the matrix factorization of social influence.

A diversity of social networks are available for recommendation systems in the current social media. Friend social networks and fan social networks are two examples of heterogeneous networks. These social networks are used because they can improve efficiency. For example, Zhang *et al.* [17] proposed a friend recommendation that can recommend some friends to user, and used many heterogeneous networks from social networks to resolve friend link problems and to choose suitable heterogeneous networks to deal with friend link resolutions. Yu *et al.* [18] aimed at providing high-quality personalized recommendations using different types of heterogeneous networks from information social networks. For example, they used a heterogeneous network based on movies and users. Social network based on movie information, including actors, directors, release date. Social network based on users' information about seen movies, the watching time. Zhu decomposed the information social network into several homogeneous networks (such as an image tag network and a video tag network) and proposed a social relevance learning method for images [19].

A Mi and Lee (2015) proposed an architecture recommendation system that allows advertisers to provide users with personalized hotel recommendations, as based on various client data associated with the user. The client data describes the user's hotel reservation preferences, which matches the hotel property data and is used to describe the characteristics of a specific hotel. Then, this system generates hotel preference based on hotel data, describing the characteristics of particular hotel offers. The recommendation system selects a specific hotel preference to be recommended to the user by comparing the hotel property data with the client data associated with the user, and provides recommendations to the user based on the quotation of the hotel selected [20].

Lin *et al.* [21] (2015) used the browsing behaviors of the user to read hotel reviews on mobile devices, and then, applied Text Mining technology to construct user interest files to provide personalized hotel recommendations. This system enables the user to search for hotels and browse hotel reviews, and each gesture performed by the user on the touch screen is recorded when reading hotel reviews. Then, a paragraph of hotel reviews, which the user had shown interest in, is identified based on the gestures the user had performed. Text Mining Technology is used to construct user interest files based on the review contents that users carefully read [21].

This Recommendation system (RSs) has been successfully applied to alleviate information overload problems and help users to make decisions. This multi-standard recommendation system is one of the RSs applied to predict user preferences by using users' multiple ratings of different aspects of the project (i.e. multi-standard ratings). Zheng (2017) proposed a novel method regarding the standard preference as a certain scenario. More specifically, this study believes that a portion of multiple standard preferences could be considered as the context, while others could deal with it in the traditional manner of a multi-standard recommendation system. This article compares the recommendation performances of three settings: using all standard scores in a traditional manner; regarding all standard preferences as scenarios; and using the selected standard scores as scenarios. The experiments in this article are conducted on the basis of two rating datasets in the real world. The experimental results show that treating the standard preference as a background could improve the effectiveness of project recommendations; however, it should be selected with care. A hybrid model using selected standard preferences as the background, and using the traditional methods of residual standard, is eventually proved to be the overall winner in this experiment [22].

Fukumoto *et al.* [23] (2015) introduced a method of collaborative filtering for hotel recommendations, which combine guest preferences. They used the results of an aspect-based sentiment analysis to recommend hotels; whether a hotel can be recommended depends on the guest's preferences related to the hotel aspect. For each aspect of the hotel, they used the triples extracted from the guest reviews to identify guest preferences. The triples represent the relationship between available aspects and their preferences. They used the positive/negative preference in certain aspects to calculate transitive associations between hotels. Finally, the hotels were scored using the Markov Random Walk model to explore the transitive associations between hotels. The empirical evaluation shows that the aspect-based sentiment analysis improved the overall performance. In addition, they found it to be effective, and found hotels that had never been lived in, but shared the same community [23].

In most business organizations, knowledge sharing is often lacking when it comes to the success of business systems. Hu *et al.* [24] (2016) investigated the factors affecting the success of Saudi business systems, where the data came from a private organization in Saudi Arabia, which was analyzed using the Partial Least Square method. The results showed that the organization culture affected the success of knowledge sharing for business systems. In addition, both intrinsic motivation and perceived usefulness had positive impact on the success of the business system, which showed that the success of the business systems was based on the concepts of knowledge sharing and user motivation [24].

The ubiquitous hotel recommendation is a very popular location-aware service type. However, the existing recommendation system has several problems. Chen (2017) proposed a method of fuzzy-weighted-average (FWA) and backpropagation-network (BPN) to overcome the ubiquitous barriers for hotel recommendations, and thus, improve its effectiveness. FWA was used to assess the overall performance of the hotel. The construction of BPN was for the overall benefits of defuzzification. In addition, a personalized preference index is proposed to address the choice of travelers for the dominated hotels. Field research is conducted regarding the effectiveness of the proposed method in a small area in Seatwen District, Taichung City, Taiwan [25].

In our study, we use Twitter friend relationships and Twitter text. Yelp contains a lot of heterogeneous social networks, such as a friend social network, a fan social network, a hotel attributes social network, etc.

III. RECOMMENDATION SYSTEMS USING HETEROGENEOUS SOCIAL MEDIA

Our study is intended to capture and preprocess user data from different social media with the ultimate goal of designing a method for creating a hotel recommendation system. The study is divided into three phases: (Fig. 1) Phase 1 is data collection, that is, collecting Twitter and Yelp data.

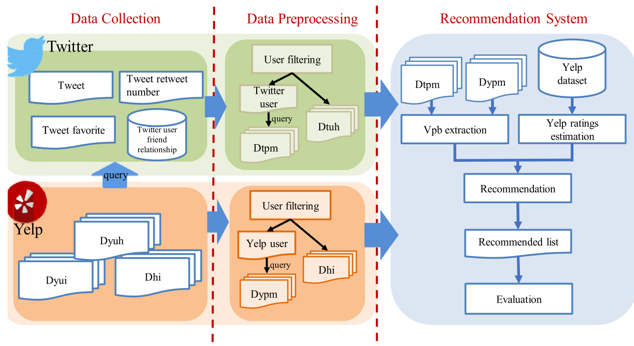


FIGURE 1. System architecture.

Phase 2 is data preprocessing, that is, setting a threshold for filtering data and capturing the user's historic postings. Phase 3 is the recommendation phase.

A. DATA COLLECTION

We assembled data from two different types of social media. The first is Twitter, the target media, and the other one is Yelp, a type of heterogeneous social media. In this paper, Twitter users are the research object, and Yelp provides auxiliary information for the hotel recommendation system (Fig. 2). Our data collection flow chart is displayed in Fig. 2. The initial Yelp data come from Yelp dataset challenge, the official data release for academic use. It contains entries for 85,901 local businesses, including hotels, restaurants, and so forth. They are spread over 4 countries and 11 cities. In the Yelp dataset challenge, every business has its attributes including check-ins, tips, and opening time. Similarly, every user has related background data such as friend social network, review count, and so forth. Data filtering is the next step. To begin with, we make "Hotels & Travel" our target. In the Yelp dataset challenge, there are a total of 731 hotels, which serve as our initial query. Hotel-related data are collected to form our Yelp dataset. This dataset is composed of three sections: users' reviews of hotels, the background of the users who leave comments, and the background of the hotels, abbreviated as "Dyuh," "Dyui," and "Dhi," respectively. Dyuh is a dataset which collects users' posts

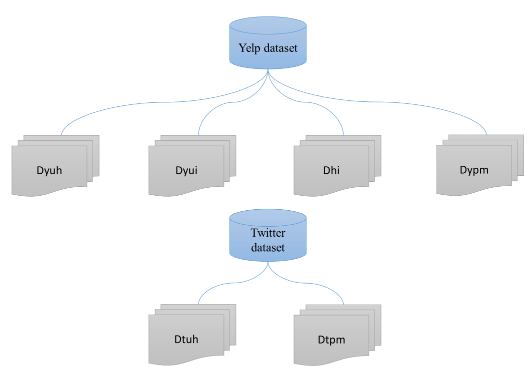


FIGURE 2. Data collection procedure.

to the hotels…Dyui is a dataset which collect users' information, such as, address, follow relationship etc. Dyui is a dataset which collect hotel information, such as opening hours, address, and so on.

