Preprint submitted to the Expert Systems with Applications (Elsevier). You can find the published version at:
https://www.sciencedirect.com/science/article/abs/pii/S0957417420310174

Tourism Recommendation System Based on Semantic Clustering and Sentiment Analysis

Zahra Abbasi-Moud[a] , Hamed Vahdat-Nejad[a*], Javad Sadri[b]

[a] Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran
zahra.abbasi@birjand.ac.ir, vahdatnejad@birjand.ac.ir

[b] Computer Science and Software Engineering Department, Concordia University, Montreal, Quebec, Canada

j_sadri@encs.concordia.ca

[*]Corresponding author

ABSTRACT

Numerous number of tourism attractions along with a huge amount of information about them on web and social platforms have made the decision-making process for selecting and visiting them complicated. In this regard, the tourism recommendation systems have become interesting for tourists, but challenging for designers because they should be able to provide personalized services. This paper introduces a tourism recommendation system that extracts users' preferences in order to provide personalized recommendations. To this end, users reviews on tourism social networks are used as a rich source of information to extract their preferences. Then, the comments are preprocessed, semantically clustered, and sentimentally analyzed to detect a tourist's preferences. Similarly, all users aggregated reviews about an attraction are utilized to extract the features of these points of interest. Finally, the proposed recommendation system, semantically compares the preferences of a user with the features of attractions to suggest the most matching points of interest to the user. In addition, the system utilizes the vital contextual information of time, location, and weather to filter unsuitable items and increase the quality of suggestions regarding the current situation. The proposed recommendation system is developed by Python and evaluated on a dataset gathered from TripAdvisor platform. The evaluation results show that the proposed system improves the f-measure criterion in comparison with the previous systems.

Keywords: Tourism recommendation system, Sentiment analysis, Context-awareness, Semantic similarity

1. Introduction

With the expansion of tourism websites and social networks, a huge amount of data and comments are produced and posted regularly (Ghane'i-Ostad, Vahdat-Nejad, & Abdolrazzagh-Nezhad, 2018; Neidhardt, Rümmele, & Werthner, 2017) .People who are planning for a trip use this data as well as reviews of other tourists as rich sources of information to select their appropriate destinations and points of interest (Hayashi & Yoshida, 2018; Renjith, Sreekumar, & Jathavedan, 2020). However, it is a major challenge for tourists to manually process large volumes of data (Borràs, Moreno, & Valls, 2014). In this regard, various tourism recommendation systems have been proposed that try to provide personalized suggestions to users. They aim to extract user preferences and present recommendations that are more in line with their preferences (Abel, Herder, Houben, Henze, & Krause, 2013). Some recommendation systems cluster users based on similarity in previously visited places and provide the same recommendations to users of each cluster (Esmaeili, Mardani, Golpayegani, & Madar, 2020; Wan, Hong, Huang, Peng, & Li, 2018). It is noteworthy

that the users' visit to a tourist attraction by itself does not provide enough information and their reviews regarding these places are also important. As a result, another set of recommendation systems leverage an analysis of comments to extract user preferences (Xiang, Du, Ma, & Fan, 2017). In this regard, user reviews are analyzed and compared with attractions' metadata, and places with the highest matching are suggested (Leal, González–Vélez, Malheiro, & Burguillo, 2017). In this type of tourism recommendation systems, the frequent keywords used in the comments are exploited, regardless of the sentiments of users. As a result, the negative words that are stressed in user's text might be mistakenly returned as their preferences. While sentiment analysis is important in the tourism domain, it has been overlooked in most cases (Alaei, Becken, & Stantic, 2019). In fact, a comprehensive recommendation system in the context of tourism should include the following features:

- It should identify preferences by looking for concepts instead of being limited to specific keywords.
- It should leverage sentiment analysis on user comments to identify their positive versus negative preferences.
- It should provide context-aware recommendations (Vahdat-Nejad, 2014), which are adapted to user's current situation.

To the best of the authors' knowledge, none of the available tourism recommendation systems meets all of the above features. Hence, this paper proposes a context-aware tourism recommendation system by sentiment analysis. In this system, text processing and sentiment analysis are leveraged to extract users' preferences, precisely. The preference extraction part is an extension to the initial paper (Abbasi-Moud, Vahdat-Nejad, & Mansoor, 2019) in which user reviews about various attractions are extracted and preprocessed. Then the preferences are extracted through semantic clustering and sentiment analysis. This research extends the initial idea by extracting features of attractions from aggregated users' reviews and proposing a personalized recommendation system. The system additionally utilizes user contextual information including location (in order to identify the attractions around him/her), time (to check when the attractions could be visited), and weather (to provide recommendations that are appropriate to the current weather situation). The proposed system is developed by Python. Therefore, an experiment is conducted on TripAdvisor[1], as a well-known travel platform, to evaluate the proposed system. In this regard, a dataset including 100 users comments and visits in 2018 has been gathered. The evaluation results reveal high efficiency of the proposed recommendation system in terms of precision, recall, and f-measure.

The structure of the paper is as follows. In the next section, a review of the related literature is provided. Section 2 introduces the proposed recommendation system. In the section 4, details of the implementation and evaluation, and in the final section, the conclusion remarks are presented.

2. Related work

[1] www.TripAdvisor.com

Most of the tourism recommendation systems are based on geographic tags(Cai, Lee, & Lee, 2018). As an example, the tags of photos shared on Flickr[1] are used to identify and cluster users who follow similar paths and this information is leveraged in order to recommend tourist attractions(Majid et al., 2013). Due to the importance of user reviews about the places they have visited, a number of recommendation systems utilize comments analysis. Moreover, context-awareness plays a major role in improving the quality of the tourist recommendation systems. In this section, at first context-aware tourism recommender systems and then tourism recommender systems that are based on user reviews are investigated, respectively.

Among many contextual information, location is the most important element used in current tourism recommendation systems (Abowd et al., 1999; Yochum, Chang, Gu, & Zhu, 2020). In this regard, the behavioral patterns of people who travel to the protected areas of the Ningaloo Marine Park in Australia were extracted using GIS (Smallwood, Beckley, & Moore, 2012). Additionally, the user's current position has been used to provide recommendations (Tumas & Ricci, 2009). These recommendations are displayed based on user preferences for both public transportation and on foot. However, the major drawback of methods that only use location as the context information is their domain constraints and their one-dimensionality.

User's various contextual elements including location, speed, and route have been exploited to provide personalized recommendations for visiting their favorite tourist destinations (Barranco, Noguera, Castro, & Martínez, 2012). Furthermore, PSiS (Anacleto, Figueiredo, Almeida, & Novais, 2014) suggests the user suitable destinations in accordance with their contextual information, including location, time, speed, direction and weather. In this system, the users' tourism history is used as their preferences to provide more accurate recommendations. A similar tourism recommender system has been proposed for Chinese language, which targets 100 prominent attractions of Taiwan (Yeh & Cheng, 2015). As individuals with different occupations, ages and nationalities usually have different interests and perspectives, the age, nationality and income of the user have been used as context elements to improve the accuracy of the recommendations(Lu, Wu, Mao, Wang, & Zhang, 2015). Finally, TripAdvisor data is utilized to predict users preferences (Pantano, Priporas, Stylos, & Dennis, 2019). In this system, each user should select at least three topics from 18 tourism topics (e.g. Eco-tourism, Nature lover, etc.). Therefore, similar set of items are suggested to the users with similar requirements and interests.

