Modeling Individuals and Making Recommendations Using Multiple Social Networks

Makbule Gulcin Ozsoy, Faruk Polat
Department of Computer Engineering
Middle East Technical University, Ankara, Turkey
{makbule.ozsoy, polat}@ceng.metu.edu.tr

Reda Alhajj
Department of Computer Science
University of Calgary
alhajj@ucalgary.ca

***Abstract*—Web-based platforms, such as social networks, review web-sites, and e-commerce web-sites, commonly use recommendation systems to serve their users. The common practice is to have each platform captures and maintains data related to its own users. Later the data is analyzed to produce user specific recommendations. We argue that recommendations could be enriched by considering data consolidated from multiple sources instead of limiting the analysis to data captured from a single source. Integrating data from multiple sources is analogous to watching the behavior and preferences of each user on multiple platforms instead of a limited one platform based vision. Motivated by this, we developed a recommendation framework which utilizes user specific data collected from multiple platforms. To the best of our knowledge, this is the first work aiming to make recommendations by consulting multiple social networks to produce a rich modeling of user behavior. For this purpose, we collected and anonymized a specific dataset that contains information from BlogCatalog, Twitter and Flickr web-sites. We implemented several different types of recommendation methodologies to observe their performances while using single versus multiple features from a single source versus multiple sources. The conducted experiments showed that using multiple features from multiple social networks produces a wider perspective of user behavior and preferences leading to improved recommendation outcome.**

Keywords*—*Recommendation systems, Individual modeling, Multiple data sources, Social networks

I. Introduction

Recommendation systems estimate users' future preferences based on their historical information. Today, many different web-based platforms, such as social networks, review web-sites, and e-commerce web-sites, use recommendation systems to serve their users. For example, Imdb which is a movie review web-site that has a service called "Recommended for you" which gives movie recommendations to its registered users. Each of these kinds of platforms captures and uses its own information to model users' preferences [14]. However, considering only the data captured locally will lead to a limited perspective which cannot be used to provide better service. Instead, more guided and informative recommendations are possible by forming a wider perspective by integrating data from multiple sources.

In general, people tend to use different web-platforms for different purposes. For example, they prefer LinkedIn for professional connections, and Facebook for personal connections [19], though both are social networking platforms. It is possible to have more complete information about each user by integrating information from multiple social networks [32]. To realize this, it is possible to benefit from the reported research on identity resolution which tries to connect identities of a single person across social networks, e.g., [9], [14], [19], [24], [32]. Jain et al. [9] stated that the solutions to the identity resolution can be adapted by different application domains, such as security, privacy and recommendation systems. Actually, some research efforts described in the literature focused on giving recommendations across domains, e.g., [10], [23], [34]. However, these cross-domain recommendation systems focused on matching items and have not considered users' preferences across platforms. Our aim in this study is to combine information collected from multiple different social networks to create an integrated model of users' preferences which may form better basis for more guided and informative recommendations. To the best of our knowledge, this is the first work that has tackled this problem by shifting the analysis from a local limited perspective to a wider global perspective which integrates data from multiple sources.

For the evaluation process, we could not find an appropriate dataset because the existing datasets used in cross-domain recommendations consist of information on common items, but no information on preferences or behavior of users. Further, the datasets used in identity resolution have information on users, but not on items that they rate or interact with. To obtain information on users as well as the items they prefer, inspiring from [32] we collected information about users from the BlogCatalog website. In this website bloggers can publicly share information about their accounts in other websites. Using the shared account information, we collected information from Flickr and Twitter, whenever the information is publicly available. We anonymized the collected data for privacy concern, and then we used it for the evaluation.[1]

To summarize, contributions of the work described in this paper may be enumerated as follows:

- We integrated information from multiple social networking platforms to model "people", not only "users". We used the constructed model to give recommendations to the target people.
- We prepared a specific dataset that contains information collected from BlogCatalog, Twitter and Flickr. This dataset contains users who have accounts in all the three platforms and their preferences/interactions in each platform.

[1] We will continue to expand this data and we plan to share it for academic research.