In our Twitter dataset, the 731 hotels mentioned in the previous step make up the initial query. The tweets that mentioned these 731 hotels on Twitter from the period 2013-2015 form the dataset named "Dtuh."

B. DATA PREPROCESSING

There are two main points in this chapter: (1) preprocessing Twitter and Yelp data to reduce data sparsity and (2) establishing a user behavior dataset. Initially, most Twitter and Yelp users commented on only one hotel among the 731; data sparsity will thus be solved through data preprocessing. To create our "user behavior dataset" (i.e., Vpb extraction), we assembled all the past posts of the users from 2013 to 2015.

C. DATASET PREPROCESSING

For the Yelp dataset filtering, Table 1 is based on Dyu. Since most Yelp users only commented on one or two hotels of the 731 hotels, we set a threshold of 6 on hotel numbers to filter the users (Fig. 3) After the update, the number of Yelp users went down to 1,617.

TABLE 1. Statistics of yelp user number.

threshold	user number	hotel review number/ user number	standard deviation
1	70,893	1.52	1.837
2	16,232	3.29	3.27
3	6,965	5.01	4.446
4	3,908	6.58	5.44
5	2,410	8.2	6.428
6	1,617	9.74	7.362

. https://www.yelp.com/dataset_challenge

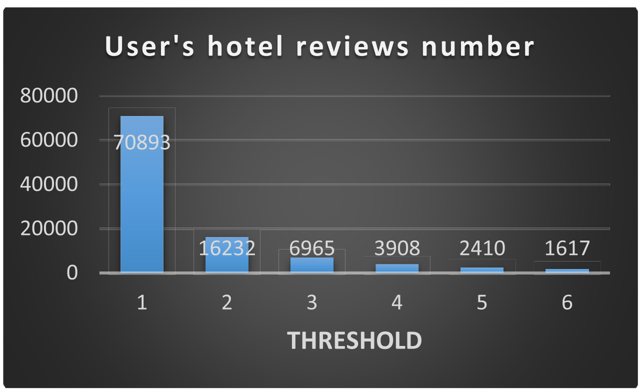


FIGURE 3. Number of user hotel reviews.

For Twitter dataset filtering, Table 2 is based on Twitter user hotels. Since most Twitter users only commented on one or two hotels, we set a threshold of 11 to filter the users.

TABLE 2. The statistics of twitter user number.

threshold	user number	hotel review number/ user number
1	70,893	1.52
11	2,484	3.29

After the update, the number of Twitter users went down to 2,485.

D. USER BEHAVIOR DATASET

After dataset preprocessing, we had a total of 1,617 Yelp users and 2,485 Twitter users left. The past posts were analyzed again for Vpb extraction. These behavior datasets were collected. We stored the collected reviews from the 1,617 Yelp users in "Dypm." The 2,485 Twitter users' tweets between 2013-2015 were saved in "Dtpm."

E. RECOMMENDATION ARCHITECTURE

The system is composed of three parts: Vpb extraction, Yelp ratings estimation, and recommendation. In the Vpb extraction, we predict each user's Vpb according to their past posting behavior. For Yelp rating estimations, we produce Msi, the auxiliary information established by the heterogeneous social media. In recommendation, we use the Vpb to make URRs. The URR is the communication bridge between Yelp and Twitter. It is what we call auxiliary information for Twitter users. The auxiliary information and Twitter's friend social network are combined to make the Twitter rating matrix, from which the Twitter recommended list can be established.

All abbreviations and the full expressions are shown in Table 3 for the convenience of readers.

TABLE 3. Noun definition.

Abbreviation	Full Name
SITA	Algorithm for supplementary information Transformation.
Myi	Yelp user information matrix.
Mhi	Yelp hotel information matrix.
Msi	Supplementary information matrix
APBA	Analytical user posting behavior algorithm
Vpb	User posting behavior vector
Vtup	The Twitter data user posting behavior vector
Vyup	The Yelp data user posting behavior vector
TRSIA	Twitter recommended list with supplementary information algorithm.
URR	User reference relationship.
RVism	Rating vector based on the information from social media.
RVug	user gone rating vector.
RVsi	Rating vector based on supplementary information.

F. YELP RATING ESTIMATION

For the Yelp rating estimation, we compile Msi, the auxiliary information from heterogeneous social media (Fig. 5). The Yelp dataset comprises Dyuh, Dyui, and Dhi. Our algorithm,

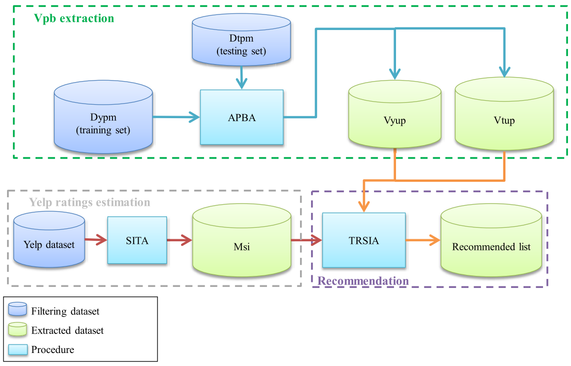


FIGURE 4. Overview of recommendation architecture in our system.

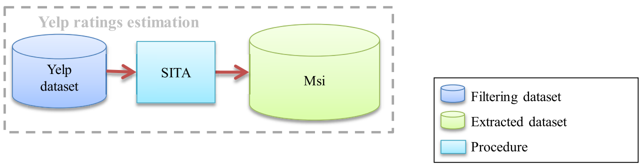
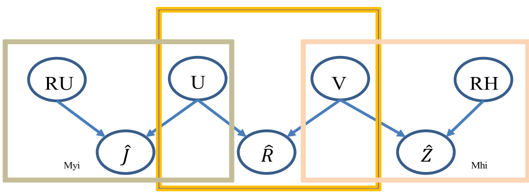


FIGURE 5. The architecture of the Yelp ratings estimation.

SITA, and the Yelp dataset yield the Msi. Figure 6 displays the SITA construction. This algorithm is composed of three parts: (1) making Myi from Dyui, (2) making Mhi from Dhi, and (3) making Msi from Mhi, Myi, and SITA



$\hat{R}$ is user to hotel matrix
U is user to latent matrix
V is latent to hotel matrix
$\hat{J}$ is user to user matrix
RU is latent to user matrix
$\hat{Z}$ is hotel to hotel matrix
RH is hotel to latent matrix

FIGURE 6. SITA.