Most of the traditional tourism recommendation systems are based on user ratings on visited attractions. As text can bear much more information than a rating, reviewing users' comments can greatly improve the

[1] www.Flickr.com

accuracy of these systems (Xiaoyao Zheng, Luo, Sun, Zhang, & Chen, 2018). Hence, researchers have been analyzing user opinions by text-mining techniques (Bao, Fang, & Zhang, 2014; Xiaolin Zheng, Ding, Lin, & Chen, 2016). In this regard, Loh et al. have designed a private chat page that asks questions from users in order to find specific vocabulary within the scope of an ontology. Then the type of attractions desired by the user is extracted based on the ontology concepts and recommendations are provided, accordingly (Loh, Lorenzi, Saldaña, & Licthnow, 2003).

POST-VIA360 (Colomo-Palacios, García-Peñalvo, Stantchev, & Misra, 2017) is a bio-inspired recommender system aims to make loyalty to the tourists after their first visit of an attraction. It uses a tourism Ontology to provide suggestions based on previous visits, current location and social aspects. Yochum et al. have created a knowledge graph for Bangkok's tourist attractions (Yochum, Chang, Gu, Zhu, & Zhang, 2018). This graph is based on the words in the text of the users' comments, each of which is considered as a concept. Then the characteristics of tourists and tourist attractions are shown as a concept vector. Finally, using the cosine similarity measurement between vectors, the correlation between tourists and attractions is calculated. Another study uses Topic modeling to extract titles from the user's reviews and other tourists' reviews about nearby attractions (Leal et al., 2017). Then, using the semantic similarity based on WordNet, the similarity of attractions is compared with the user's preferences and the most similar attractions are recommended.

Considering users' sentiments is vital in extracting their preferences from their reviews to avoid identifying negative points as preferences. To this end, this paper proposes a tourism recommendation system that exploits sentiment analysis and text mining to identify user preferences as well as key features of tourist attractions. In fact, the proposed method augments previous research by considering sentiment analysis in text mining on tourist's reviews. Besides, in contrary to the reviewed papers, the proposed system is context-aware, in which several contextual elements including time, location, weather condition, user's preferences, and attractions features are taken into account.

3. Proposed system

The proposed tourism recommender system consists of three stages. In the first stage, user preferences are extracted from their comments and reviews. Similarly, in the second stage, the characteristics of tourist attractions are extracted from the reviews of tourists regarding them. Finally, in the third stage, appropriate recommendations are presented based on the contextual information as well as the similarity between user preferences with the characteristics of the tourist attractions. The contextual information used in this method includes weather, time, location, and user preferences. The pseudo code of the proposed recommender system is presented in algorithm 1. The detail is discussed below.

Preprint submitted to the Expert Systems with Applications (Elsevier). You can find the published version at:
https://www.sciencedirect.com/science/article/abs/pii/S0957417420310174

Algorithm1: The proposed recommendation algorithm

Inputs:

PoI Data: A = (<PoI a, Location l, Reviews $r_{all\ tourists}$ >)

User Data: U = (<Location l, Reviews r_{user} >)

Context Data: C = (<Location l, Weather w, Time t>)

Output:

Ordered list of recommended PoIs per user: PoI_u = $[PoI_a, \ldots, PoI_n]$

Step1:

Extract user preferences

Step2:

In weather w

for each nearby user PoI **do:** Top_{PoI} ← Five most repetitious concepts of reviews of all tourists

Step3:

for each nearby user PoI **do** similarity-PoI=Compute Similarity(user preferences, Top_{PoI})

return PoIs sorted by similarity-PoI

if in l,t, (w=rainy or w=stormy or w=snowy)

give priority to indoor places as sorted above

else return previous results

Evaluation parameters: Precision, Recall and F-measure

3-1. Extracting preferences

Four main steps including preprocessing, semantic graph formation, clustering, and preferences extraction are performed in this stage (see Fig. 1).