- We implemented several recommendation methodologies to observe their performance. We mainly used collaborative filtering, multi-objective optimization based recommendation [22], hybrid recommendation and social-historical model [6] based recommendation methods.
- We compared the performance of the different recommendation methodologies on single feature versus multiple features and on a single source versus multiple sources. The results are reported in Section IV.

The rest of the paper is structured as follows: The collected and prepared multi-source dataset is described in Section II. The employed methodology is presented in Section III. The conducted experiments and their results are discussed in Section IV. Section V provides an overview of the related work. Section VI is conclusions and future work.

II. Multi-source Dataset

Today, most social networking platforms and e-commerce websites provide application programming interfaces (APIs) to researchers and developers in order to collect data from these platforms which allow users to make the information about themselves publicly available, or share it only with specific users or user groups. In some of these websites, beside sharing their personal information (e.g., nickname, real name, city, age, etc.), users can also share their account addresses on other platforms.

Inspiring from [32], we referred to BlogCatalog which is a web-site where users can publicly share their accounts on other web-sites and social networks in a section called “My Communities”. The shared accounts may exist on social networking platforms like Digg, Facebook, Flickr, Twitter, etc. The number of users from each social network are reported in Figure 1; there are three clusters of social networks based on the number of users. From the first cluster, which contains the social networks with many users, we used BlogCatalog and Twitter. We could not use Technorati since it is not in use anymore. From the second cluster, which contains social networks with average number of users, we preferred to use Flickr, since its API is easy to use and it permits the collection of public data without asking the permission of the target users. Since the number of users in the last cluster is limited, we decided not to use any of those social networks in this study. As a next step we plan to expand the dataset to cover other APIs. We collected the publicly available data from the selected platforms; namely BlogCatalog, Twitter and Flickr; on 19-20 February 2015.

Before going into the details of the data we collected, we want to give a brief explanation of the web-sites we are using. BlogCatalog is a blog advertisement web-site, which provides services to bloggers to share information about themselves and their blogs. It also provides services to readers to search for blogs of their interest. It groups the blogs by: (1) their categories, e.g., art, music, food, etc.; (2) bloggers’ cities, and (3) bloggers’ interests, e.g., movies, books, bands, etc. It provides discussion forums where bloggers can interact with each other. On each blogger’s page, a blogger can give brief explanation about himself/herself, information about his/her city, and information about his/her communities, i.e., his/her accounts on other platforms. Also on each of the blogger’s pages, the recent visitors of the page, the blogs she/he has, the followers, followees and reading list are available.

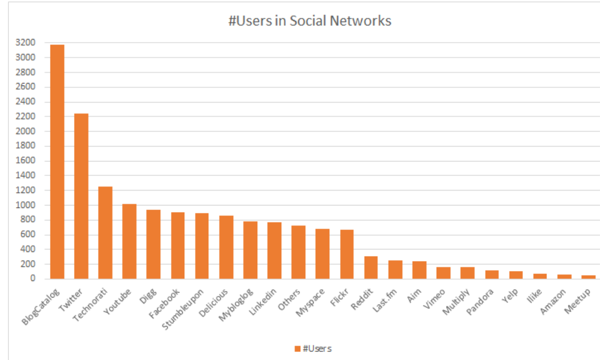

Fig. 1: Number of users on each social network

Twitter is a social networking web-site, in which users can share short messages (tweets) with the public. Registered users can send and read these messages and connect to each other. Each user has his/her own page, where he/she can give brief information on himself/herself, share his/her city, etc. On this page, also information on the tweets he/she has written, the followers and followees, favorites and lists are available. Twitter also gives information on what users mostly tweet about by trending topics service and it verifies the accounts to confirm that the account is not fake, usually for popular or notable people in politics, music, etc. Flickr is an image and video hosting web-site where users can share their photographs, label them with tags, titles and descriptions. Each user has a page which includes the photos he/she shares, the albums he/she created, and the photos he/she favored. The web-site also has groups feature, in which users can form groups, share their photos with others, and discuss subjects of their choice.