G. THE MYI ESTABLISHMENT

We use Dyui and Dyuh to create Myi, which refers to the target $\hat{J}$ matrix in Figure 6. Figure 7 is an example for Myi, which is a matrix. To create Myi, we use two feature extraction methods. One is the BOW method made using the TweetTokenizertool.[1] Every Yelp user in Dyuh has its own assembly. This assembly exists in all the reviews of this user from the Yelp dataset challenge. We use the TweetTokenizertool to turn reviews into tokens. By tokens we mean the user features extracted/collected by BOW. The second method, which is called the other information method, as designed for this work, fetches features from Dyui. The features include

[1] http://www.cs.cmu.edu/~ark/TweetNLP/

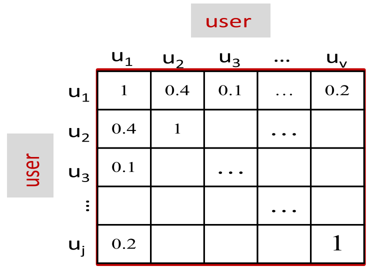


FIGURE 7. The sample of Myi.

the residential area of the users and names of restaurants, to name a few. We use KLD similarity, which is one of the formula used to calculate similarity. The similarity is calculated to reach the inner values of Myi. The user features come from the two feature extraction methods mentioned above. Finally, Myi is established.

H. THE ESTABLISHMENT OF MHI

We use Dyui and Dyuh to create Mhi, the target $\hat{Z}$ matrix in Figure 7. To create Mhi, we have to fetch the features of every hotel. To this end, we use two feature extraction methods. One is the BOW method, made by TweetTokenizertool. Every hotel in Dyuh has its own assembly. This assembly exists in all the reviews of this hotel from the Yelp dataset challenge. We use the TweetTokenizertool to turn reviews into tokens. By tokens we mean the hotel features we collected using the BOW method. In the second method, We directly fetch the hotel features such as, opening time, vacancy... etc. We use KLD similarity, which is one of the formulae used to calculate similarity. The similarity is calculated to obtain the inner values of Myi. The user features come from the two feature extraction methods mentioned above. Finally, Myi is established.

I. THE MSI ESTABLISHMENT

We create the Yelp rating matrix in this section. The Yelp rating matrix is established through predicting the initial Yelp rating matrix constructed from the unknown value in Dyuh. First, we use the three known matrices, Myi, Mhi, and the initial Yelp rating matrix. These three matrices correspond to J, Z, and R. Through SITA and these three matrices, J, Z, and R, respectively, we are able to make the Yelp rating matrix. In SITA, there are three matrices to be predicted, namely $\hat{j}$, $\hat{R}$, and $\hat{Z}$. These $\hat{j}$, $\hat{R}$, and $\hat{Z}$ matrices aim at J, Z, and R in order to predict the inner unknown value(s) of $\hat{R}$. After training SITA, we get the Yelp rating matrix, namely Msi, which is $\hat{R}$.

Next, we introduce our SITA algorithm.

SITA uses matrix factorization, a powerful collaborative filtering method. Matrix factorization doesn't require any information about items or users. It only requires an algorithm and the known values in the initial rating matrix to create a complete rating matrix. An example is shown in Figure 8. In this example, R is the initial matrix. Matrix factorization

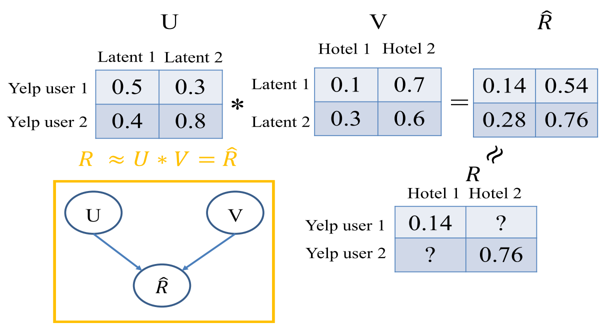


FIGURE 8. The example of matrix factorization.

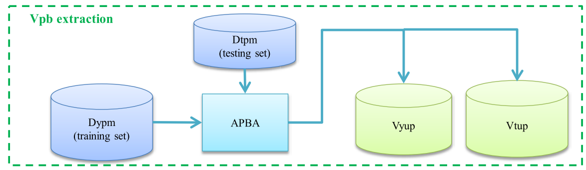


FIGURE 9. The architecture of the Vpb extraction.

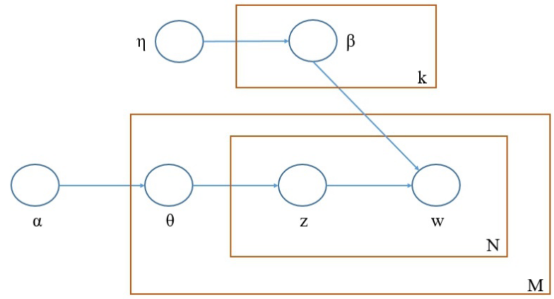


FIGURE 10. Graphical model of the APBA.

produces a U and a V matrix from the known value in the R matrix. Based on the multiplication of the U matrix and V matrix, the $\hat{R}$ matrix is used to predict the unknown value in the R matrix.

Matrix factorization does not require any reference from the users or the items. Actually, when essential information is included in the users or items, matrix factorization produces worse results than other collaborative filtering methods. Our SITA algorithm overcomes this drawback via a rating matrix. SITA is attached in the Appendix. In this section, we present the Yelp rating matrix according to the SITA, Myi, and Mhi. The $\hat{J}$ matrix is the product of U and RU; the $\hat{R}$ matrix is the product of RH and V, and the $\hat{Z}$ matrix is the product of U and V. The loss function for SITA is:

$$\begin{aligned}\text{Loss function} = &(R-UV)^2+(J-U*RU)^2\\&+(Z-RH*V)^2\\&+\beta\left(\|U\|^2+\|V\|^2+\|RU\|^2+\|RH\|^2\right).\end{aligned}$$

The R denotes the initial Yelp rating matrix. J denotes Myi. Z denotes Mhi.

We train SITA through the loss function. After training SITA, we obtain the target Yelp rating matrix $\hat{R}$, which is the $\hat{R}$ product of U and V.

Msi is completed when the Yelp rating estimations are completed.

J. VPB EXTRACTION

User's posting data are input to predict Vpb via the Latent Dirichlet allocation (LDA). Figure 10 displays the Vpb extraction, which consists of two parts: the training set and the testing set.

In the training set, we input Dypm. Every Yelp user has a document that contains that user's past posts.

These documents are used as inputs to edit the LDA and train the APBA, as shown in Figure 10. The purpose of the APBA is to acquire the document topics, user topic distribution, and the per-topic word distribution, etc.

Here, α denotes the parameter of the Dirichlet prior to the per-document topic distribution(s); β denotes the parameter of the Dirichlet prior to the per-topic word distribution. In addition, $\boldsymbol{\theta}$ denotes the topic distribution for user m; η denotes the word distribution for topic k, and z denotes the topic for the n-th word for user m. Finally, w denotes a specific word; k denotes a topic; N denotes a word, and M denotes user posting behavior data.

After APBA has been trained, every Yelp user has a topic distribution. Then, we look upon the topic distribution as Vpb. Figure 11 I is a demonstration of a Yelp user's Vpb, where Vyup is simply the collection of all Vpb vectors.

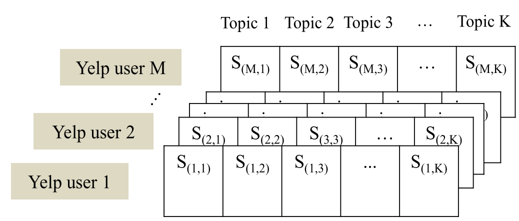


FIGURE 11. Vpb of Yelp users.

Here, m is the serial number of a Yelp user, and k is the serial number of a topic. In addition, S (m, k) denotes the score for the Kth topic by the Mth user.