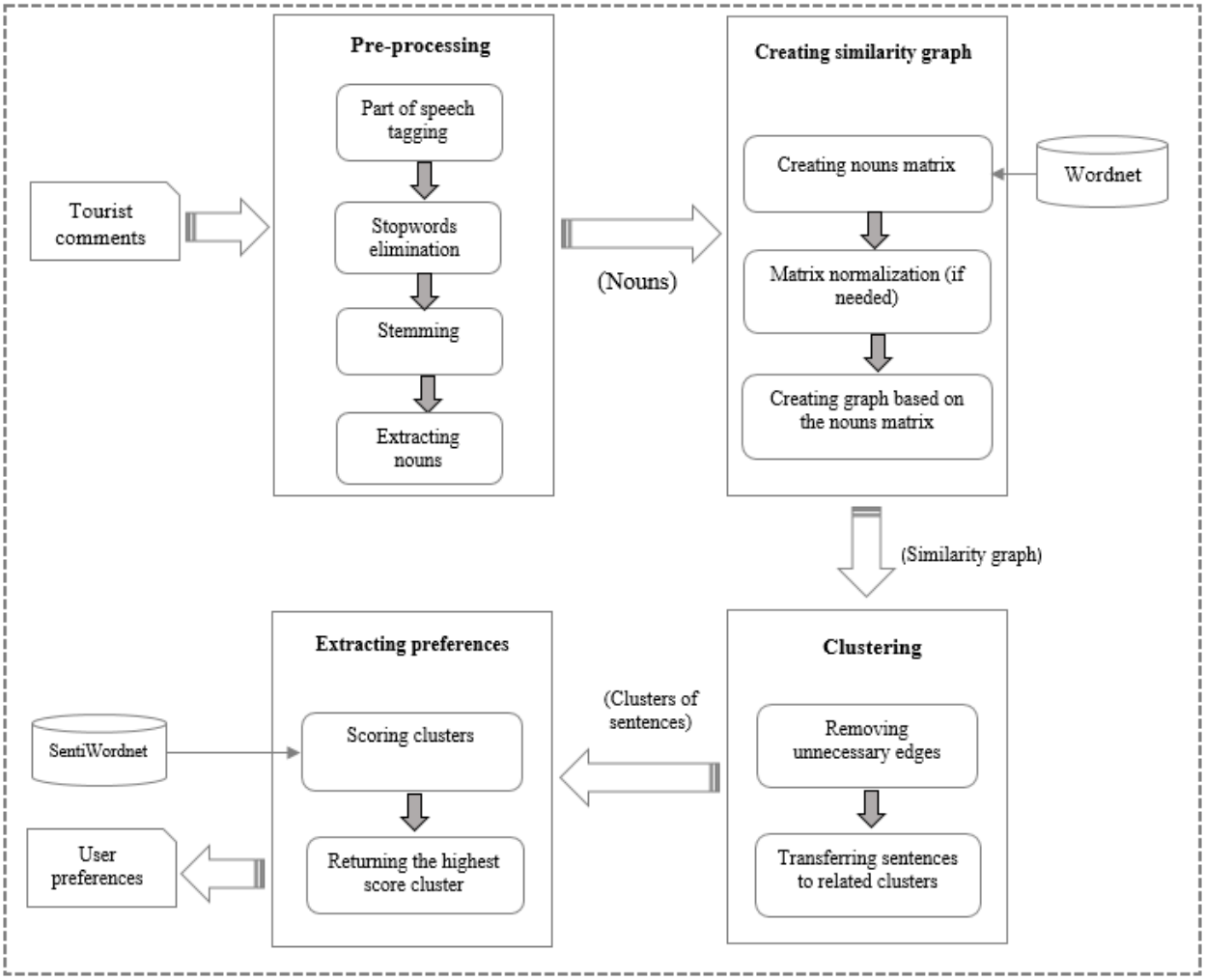

Fig. 1- Extracting a user's preferences

The pre-processing is performed to convert an initial document into a suitable processing form. Fig. 2 shows the scheme of this step. The operations are described below.

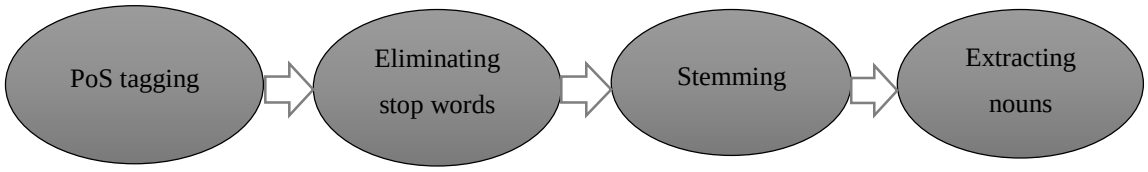

Fig. 2- Pre-processing operations

Part of Speech (PoS) tagging: In this sub-stage, the constituents of the sentence, including noun, verb, etc. are identified and tagged. This helps to extract information from sentences.

Stop words elimination: In a sentence, the words that do not have any specific meaning are called stop words. By eliminating them, only the words with useful information remain.

Stemming: Stemming refers to the process of converting a word into its base or stem. In this sub-stage, words such as cats and catlike are turned into their stem, i.e. cat. To this end, Wordnet is used, which keeps the stem form of all words.

Extracting nouns: Nouns are the most informative constituents of a sentence. Using only nouns increases the clustering efficiency in comparison with the use of all the words in a text(Fodeh, Punch, & Tan, 2011). As a result, only the words that are tagged as noun are extracted in this sub-stage for further processes.

After pre-processing, the noun similarity matrix is constituted. It is a symmetric matrix in which rows and columns correspond to the extracted nouns (as shown in Fig. 3). The semantic similarity between each pair of entries forms the corresponding element in the matrix.

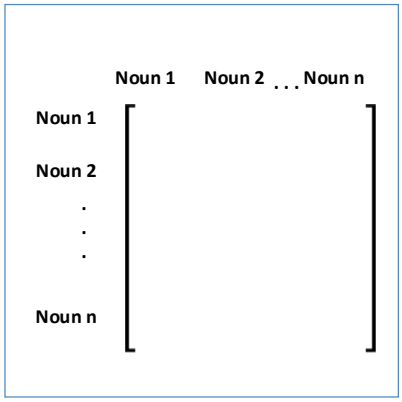

Fig. 3- Nouns matrix

To constitute the matrix, a previously proposed hybrid semantic similarity measure (Wei, Lu, Chang, Zhou, & Bao, 2015) is exploited. This measure considers all the direct and indirect relationships between concepts in the Wordnet to increase the accuracy in calculating the similarity. It addresses the defects of the Wu-Palmer semantic similarity (Wu & Palmer, 1994) (Lack of the direct relationships between concepts) and the Extended gloss overlaps (Banerjee & Pedersen, 2003) (Lack of similarity between the concepts that have a direct relationship in the structure of WordNet, but are not similar in their definitions).

Afterward, the elements of the matrix are normalized so that the method can be evaluated with different semantic similarity measurement criteria. To this end, the values of all matrix elements are divided by the maximum element.

Finally, the graph for this matrix is created, the vertices of which represent the extracted nouns of the preprocessing stage and the weights of the edges represent the semantic similarity of the two meeting vertices (nouns). If there is not semantic similarity between a pair of vertices, these two vertices are not

connected. Then, the edges that have a weight less than a specified threshold are eliminated. As a result, the graph may be converted to several connected sub-graphs. Each sub-graph is considered as a cluster in which each noun has a reasonable semantic similarity with at least one other noun.

Finally, each sentence is transferred to the clusters that contain any of its constituent words. Hence, based on the nouns in a sentence, it might be placed in more than one cluster. For example, if the "X" cluster contains a, b, c, and d nouns, and the "Y" cluster contains e and f, the “abg” sentence is located in the cluster X, while the “afh” sentence is assigned to both clusters.

In continue, each cluster is scored based on the result of sentiment analysis as well as the frequency of nouns. In this research, sentiment analysis is performed semantically with the help of the Sentiwordnet (Baccianella, Esuli, & Sebastiani, 2010). Since a word in different situations may have different meanings and sentimental loads, average of the positive as well as negative loads of synsets of the word is used as the positive and negative load of that word, respectively. The score of each sentence is computed by deducting the negative scores from the positive ones. The emotional load of emoticons is also taken into account. Positive emoticons of each sentence are scored +1; while negative emoticons are scored -1 (Neidhardt et al., 2017).

As equation (1) expresses, the sentiment analysis score of each cluster is equal to the average of sentiment analysis score of its sentences.

$$Score_{Sentiment\ Analysis}(cluster_i) = \frac{\sum Score\ of\ each\ sentence\ of\ (cluster_i)}{Total\ number\ of\ sentences\ in\ (cluster_i)} \quad (1)$$

In a text, repeated words (except for stop words) are usually more important than less frequent words (Binwahlan, Salim, & Suanmali, 2010). Therefore, the frequency of nouns of each cluster has been involved in scoring the clusters. Equation (2) shows the formula for computing the ith cluster score.

$$Score_{cluster_i} = TF_{cluster_i} \times Score_{Sentiment\ Analysis}(cluster_i) \quad (2)$$

Where $TF_{cluster\ i}$ is equal to the total number of repetitions of nouns in the cluster i. For example, for a cluster i with noun frequencies of 2, 2, 5, and 6, the value of $TF_{cluster\ i}$ is 15.

Finally, the cluster that gains the highest score is considered as the selected cluster. In addition, sufficiently closed clusters to the selected cluster (less than 10-point score difference) are also considered as selected clusters. The set of nouns in these clusters represent the preferences of the user.

3-2. Extracting attractions features

After extracting users' preferences, the features of tourism attractions should be extracted. In order to consider the quality of recommendations, the attractions that have received less than 3 stars by tourists are ignored. Besides, tourism attractions have different characteristics in different weather conditions. For example, while the Alps in snowy days are known for snowy landscape and cold weather; in sunny days of summer, they are known for lush landscape and moderate weather. Hence, for every tourism attraction, the set of reviews of tourists, who have visited the attraction, are collected and preprocessed in five different weather conditions including snowy, rainy, sunny, stormy and partly cloudy. Then, the top five most repeated words of each attraction in each of these five weather conditions are extracted and are considered as the attraction's features. Fig. 4 shows the proposed steps.

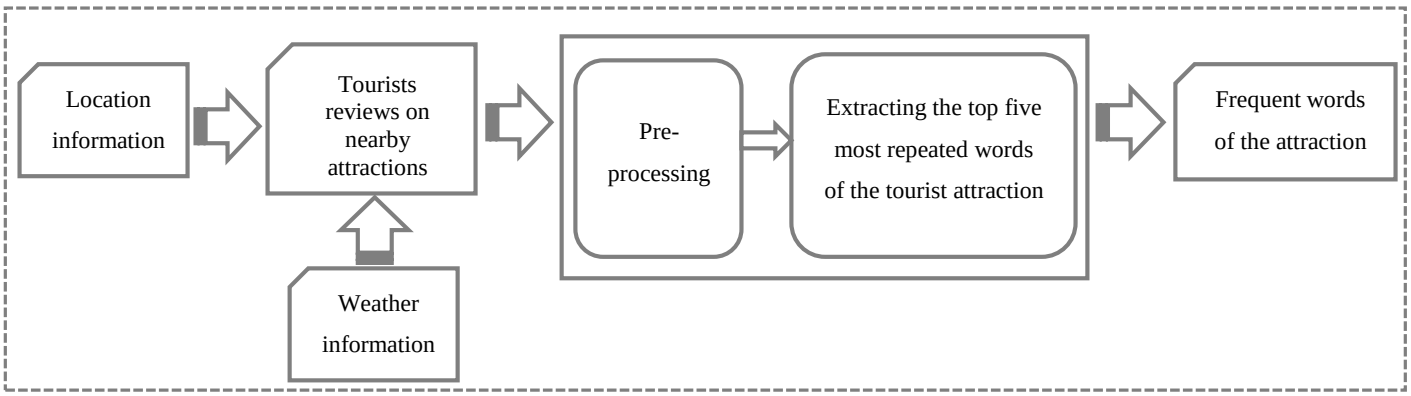

Fig. 4- Extracting an attraction's features

3-3. Recommendation system

The main idea of the proposed recommendation system is to compare a tourist's preferences with the nearby attractions' features and return the most similar attractions. To compute the similarity of a user's preferences with an attraction's features, the maximum similarity of any of the user's preference elements regarding all the features of the attraction is calculated. Therefore, these maximum similarities are averaged for all preference elements of the user (Equation (3))(Haase, Siebes, & Van Harmelen, 2004).

$$Sim(P,F) = \frac{1}{|P|}\sum_{P_i \in P} \underset{F_j \in F}{Max\ similarity}(P_i\ , F_j) \tag{3}$$

In the above Equation, P is the set of user's preferences and F is the set of attraction's features. The semantic similarity criterion is the same as the measure used for the user preference extraction.

Finally, if the weather condition is snowy, rainy or stormy, indoor locations will be prioritized over outdoor locations to reduce visiting troubles. On the other hand, if the weather is expected to be fine, the original results will be shown. The general scheme of the recommendation is shown in Fig. 5.

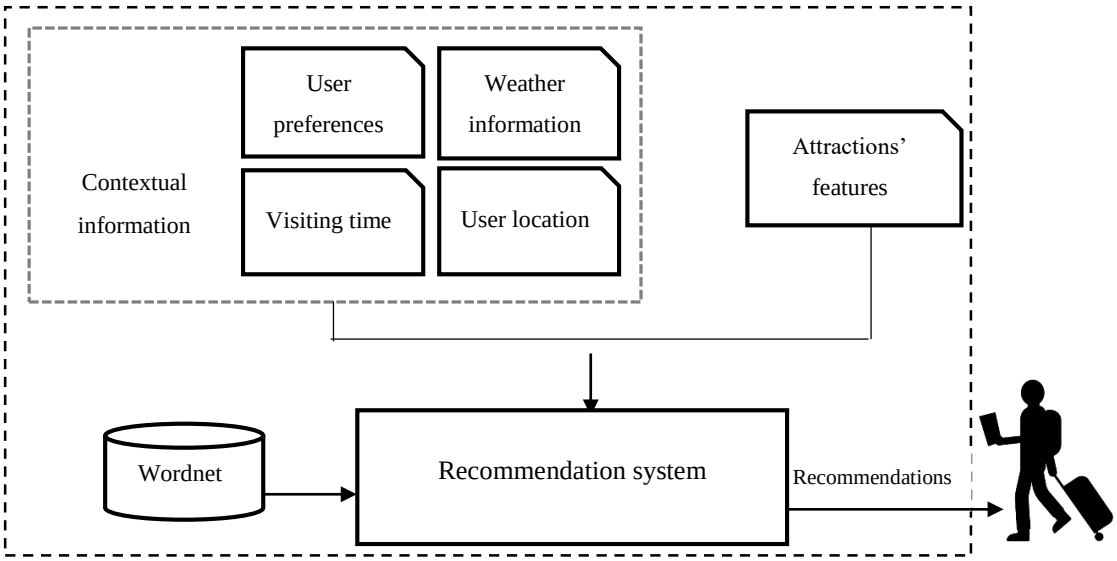

Fig. 5- General scheme of attraction recommendation

4. Evaluation

The proposed recommender system has been developed by the python3 language and Anaconda[1] platform. As TripAdvisor is the most well-known tourism platform(Pantano, Priporas, & Stylos, 2017), which has been used to collect a dataset for evaluating the proposed recommender system. The main part of the dataset (training data) includes the reviews of 100 tourists with different ages and nationalities on various tourist attractions in the course of six months (January-June 2018). The first travel of the users after this period (from July 2018) along with all of its attraction visits are considered as the test data. Finally, the sentiment analysis process has been performed using Sentiwordnet 3.0(Baccianella et al., 2010).At first, the proposed sentiment analysis method is validated by comparing it with the approaches that are based on Support Vector Machine (SVM) and Bayesian network (which are two popular methods in the field of sentiment analysis(Rana & Singh, 2016)). To this end, the reviews of tourists are categorized into either positive or negative classes. Besides, this classification is also performed by considering ratings of users. In this case,

[1] www.anaconda.org

if the user has given a score of 4 or 5 bubbles (very good, excellent[1]), it is considered positive and if the score is equal to 1 or 2 bubbles (terrible, poor), it is considered negative (Valdivia, Luzón, & Herrera, 2017).Afterwards, the percentage of positive and negative comments are computed by each approach and the result is shown in Fig. 6. It reveals that the proposed approach achieves the most similar results to the ratings provided by users.