In the testing stage of the Vpb extraction, we input Dtpm. We use Dtpm and the trained APBA to predict every Twitter user's topic distribution. Then, we consider these topic distributions to be the Vpb and collect all the Vpb vectors to form Vtup.

K. RECOMMENDATION

We create the recommendation list for Twitter users in this section. First, we form a communication bridge between Yelp and Twitter users via Vtup and Vyup. Then, we design an algorithm to make Twitter user rating vectors using Msi, URR, and the friend social network from Twitter itself.

Finally, we use the Twitter user rating vectors to create the recommendation list for each user. Figure 12 displays the corresponding flow chart. The combined model is composed of three phases: (1) Establishing the URR, (2) building Twitter user rating vectors, and (3) building the recommendation list.

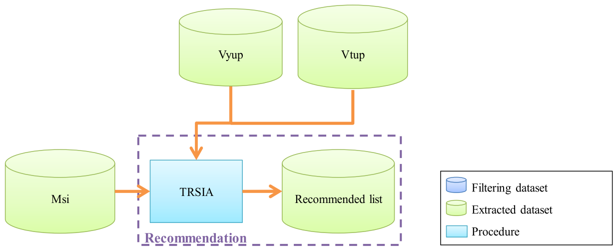


FIGURE 12. Recommendation architecture.

L. URR ESTABLISHMENT

In this section, we introduce how to create URR with Vtup and Vyup. Figure 13 displays one part of TRSIA.

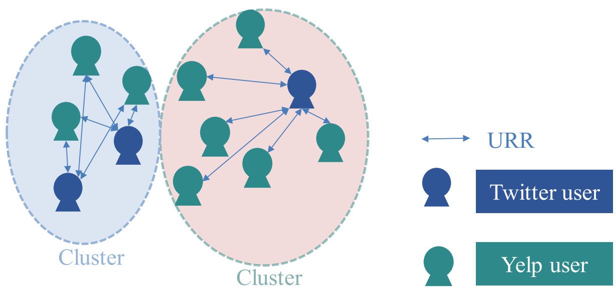


FIGURE 13. Example of URR establishment.

To begin with, we input Vtup and Vyup. Every Yelp user and Twitter user has a Vpb. We use K-mean clustering to classify the users. The best result are obtained when the users are divided into 70 groups. K-mean clustering is used because when a Yelp user and a Twitter user have the same posting habits, they can be classified into the same group.

In the second step, we create the relations between the Twitter users and the Yelp users who are classified in the same group. Their relationship is recorded according to the URR (Fig. 13).

By creating a significant amount of URR records, more auxiliary information is shared from Yelp to Twitter users. Normally, there are few relationships between Yelp and Twitter in the real life. If a person holds IDs for both Yelp and Twitter, this ID will form a relation called "same user relation."

M. TWITTER USER RATING VECTOR ESTABLISHMENT

In this section, we construct the Twitter user rating matrix based on Msi, URR, and the friend social network from Twitter itself. Figure 14 illustrates a friend social network URR chart for the current system.

In Figure 14, note denotes a use, side denotes a friend relationship, sky blue objects represent data from Yelp, orange

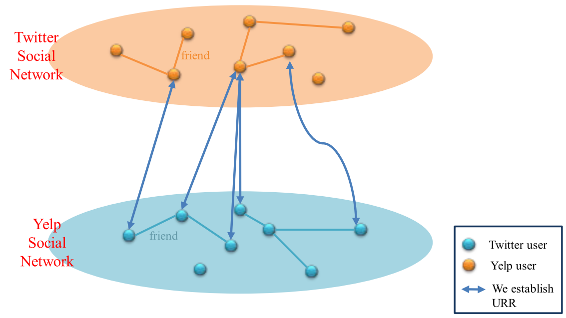


FIGURE 14. The example consists of Yelp friend social network, Twitter friend social network, and the URR.

objects are from Twitter and the blue objects denote the URR. The formula for making the Twitter user rating vector is composed of two components: (1) the RVsi of a Twitter user and (2) the RVism (the rating vector for every Twitter user).

$$S\,(u) = \alpha^{*} S_e\,(u) + (1-\alpha) * S_l\,(u)\,. \tag{1}$$

Here, S(u) denotes the rating vector for Twitter user u;$S_e\,(u)$ denotes the RVsi for Twitter user u, and $S_l\,(u)$ denotes the RVism for Twitter user u. In addition, α denotes the weight controlling the performance of these two components: the RVsi of a Twitter user and the RVism of a Twitter user.

1) RVSI OF A TWITTER USER

In this section, we use the Yelp rating matrix and the URR to design the algorithm for making a rating vector for every Twitter user. We name the group of rating vectors "RVsi," as shown in Equation 3-1. Figure 15 is an example of how to establish the RVsi. In Figure 15, we can see that Twitter user 1 has three URRs: one between Twitter user 1 and Yelp user 1, one between Twitter user 1 and Yelp user 2, and one between Twitter user 1 and Yelp user 3. Then, we use the rating vectors of Yelp user 1, Yelp user 2, and Yelp user 3 from Msi and Equation 3-1 to calculate Twitter user 1 的 RVsi. These rating vectors from Msi are below $S_p(Yi)$. Finally, we use cosine similarity to calculate the weight of $S_p(Yi)$. When the Vpbs of a Twitter user and a Yelp user are similar, the weight goes up, and vice versa. After normalization, the total weight of every Twitter user is 1. Based on the Msi and cosine similarity, we complete the formula as follows:

$$S_e\,(u) = \sum\nolimits_{i=1}^{k} \frac{Similarity(Tu, Yi)}{\sum\nolimits_{r=1}^{k} Similarity(Tu, Yr)} * S_p(Yi). \tag{III-1}$$

Here $S_p(Yi)$ denotes the rating vector for the Yelp user i from the Yelp rating matrix. Yi denotes Yelp user i, and Tu denotes Twitter user u. $Similarity(Tu, Yi)$ denotes the cosine similarity between the Vpb of Yelp user i and the Vpb of Twitter user u.

2) RVISM OF A TWITTER USER

In this section, we use the friend social network of Twitter itself to design an algorithm_$S_l\,(u)$ for making the rating vector for every Twitter user. We call this rating vector "RVism." $S_l\,(u)$ is shown in Equation 3-2. First, we extend the Twitter

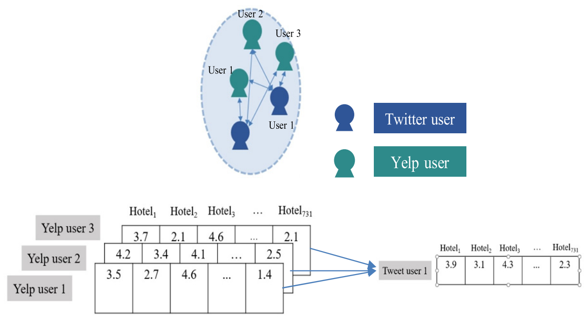


FIGURE 15. The example of RVsi.

friend social network. Since the network in Twitter is sparse, we design a rule to update the friend social network. The rule is to build a new friend relation between two users who have a mutual friend. Figure 16 is an updated example. In Figure 16, Twitter users u and u2 do not have any friend relationship, but they will have one after the update. Next, we establish a rating vector for every Twitter user. There are 731 rating vectors corresponding to 731 hotels. When users comment on a hotel, the corresponding number in the hotel index becomes 1. With no comments, the corresponding number in the hotel index becomes 0.This rating vector is called "RVug," which is also seen as the $S_r(T_k)$ in Equation 3-2. Finally, we use the new Twitter friend social network and the RVug to design a formula for $S_l(u)$ as follows:

$$S_l(u) = \sum_{i=1}^{k} \frac{1}{k} * [\beta * S_r(T_k)], \quad T_k \in friends. \quad \text{(III-2)}$$

Here $S_r(T_k)$ denotes the RVug for Twitter user K. T_k denotes the friends of Twitter user u, and k denotes the friend count of Twitter user u. In addition, β denotes the high score for heterogeneous social media. For instance, in Yelp, the highest rating for a hotel is 5. In this case, β will be 5. Different social media have different rating standards. β is used to standardize the evaluation.