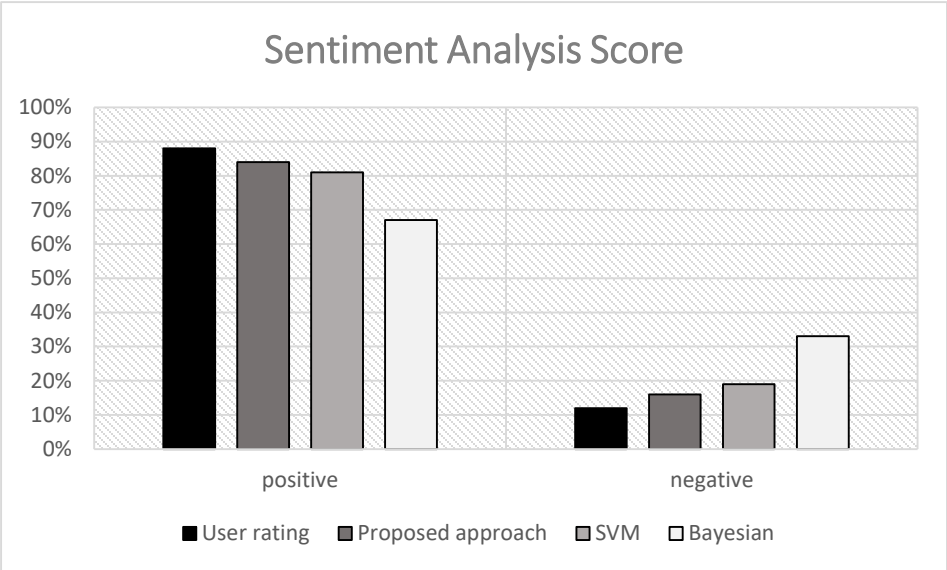

Fig. 6- Validating the proposed sentiment analysis method

Next, the utilized hybrid semantic similarity measure, which is a combination of the two criteria of Wu-Palmer (Wu & Palmer, 1994) and Extended Gloss Overlaps (Banerjee & Pedersen, 2003), is evaluated. To this end, it is compared with each of these two methods in clustering the nouns of user comments. On the other hand, the Cosine, Jaccard (Liu, Hu, Mian, Tian, & Zhu, 2014) and hybrid (Wei et al., 2015) semantic similarity measures are investigated to calculate the similarity of user preferences with features of tourist attractions.

If the radius of user's current city is φ km, three values of $\frac{\varphi}{4}$, $\frac{\varphi}{2}$ and φ are considered as the allowed distance for suggestion. Besides, three different modes are considered for evaluation, which are suggesting one (Top1), three (Top3), and five (Top5) recommendations, respectively. For each of these scenarios, three criteria of precision, recall, and f-measure are evaluated. Precision estimates what percentage of recommendations are actually visited by the user. It should be noted for top3 and top5 that if the user visits

any of the recommended attractions, the recommendation is regarded as successful. Similarly, recall indicates what percentage of actually visited attractions are recommended by the system. Figure 7 shows the evaluation results for top1. The values of the horizontal axis (A-B) represent the semantic similarity measures used for clustering nouns of user comments and the measure for similarity estimation of preferences with features of attractions, respectively. In fact, A represents the measure used in the clustering of nouns, and B represents the measure used in estimating the similarity between users preferences with attractions features.

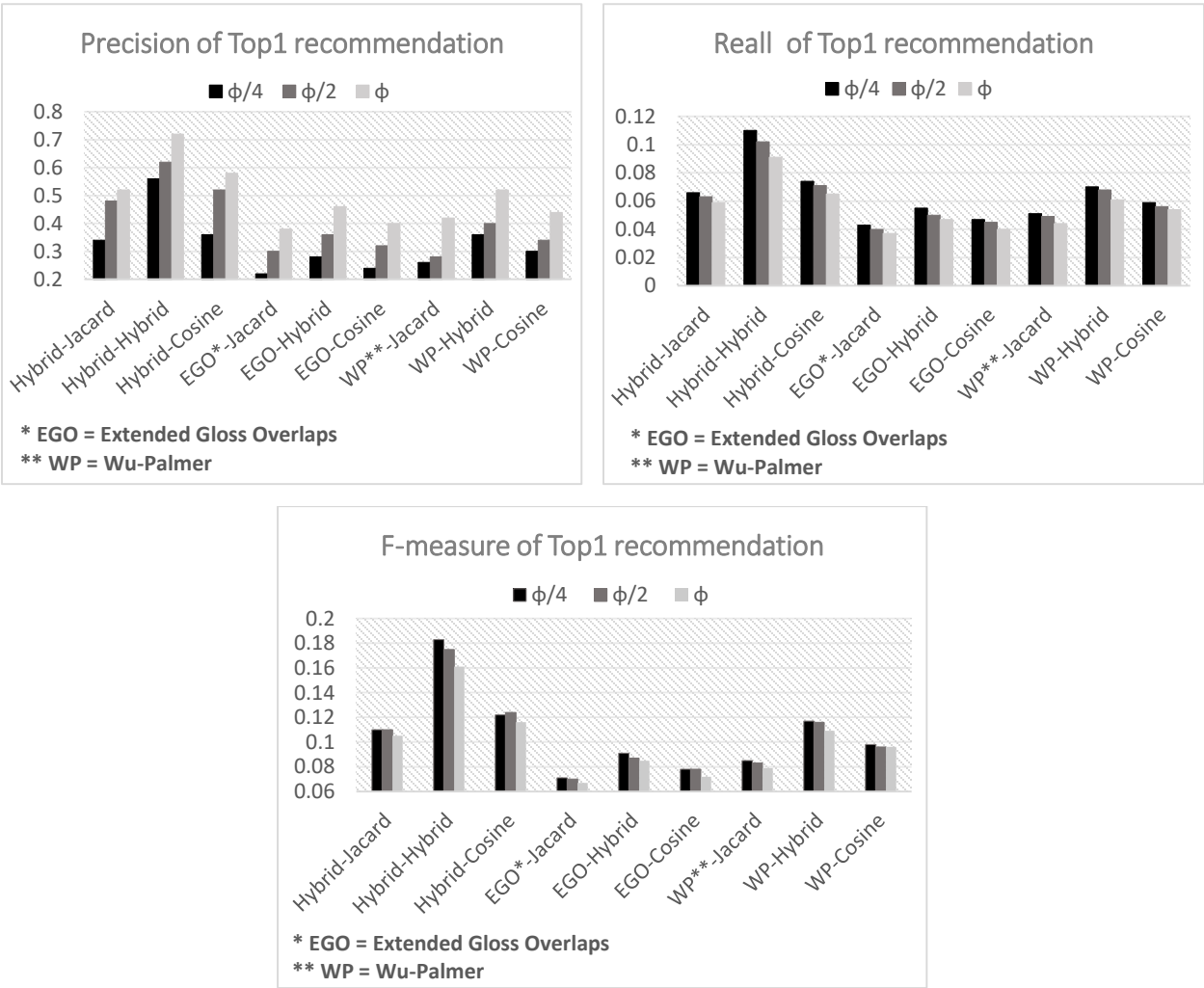

Fig7: Top1 recommendation (precision, recall, f-measure)

The results indicate the superiority of the hybrid criteria in comparison with other similarity metrics for top1. In fact, the use of hybrid methods results in the best precision, recall, and f-measure for different radii. Correspondingly, figures 8 and 9 show similar results for top3 and top5, respectively.

Preprint submitted to the Expert Systems with Applications (Elsevier). You can find the published version at:
https://www.sciencedirect.com/science/article/abs/pii/S0957417420310174

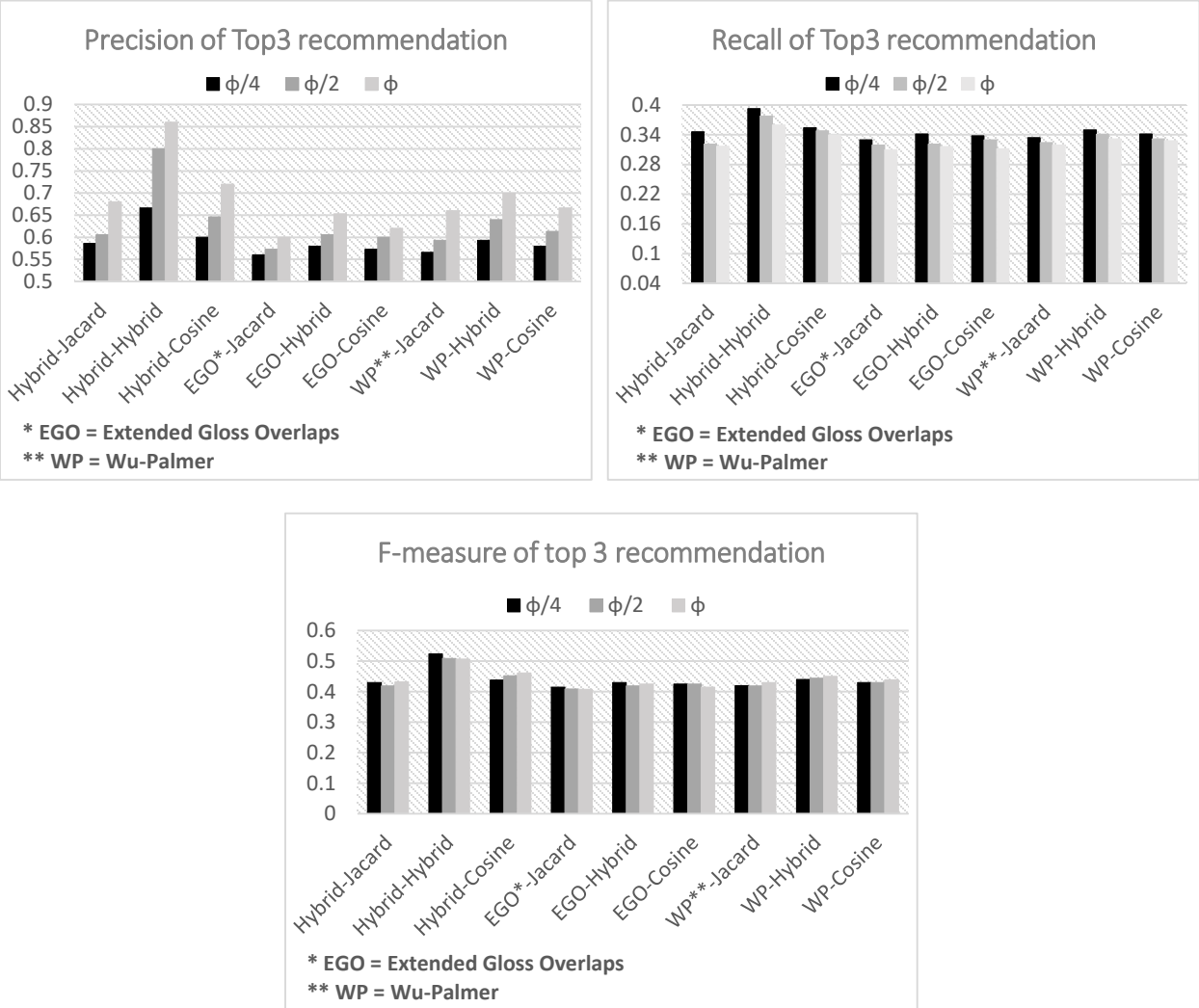

Fig. 8- Top3 recommendation (precision, recall, f-measure)