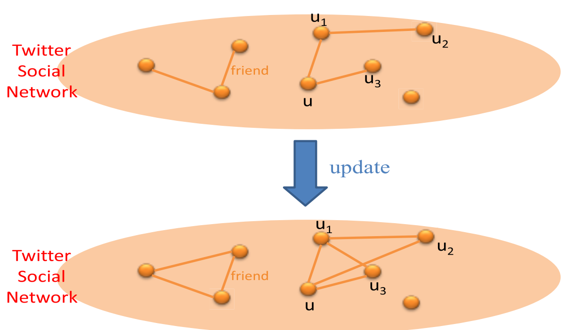


FIGURE 16. The example of updating Twitter friend social network.

N. RECOMMENDED LIST ESTABLISHMENT

By now, every Twitter user has a rating vector. Our system uses the rating vectors to make a recommended list. Figure 17 is a simple example. Sensitive is a term that

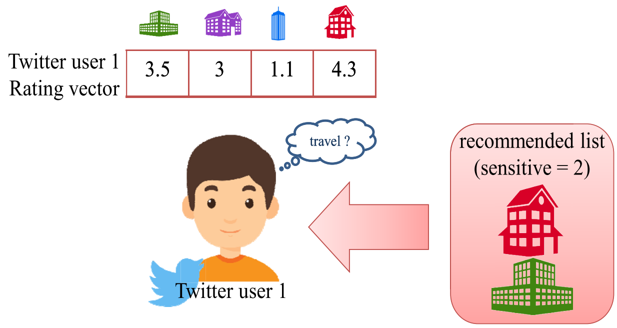


FIGURE 17. The example of a Twitter user recommended list.

represents the recommended hotel count according to the system. Using the rating vectors, the system predicts that the green and the red hotels are the top two preferred by Twitter user 1.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL DESIGN

We conducted two types of experiments: (1) URR-related and (2) recommended list-related. Through the first kind of experiments, we wanted to assess the validity of the establishment of the URR. Through the second kind of experiments, we wanted to assess the performance of our recommendation system. We used the mean reciprocal rank (MRR) of the recommended efficiency and the Recall to evaluate the performance of our recommendation system. The MRR and Recall are defined as follows:

$$\mathrm{MRR} = \frac{1}{\mathrm{m}} \sum_{\mathrm{j}=1}^{\mathrm{m}} \left(\frac{1}{\mathrm{rank}\left(\mathrm{u_j}, \mathrm{g}\right)} \right) |\mathrm{g} \in \mathrm{test}\left(\mathrm{u_j}\right) \quad \text{(IV-3)}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \quad \text{(IV-4)}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP}. \quad \text{(IV-5)}$$

In Equation 4-1, m denotes the number of Twitter users; $\mathrm{u_j}$ denotes Twitter user j, $\mathrm{rank}\left(\mathrm{u_j}, \mathrm{g}\right)$ denotes the ordering number of the first correctly recommended item for user j, and $\mathrm{test}\left(\mathrm{u_j}\right)$ denotes the recommended list for user j. In Equation 4-2, true positive (TP) denotes the number of correct items recommended by our system. False negative (FN) denotes the number of correct items missed by our system. In Equation 4-3, false positive (FP) denotes the number of wrong items recommended by our system.

We used the mean squared error (MSE) to calculate the error of our Yelp rating estimation as follows:

$$\mathrm{MSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(\tilde{Y}_i - Y_i\right)^2}{n}}.$$

Here $\tilde{Y}_i$ denotes the predicted value of a test set in Msi; Y_i denotes the real value of a test set in Msi, and n denotes the number of test values in Msi.

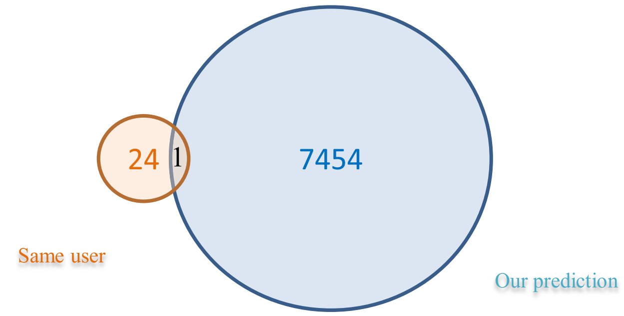


FIGURE 18. Venn diagram of URR with heterogeneous social media method.

B. EXPERIMENTS RELATED TO ESTABLISHING THE URR

We intended to prove the rationality of our URR concept in the first kind of experiments with realistic same user relations. When same user relation exists, the user in both Yelp and Twitter is the same user. With the same user relation, the information for the same user appearing simultaneously in Yelp and Twitter can both be used. Because they are in fact the same user, the posting habit can also be assumed to be similar. The URR can easily establish a relation that is same as that of the same user relation. Due to the two factors above, we marked the same user relation manually to verify the URR. In the Yelp dataset challenge, there is no information on the same user relation; hence we had to mark them manually. Upon confirming 25 same user relations, we confirmed 7,455 URRs. Figure 18 is the Venn diagram of our result.

C. EXPERIMENTS ON RECOMMENDED LISTS

We picked the best result using two similarity methods (KLD and Jaccard) and two feature extraction methods (BOW and hybrid) for the Yelp rating estimations (Myi and Mhi). Hence, MSE is the reference. We used the MSE to see which similarity method and feature method was better for training the SITA in order to optimize the SITA. We arrived at three combinations: (1) KLD + BOW, (2) Jaccard + BOW, (3) KLD + hybrid. The three results for the combinations are as follows:

In Figure 19, we display the best way to create Myi and Mhi through the KLD similarity method and the hybrid feature extraction method.

For accuracy comparison, we used six methods: (1) our method (min distance), (2) our method (URR), (3) MF + Vpb extraction, (4) social network, (5) popularity, (6) our method. In the first method, we used the minimal distance algorithm. This algorithm creates a relation between users who have similar posting behaviors in Yelp and Twitter. In the second method, we set $\alpha = 1$ in the recommendation. In the third method, we used traditional matrix factorization instead of SITA for Yelp rating estimation. In the fourth method, we set $\alpha = 0$ in the recommendation. In the fifth method, we recommended hotel items according to their real-life popularity (the baseline method). In the sixth method, we set

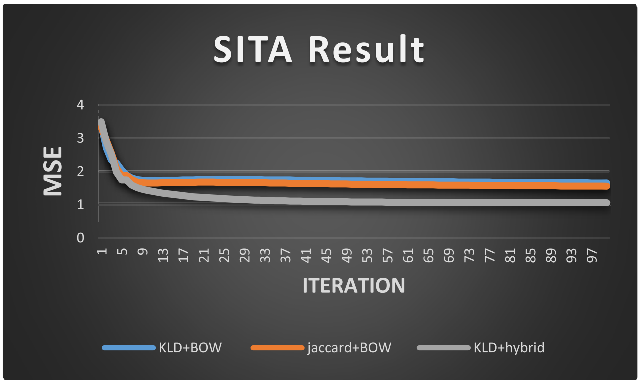


FIGURE 19. MSE result on different methods.