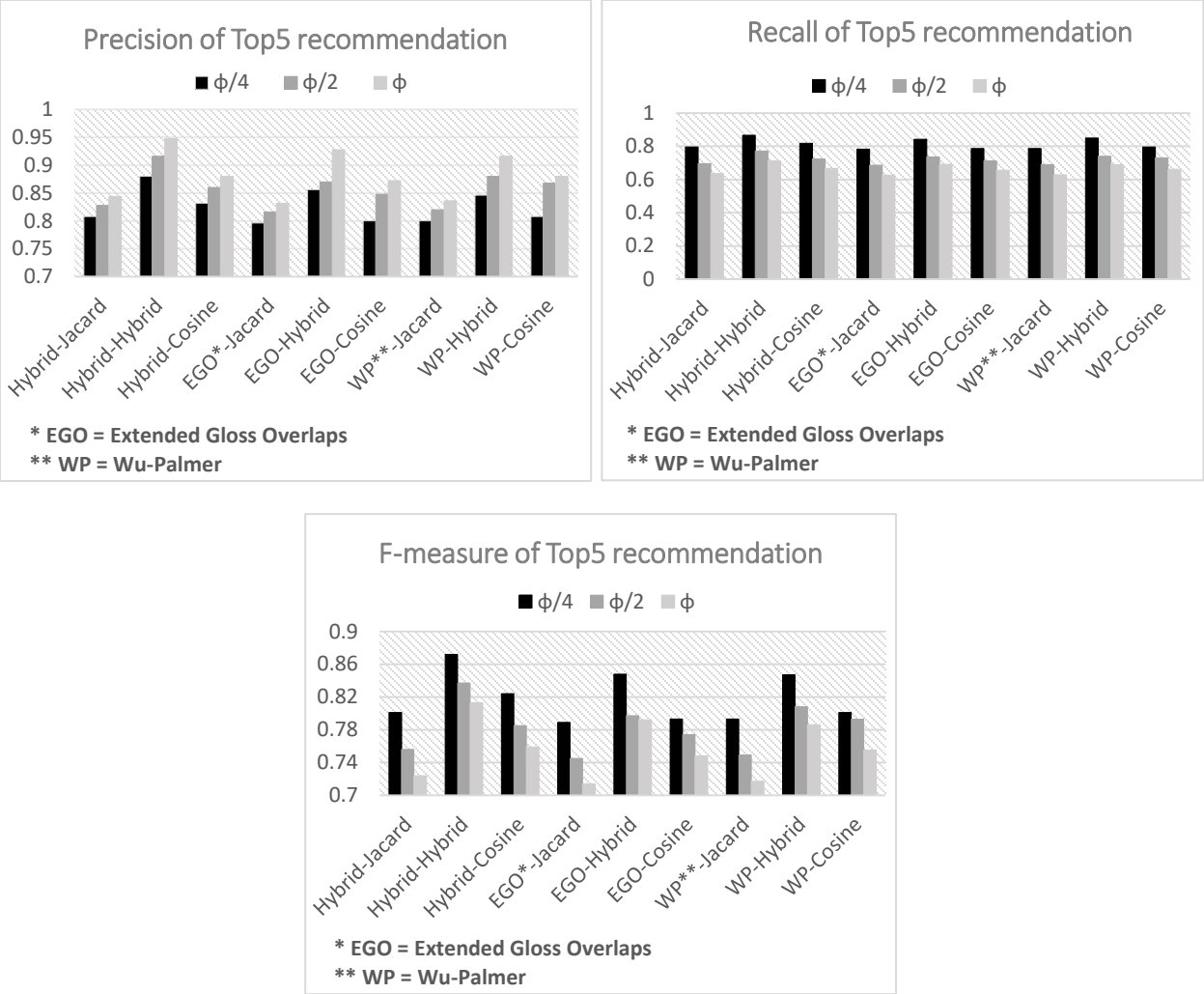

Fig. 9- Top5 recommendation (precision, recall, f-measure)

When the number of recommendations increases, the precision (the possibility that the user visits at least one of them) as well as recall (the possibility that a visited attraction has been recommended) increases. Since f-measure is dependent on precision and recall, it is also greater for top5 comparing with top3, which in turn yields better f-measure comparing with top1.

Finally, the proposed system is compared with two other similar systems proposed by Leal et al. (Leal et al., 2017) and Loh et al. (Loh et al., 2003). Due to the similarity of the basis of these studies with our idea, they have been used to evaluate the effectiveness of the proposed method. As mentioned in the related work section, Loh et al. use the keywords of the user's reviews in a private chat and consider them as the user's preferences and offer suggestions based on them. In the Leal's proposed method, user preferences and features of tourist attractions are extracted using the topic modeling and suggestions are presented based on the similarity between preferences and features.

In this research, f-measure is the most comprehensive criterion, hence the comparison is based on it. Fig. 10 shows the results of comparing f-measure for top1, top3, and top5 as well as the radius of φ/4. This comparison is based on the highest f-measure score obtained in the evaluation.

According to the evaluation results, the proposed system has achieved desired values for the evaluation measures. Also, since most recommender systems offer more than one recommendation, if the user receives more offers, the proposed system will perform much better than similar systems.

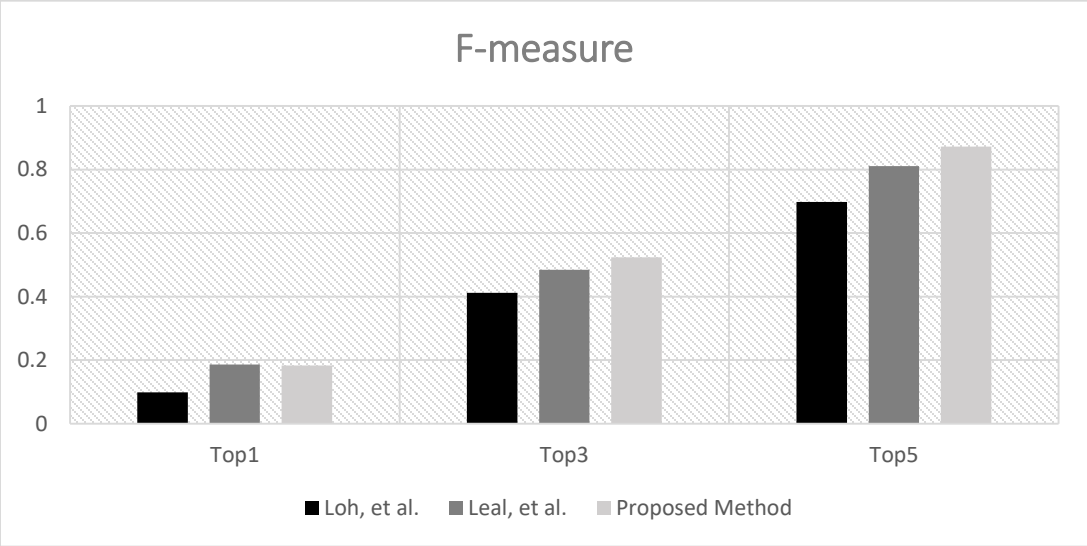

Fig. 10- Comparison of the proposed system with similar methods

5. Conclusion

In this paper, a context-aware tourism recommendation system has been presented, which extracts a user's preferences by performing semantic clustering as well as sentiment analysis on their comments and reviews. Similarly, the features of an attraction are extracted from aggregated reviews of users about it. Subsequently, nearby attractions are ranked according to their similarity with the user's preferences as well as current contextual information. Finally, an experiment has been conducted on TripAdvisor website and reviews published by one hundred users were extracted. For each user, the system has been trained by the data from the first 6 months, while the first trip after this period along with all of its attractions visits were considered for testing the system. Evaluation results have shown that the proposed system has outperformed comparable systems in terms of f-measure.

Although time, location, weather, users' preferences, and attractions' features have been exploited as vital contextual information, other contextual information such as the traffic of routes, various conditions of

individuals as well as group users and their environment have been neglected. Augmenting the proposed system by taking into account all the contextual information could enhance the recommendation process as well as user satisfaction and convenience. As users usually travel and visit attractions in groups, extending the proposed recommendation system for group scenarios is an important future research direction.

Acknowledgement

It is a pleasure for authors to thank Mr. Saeed hosseinabadi for his valuable remarks and helps.

References

Abbasi-Moud, Z., Vahdat-Nejad, H., & Mansoor, W. (2019). *Detecting Tourist's Preferences by Sentiment Analysis in Smart Cities.* Paper presented at the 2019 IEEE Global Conference on Internet of Things (GCIoT).