$\alpha = 0.7$ in the recommendation. The greatest difference between method the second method and the sixth lies in the fact which the second method only refer to the auxiliary information provided by Yelp to build the system; whereas the sixth method is not only based on auxiliary information, but also the friend social network of Twitter itself.

Figure 20 shows that our MRR method is able to increase the sensitive rapidly from 1 to 10. A higher MRR reflects that the system can quickly recommend the correct item. The MRR result for our method is 1.5 times higher than that for the popularity method, when both have the best result. Our system was thus proven to be able to recommend appropriate hotels for the users at the fastest speed. The system is also capable of handling a situation in which the user requires several hotels at the same time. Overall, our system can recommend the correct items faster than other methods. Table 4 shows all methods according to their MRR records. Comparing our method to other methods (URR), it was confirmed that our system, with heterogeneous social media auxiliary information added, can accurately choose the users' ideal hotel with only one or two recommended hotels provided by the system. The MRR is improved greatly. High recall means that the most correct items are identified by our system. We found the popularity method to have the best result in Figure 21 when the sensitivity = 5. Probably, some hotels

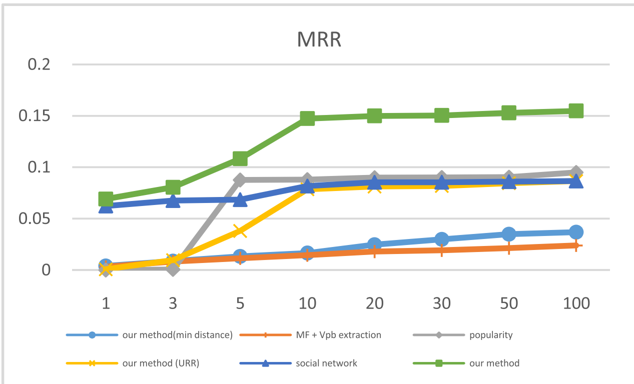


FIGURE 20. MRR tested on all of methods.

TABLE 4. The detail of MRR tested on all of methods.

	our method(min distance)	MF+ Vpbextraction	popularity	URR	social network	our method
1	0.00402	0.00322	0	0.0008	0.06237	0.06922
3	0.00872	0.00832	0.0006	0.00966	0.06767	0.08055
5	0.01325	0.01149	0.08761	0.03805	0.06856	0.10834
10	0.01654	0.01444	0.08802	0.07854	0.08175	0.14737
20	0.02471	0.01803	0.08991	0.08124	0.08539	0.14998
30	0.02991	0.01917	0.09007	0.08171	0.08545	0.15041
50	0.03484	0.02133	0.09038	0.08445	0.08599	0.15293
100	0.03684	0.02397	0.09494	0.08657	0.08659	0.15482

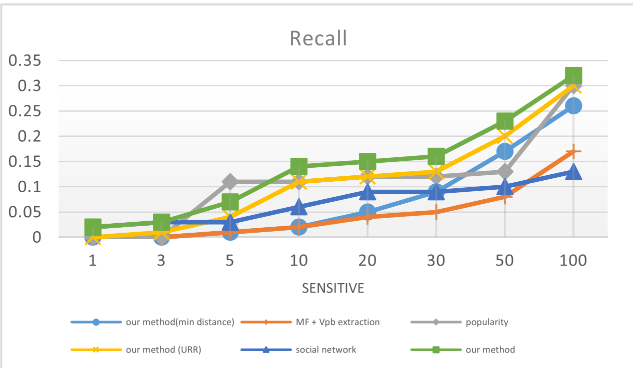


FIGURE 21. Recall tested on all methods.

have high ratings and topicalities. For instance, hotel with Death of Elisa Lam belongs to topicalities in 2013. Hence, Twitter users are interested in these hotels. At the same time, when the sensitivity is 1, the recall result by the popularity method is not good because these recommended items have high ratings without topicalities. Figure 22 shows the precision of all methods. High precision means that a system can precisely recommend correct items to a user. For instance, when the sensitivity is between 1 and 20, the precision of our method is greater than that of the other method (URR). That is to say, adding heterogeneous social media auxiliary information can improve the efficiency to a great extent.

D. CASE STUDY WITH DIFFERENT USERS

In this section, we discuss some interesting cases with two user groups. We divided Twitter users into two groups by a threshold defined as the number of hotels commented upon. One group with frequent users was labeled the "travel lovers group." The other group had relatively infrequent Twitter users, so we called this group "other group." Members in the travel lovers group commented on more than 10 hotels.

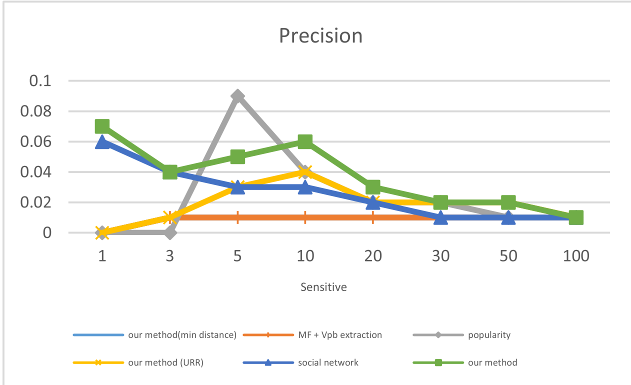


FIGURE 22. Precision tested on all of methods.

In the other group, users commented on less than 10 hotels. The differences are shown in Figures 4 and 25.

In Figure 23, it can be seen that our system can recommend the correct items to members in the other group more quickly. Furthermore, from Figures 23 and 24, we learn that when the sensitivity =1, our system is more accurate for the other group. This means that our system precisely and quickly recommends a correct hotel to infrequent Twitter users. Nevertheless, from dividing all the users into two groups, the result shows that the 'other' group acquires less information compared to the 'travel lovers' group. In other words, a cold start problem exists in the 'other' group. For a traditional recommendation system, the MRR results are confined to the cold start problem mentioned above, which bring about a better result on the travel lovers group as well. For our system, this MRR result in Figure 23 indicates that our system can deal with the cold start problem. The threshold is set as 10 hotels to divide the users into the two groups. Hence, when comparing the efficiency of the two parties, an F-score of less than 10 hotels needs to be taken into consideration, for which the sensitivity is less than 10. Figure 24 also shows that the F-score result for the travel lovers group is bigger than that for the other group when the sensitivity is lower than 10. The F-score thus further indicates that our system successfully resolves the cold start problem.

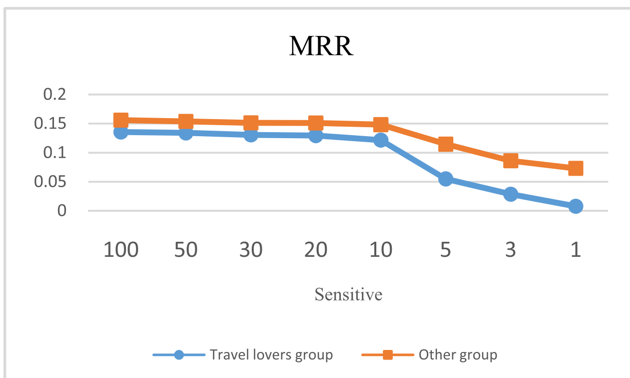


FIGURE 23. MRR results for the two Twitter user groups.

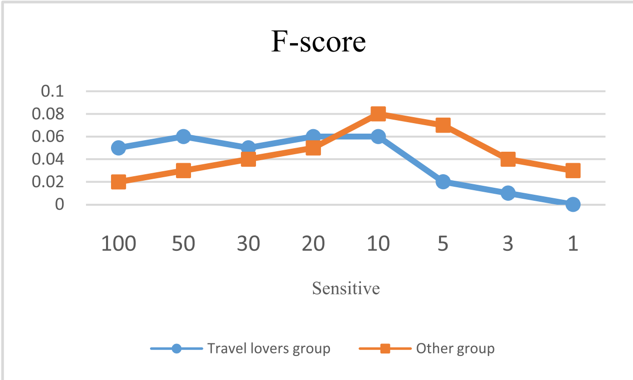


FIGURE 24. **F-score result for the two Twitter user groups.**

V. CONCLUSIONS AND PERSPECTIVES

A. CONCLUSIONS

It should be noted that there is a limitation in our system. Users have to set the social media ranking and mark the target items as heterogeneous social media personally, for instance, Yelp, Amazon, etc. These social media platforms include a variety of subjective comments and rankings. Thus, when transferring them as auxiliary information, a large amount of reliable information might improve the accuracy. If the aim is to use any other social media as the heterogeneous social media, the SITA should be edited once again.

B. PERSPECTIVES

We summarize our perspectives as follows:

1) Improve the performance of the Vpb extraction.

We used the LDA method to build user Vpbs in the form of Vpb extraction. Through the first experiment, we learned that the LDA method we used may not be the most appropriate technique. Deep learning is another choice that we could be taken into consideration that might make it possible to obtain a better result.

2) Enhance performance by applying more heterogeneous social media. In our study, we used only two social media platforms to build a recommendation system. More social media platforms could be used to build the auxiliary information and potentially obtain a better result.

3) Regard image-based social media as the heterogeneous social media. In this paper, we used Yelp as the heterogeneous social media. Yelp is a text-based form of social media. Image-based social media, such as Instagram, was not used as heterogeneous social media to develop the auxiliary information. If an image-based social media platform were to be used, it might improve our system.

It is hoped that our system can assist people in choosing hotels when they want to travel. It is also hoped that our heterogeneous social media method can help computer science researchers to extend a path in the recommendation system field.

APPENDIX

C. SITA

In this section, we describe the architecture of SITA. The J, Z, and R matrices are considered the true matrices. The $\hat{J}$, $\hat{Z}$, and $\hat{R}$ matrices are the predictive matrices.

$$\hat{R} = \sum\nolimits_{k=1} \mathrm{U_{ik}} * \mathrm{V_{kj}}$$
$$\hat{J} = \sum\nolimits_{k=1} \mathrm{U_{ik}} * \mathrm{RU_{kj}}$$
$$\hat{Z} = \sum\nolimits_{k=1} \mathrm{RV_{ik}} * \mathrm{V_{kj}}$$

The error formulas are the following:

$$\begin{aligned} \mathrm{e_{ij}^2} &= \left(R-\hat{R}\right)^2 + \left(J-\hat{J}\right)^2 + \left(Z-\hat{Z}\right)^2 \\ &= \left(r_{ij}-\sum\nolimits_{k=1} U_{ik} * V_{kj}\right)^2 + \left(j_{ij}-\sum\nolimits_{k=1} U_{ik} * RU_{kj}\right)^2 \\ &\quad + \left(z_{ij}-\sum\nolimits_{k=1} \mathrm{RV_{ik}} * \mathrm{V_{kj}}\right)^2) \end{aligned} \tag{V-6}$$

In this difference, r_{ij} denotes the row i and the column j rating in the R matrix, $\hat{r_{ij}}$ denotes the row i and the column j rating in the $\hat{R}$ matrix, and k denotes the number of latencies. In addition, j_{ij} denotes the row i and the column j rating in the J matrix and $\hat{j_{ij}}$ denotes the row i and the column j rating in the $\hat{J}$ matrix. Moreover, z_{ij} denotes the row i and the column j rating in the Z matrix and $\hat{z_{ij}}$ denotes the row i and the column j rating in the $\hat{Z}$ matrix. We set k = 100. We minimize the error by differentiating the squared error with respect to four variables ($\mathrm{U_{ik}}$, $\mathrm{V_{kJ}}$,$\mathrm{RV_{ik}}$,and $\mathrm{RU_{kj}}$) in Equation 5-1.

$$\frac{\partial}{\partial \mathrm{U_{ik}}}\mathrm{e_{ij}^2} = -2[\left(r_{ij}-\hat{r_{ij}}\right) * \left(V_{kj}\right) + \left(j_{ij}-\hat{j_{ij}}\right) * (RU_{kj})]$$

$$\frac{\partial}{\partial \mathrm{V_{kJ}}}\mathrm{e_{ij}^2} = -2[\left(r_{ij}-\hat{r_{ij}}\right) * (U_{ik}) + \left(z_{ij}-\hat{z_{ij}}\right) * (RV_{ik})]$$

$$\frac{\partial}{\partial \mathrm{RV_{ik}}}\mathrm{e_{ij}^2} = -2[\left(z_{ij}-\hat{z_{ij}}\right) * (V_{kj})]$$

$$\frac{\partial}{\partial R\mathrm{U_{kj}}}\mathrm{e_{ij}^2} = -2[\left(j_{ij}-\hat{j_{ij}}\right) * (U_{ik})]$$

Now, we have the minimal error with respect to those four variables, and we use it to design updating rules for those four variables ($\mathrm{U_{ik}}$,$\mathrm{V_{kJ}}$, $\mathrm{RV_{ik}}$ and $\mathrm{RU_{kj}}$):

$$\mathrm{U'_{ik}} = \mathrm{U_{ik}} - \alpha\frac{\partial}{\partial \mathrm{U_{ik}}}\mathrm{e_{ij}^2}) \tag{V-7}$$

$$\mathrm{V'_{kj}} = \mathrm{V_{kJ}} - \alpha\frac{\partial}{\partial \mathrm{V_{kJ}}}\mathrm{e_{ij}^2}) \tag{V-8}$$

$$\mathrm{RV'_{ik}} = \mathrm{RV_{ik}} - \alpha\frac{\partial}{\partial \mathrm{RV_{ik}}}\mathrm{e_{ij}^2}) \tag{V-9}$$

$$\mathrm{RU'_{kj}} = \mathrm{RU_{kj}} - \alpha\frac{\partial}{\partial R\mathrm{U_{kj}}}\mathrm{e_{ij}^2} \tag{V-10}$$

Where $\mathrm{U'_{ik}}$ denotes the updating of $\mathrm{U_{ik}}$, $\mathrm{V'_{kj}}$ denotes the updating of $\mathrm{V_{kJ}}$, $\mathrm{RV'_{ik}}$ denotes the updating of $\mathrm{RV_{ik}}$, and $\mathrm{RU'_{kj}}$ denotes the updating of $\mathrm{RU_{kj}}$, all according to the squared error.