Abel, F., Herder, E., Houben, G.-J., Henze, N & ,.Krause, D. (2013). Cross-system user modeling and personalization on the social web. *User Modeling and User-Adapted Interaction, 23*(2-3), 169-209.

Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggles, P. (1999). *Towards a better understanding of context and context-awareness.* Paper presented at the International symposium on handheld and ubiquitous computing.

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research, 58*(2), 175-191.

Anacleto, R., Figueiredo, L., Almeida, A., & Novais, P. (2014). Mobile application to provide personalized sightseeing tours. *Journal of Network and Computer Applications, 41*, 56-64.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). *Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining.* Paper presented at the Lrec.

Banerjee, S., & Pedersen, T. (2003). *Extended gloss overlaps as a measure of semantic relatedness.* Paper presented at the Ijcai.

Bao, Y., Fang, H., & Zhang, J. (2014). *Topicmf: Simultaneously exploiting ratings and reviews for recommendation.* Paper presented at the Twenty-Eighth AAAI conference on artificial intelligence.

Barranco, M. J., Noguera, J. M., Castro, J., & Martínez, L. (2012). A context-aware mobile recommender system based on location and trajectory *Management intelligent systems* (pp. 153-162): Springer.

Binwahlan, M. S., Salim, N., & Suanmali, L. (2010). Fuzzy swarm diversity hybrid model for text summarization. *Information processing & management, 46*(5), 571-588.

Borràs, J., Moreno, A., & Valls, A. (2014). Intelligent tourism recommender systems: A survey. *Expert Systems with Applications, 41*(16), 7370-7389.

Cai, G., Lee, K., & Lee, I. (2018). Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos. *Expert Systems with Applications, 94*, 32-40.

Colomo-Palacios, R., García-Peñalvo, F. J., Stantchev, V., & Misra, S. (2017). Towards a social and context-aware mobile recommendation system for tourism. *Pervasive and Mobile Computing, 38*, 505-515.

Esmaeili, L., Mardani, S., Golpayegani, S. A. H., & Madar, Z. Z. (2020). A novel tourism recommender system in the context of social commerce. *Expert Systems with Applications, 149*, 1133 .٠١

Fodeh, S., Punch, B., & Tan, P.-N. (2011). On ontology-driven document clustering using core semantic features. *Knowledge and information systems, 28*(2), 395-421.

Ghane'i-Ostad, M., Vahdat-Nejad, H., & Abdolrazzagh-Nezhad, M. (2018). Detecting overlapping communities in LBSNs by fuzzy subtractive clustering. *Social Network Analysis and Mining, 8*(1), 23.

Haase, P., Siebes, R., & Van Harmelen, F. (2004). *Peer selection in peer-to-peer networks with semantic topologies.* Paper presented at the International Conference on Semantics for the Networked World.

Hayashi, T., & Yoshida, T. (2018). *Development of a Tour Recommendation System Using Online Customer Reviews.* Paper presented at the International Conference on Management Science and Engineering Management.

Leal, F., González–Vélez, H., Malheiro, B., & Burguillo, J. C. (2017). *Semantic profiling and destination recommendation based on crowd-sourced tourist reviews.* Paper presented at the International Symposium on Distributed Computing and Artificial Intelligence.

Liu, H., Hu, Z., Mian, A., Tian, H., & Zhu, X. (2014). A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-Based Systems, 56*, 156-166.

Loh, S., Lorenzi, F., Saldaña, R., & Licthnow, D. (2003). A tourism recommender system based on collaboration and text analysis. *Information Technology & Tourism, 6*(3), 157-165.

Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems, 74*, 12 .٣٢-

Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., & Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science, 27*(4), .٦٨٤-٦٦٢

Neidhardt, J., Rümmele, N., & Werthner, H. (2017). Predicting happiness: user interactions and sentiment analysis in an online travel forum. *Information Technology & Tourism, 17*(1), 101-119.

Pantano, E., Priporas, C.-V., & Stylos, N. (2017). 'You will like it!'using open data to predict tourists' response to a tourist attraction. *Tourism Management, 60*, 430-438.

Pantano, E., Priporas, C.-V., Stylos, N., & Dennis, C. (2019). Facilitating tourists' decision making through open data analyses: A novel recommender system. *Tourism Management Perspectives, 31*, 323-331.

Rana, S., & Singh, A. (2016). *Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques.* Paper presented at the 2016 2nd International Conference on Next Generation Computing Technologies (NGCT).

Renjith, S., Sreekumar, A., & Jathavedan, M. (2020). An extensive study on the evolution of context-aware personalized travel recommender systems. *Information processing & management, 57*(1), 102078.

Smallwood, C. B. ,Beckley, L. E., & Moore, S. A. (2012). An analysis of visitor movement patterns using travel networks in a large marine park, north-western Australia. *Tourism Management, 33*(3), 517-528.

Tumas, G., & Ricci, F. (2009). Personalized mobile city transport advisory system. *Information and communication technologies in tourism 2009*, 173-183.

Vahdat-Nejad, H. (2014). Context-aware middleware: A review *Context in computing* (pp. 83-96): Springer.

Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment analysis in tripadvisor. *IEEE Intelligent Systems, 32*(4), 72-77.

Wan, L., Hong, Y., Huang, Z., Peng, X., & Li, R. (2018). A hybrid ensemble learning method for tourist route recommendations based on geo-tagged social networks. *International Journal of Geographical Information Science, 32*(11), 2225-2246.

Wei, T., Lu, Y., Chang, H., Zhou, Q., & Bao, X. (2015). A semantic approach for text clustering using WordNet and lexical chains. *Expert Systems with Applications, 42*(4), 2264-2275.

Wu, Z., & Palmer, M. (1 .(٩٩٤ *Verbs semantics and lexical selection.* Paper presented at the Proceedings of the 32nd annual meeting on Association for Computational Linguistics.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58*, 51-65.

Yeh, D.-Y., & Cheng, C.-H. (2015). Recommendation system for popular tourist attractions in Taiwan using Delphi panel and repertory grid techniques. *Tourism Management, 46*, 164-176.

Yochum, P., Chang, L., Gu, T., & Zhu, M. (2020). Linked Open Data in Location-Based Recommendation System on Tourism Domain: A Survey. *IEEE Access, 8*, 16409-16439.

Yochum, P., Chang, L., Gu, T., Zhu, M., & Zhang, W. (2018). *Tourist Attraction Recommendation Based on Knowledge Graph.* Paper presented at the International Conference on Intelligent Information Processing.

Zheng, X., Ding, W., Lin, Z., & Chen, C. (2016). Topic tensor factorization for recommender system. *Information Sciences, 372*, 276-293.

Zheng, X., Luo, Y., Sun, L., Zhang, J., & Chen, F. (2018). A tourism destination recommender system using users' sentiment and temporal dynamics. *Journal of Intelligent Information Systems, 51*(3), 557-578.