The above Equation 5-2 to 5-5 lack regularization and suffer from overfitting. Therefore, we have to add a parameter β for modifying the squared error. The new squared error is represented as follows:

$$e_{ij}^2 = \left(R - \hat{R}\right)^2 + \left(J - \hat{J}\right)^2 + \left(Z - \hat{Z}\right)^2 + \beta * (\|U\|^2 + \|V\|^2 + \|RU\|^2 + \|RV\|^2)$$

Due to the addition of the parameter β, the updating formulas for the four variables (U_{ik}, V_{kJ}, RV_{ik} and RU_{kj}) also change as follows:

$$\begin{aligned} U'_{ik} &= U_{ik} - \alpha \frac{\partial}{\partial U_{ik}} e_{ij}^2 \\ &= U_{ik} - \alpha^*(\text{Equation 3-6}) - 2 * \alpha^* \beta * (U_{ik}) \\ V'_{kj} &= V_{kJ} - \alpha \frac{\partial}{\partial V_{kJ}} e_{ij}^2 \\ &= V_{kJ} - \alpha^*(\text{Equation 3-7}) - 2 * \alpha^* \beta * (V_{kJ}) \\ RV'_{ik} &= RV_{ik} - \alpha \frac{\partial}{\partial RV_{ik}} e_{ij}^2 \\ &= RV_{ik} - \alpha^*(\text{Equation 3-8}) - 2 * \alpha^* \beta * (RV_{ik}) \\ RU'_{kj} &= RU_{kj} - \alpha \frac{\partial}{\partial RU_{kj}} e_{ij}^2 \\ &= RU_{kj} - \alpha^*(\text{Equation 3-9}) - 2 * \alpha^* \beta * (RU_{kj}) \end{aligned}$$

We set $\alpha = 0.007$ and $\beta = 0.02$ for SITA.

REFERENCES

[1] P. Bhattacharya, M. B. Zafar, N. Ganguly, S. Ghosh, and K. P. Gummadi, "Inferring user interests in the Twitter social network," in *Proc. 8th ACM Conf. Recommender Syst.*, Foster City, CA, USA, 2014, pp. 357–360.

[2] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The Adaptive Web: Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Germany: Springer, 2007, pp. 325–341.

[3] J. He and W. W. Chu, "A social network-based recommender system (SNRS)," in *Data Mining for Social Network Data*, J. Xu, D. Hicks, and H. Chen, Eds. Boston, MA, USA: Springer, 2010, pp. 47–74.

[4] B. Berjani and T. Strufe, "A recommendation system for spots in location-based online social networks,"presented at the 4th Workshop Social Netw. Syst., Salzburg, Austria, 2011.

[5] S. Bostandjiev, J. O'Donovan, T. Höllerer, "TasteWeights: A visual interactive hybrid recommender system," presented at the 6th ACM Conf. Recommender Syst., Dublin, Ireland, 2012.

[6] Z. Shen, K.-L. Ma, and T. Eliassi-Rad, "Visual analysis of large heterogeneous social networks by semantic and structural abstraction," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 6, pp. 1427–1439, Nov. 2006.

[7] Y. Dong *et al.*, "Link prediction and recommendation across heterogeneous social networks," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 181–190.

[8] J. Hannon, K. McCarthy, and B. Smyth, "Finding useful users on Twitter: Twittomender the followee recommender," in *Advances in Information Retrieval*, P. Clough *et al.*, Eds. Berlin, Germany: Springer, 2011, pp. 784–787.

[9] Y. Kim and K. Shim, "TWILITE: A recommendation system for Twitter using a probabilistic model based on latent Dirichlet allocation," *Inf. Syst.*, vol. 42, pp. 59–77, Jun. 2014.

[10] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Analyzing user modeling on Twitter for personalized news recommendations," in *Proc. Int. Conf. User Modeling, Adaptation, Personalization*, 2011, pp. 1–12.

[11] F. Abel, Q. Gao, G. J. Houben, and K. Tao, "Analyzing temporal dynamics in Twitter profiles for personalized recommendations in the social Web," presented at the 3rd Int. Web Sci. Conf., Koblenz, Germany, 2011.

[12] D. Yang, D. Zhang, Z. Yu, and Z. Wang, "A sentiment-enhanced personalized location recommendation system," presented at the 24th ACM Conf. Hypertext Social Med., New York, NY, USA, 2013.

[13] J. Hannon, M. Bennett, and B. Smyth, "Recommending Twitter users to follow using content and collaborative filtering approaches," presented at the 4th ACM Conf. Recommender Syst., Barcelona, Spain, 2010.

[14] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu, "Collaborative personalized tweet recommendation," presented at the 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Portland, OR, USA, 2012.

[15] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, 2004.

[16] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," presented at the 10th Int. Conf. World Wide Web, Hong Kong, 2001.

[17] J. Zhang, P. S. Yu, and Z.-H. Zhou, "Meta-path based multi-network collective link prediction," presented at the 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA, 2014.

[18] X. Yu *et al.*, "Personalized entity recommendation: A heterogeneous information network approach," presented at the 7th ACM Int. Conf. Web Search Data Mining, New York, NY, USA, 2014.

[19] S. Zhu, S. Aloufi, J. Yang, and A. El Saddik, "On the learning of image social relevance from heterogeneous social network," *Neurocomputing*, vol. 210, pp. 269–282, Oct. 2016.

[20] H. Mi and E. M. Lee, "Hotel recommendation engine based on customer data from multiple online sources," U.S. Patent 14 195 711, Mar. 3, 2014.

[21] K.-P. Lin, C.-Y. Lai, P.-C. Chen, and S.-Y. Hwang, "Personalized hotel recommendation using text mining and mobile browsing tracking," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 191–196.

[22] Y. Zheng, "Situation-aware multi-criteria recommender system: Using criteria preferences as contexts," in *Proc. Symp. Appl. Comput.*, 2017, pp. 689–692.

[23] U. Fukumoto, H. Sugiyama, Y. Suzuki, and S. Matsuyoshi, "Exploiting guest preferences with aspect-based sentiment analysis for hotel recommendation," in *Proc. IC3K*, 2015, pp. 31–46.

[24] Y.-H. Hu, P.-J. Lee, K. Chen, J. M. Tarn, and D.-V. Dang, "Hotel recommendation system based on review and context information: A collaborative filtering appro," in *Proc. PACIS*, 2016, p. 221.

[25] T. Chen, "Ubiquitous hotel recommendation using a fuzzy-weighted-average and backpropagation-network approach," *Int. J. Intell. Syst*, vol. 32, no. 4, pp. 316–341, 2017.

JUI-HUNG CHANG is currently a full-time professor with the Computer and Network Center, Department of Computer Science Information Engineering, National Chen Kung University. Her current research interests include data base management, digital learning system, system analysis and object oriented design, program design, data mining, cloud calculation application, and project management and planning, travel information prediction and recommendation, geographical system, agricultural statistical big data, and biological big data mining with industry academy cooperation plan and research projects.

CHEN-EN TSAI received the degree from the Department of Computer Science and Information Engineering, National Cheng Kung University. He was with Intelligent Information Service Research Laboratory during graduate years. His research interests include machine learning, text mining, deep learning, and application of artificial intelligence.

JUNG-HSIEN CHIANG is currently the Distinguished Professor with the Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan. He has been involving in the fields of text mining and artificial intelligence for over 20 years. His current research interests include users profile modeling in social media, text mining, pattern clustering, and modeling of systems biology problems.