To prepare the multi-source based social networking dataset, first we found the mapping of user-ids across the selected platforms. Then, we collected the data about the common users using the APIs of the platforms. We collected only publicly available information. We anonymized the collected data to avoid privacy issues. From BlogCatalog, we collected the following information: (1) userids, (2) cities of the users, (3) regions of these cities, e.g., North America, Europe, etc., and (4) followers and followees of the users. From Twitter, we collected the following information (1) userids, (2) creation date of the account, (3) verification information of the account, (4) favorites count of the user, (5) friends count of the user, and (6) followers and followees of the user. From Flickr, we collected the following information: (1) userids, (2) first date of their photo sharing, (3) contacts of the users, (4) photos that the users favored, (5) number of views, favorites, comments and tags of those photos, and (6) groups that the user is member of and the count of members, photos and topics of those groups. We anonymized the data by assigning our own ids, which are unrelated to the ids assigned by the accessed websites.

From the BlogCatalog we collected information of 22291 users. However only 3179 of them explicitly indicate their accounts on other social networks. Among the BlogCatalog users, only 2187 of them publicly share their Twitter accounts and only 671 of them publicly share their Flickr accounts. Note that for some of the identified users it is not possible to collect any information, i.e. they close their accounts or they do not publicly share their information. There are 241 users who have accounts in all of the above mentioned three platforms and whose information is reachable. The 241 users selected from BlogCatalog indicated that they are from 66 different cities, which are located in 6 different regions. From the 241 users, 133 of the BlogCatalog users have followees and 156 of them have followers. However, only 31 of the followees and 13 of the followers have accounts in Twitter and Flickr. Regarding Twitter, 237 of the 241 users have followers and 234 have followees; 70 of these are followees and none of the followers have accounts in both BlogCatalog and Flickr. From Flickr, 160 of the 241 users have at least one contact. However, only 5 of the contacts are among the selected 241 users; 126 of the 241 Flickr users are members of at least one group, 123 of them are also among the ones who have at least one contact. The total number of groups in the produced dataset is 4802. Finally, 105 of the 241 users have at least one favorite photo, and the total number of distinct photos favored by a member is 5067. These photos have 17611 different tags in total.

We conjecture that the constructed dataset can be used for several different purposes; such as tag prediction, item recommendation, link prediction, identity prediction and location prediction. For example, the behavior of a user in a single network or multiple social networks can be used to predict his/her hometown. Further, this information can be used by researchers and practitioners working on recommendation systems, privacy and security, among other domains.

III. Recommendations Using Multiple Data Sources

After having the dataset, the next step is to set the objective of the recommendation system in order to decide on the social network(s) and features to be used. Obviously, the diversity and richness of the available information increase the number of alternatives which could be considered in developing a recommendation system. For this paper, we set the objective as making recommendations to Flickr users regarding the groups they may join in the future.

There are various types of recommendation methodologies described in the literature. In this work, we used collaborative filtering, multi-objective optimization based, hybrid and social-historical model based recommendation methods to observe the effect of using data from multiple data sources.

Collaborative filtering based recommendation: We used user-based collaborative filtering. In this approach, the similarities among users are calculated and the most similar users to the target user are assigned as neighbors. Then neighbors' past preferences are used to give a recommendation to the target user.

Multi-objective optimization based recommendation: Here neighbors are determined by employing our multi-objective optimization based method proposed in [22], which uses Pareto dominance for recommendation. Then, an approach similar to collaborative filtering is followed to give recommendations by incorporating neighbors' past preferences. The neighbor selection process in this method proceeds as follows: first user-user similarities are calculated separately, and then these similarities are used to decide on non-dominated users who are assigned as the neighbors.

Hybrid recommendation: We used an item based hybrid approach which combines the output of different recommendation methods. Several different techniques for hybridization are explained in [1]. In this work we decided to combine only collaborative filtering methods. We combined and ranked the recommended items by the number of votes they received.

Social-historical model based recommendation: We used the method proposed in [6]. This method was actually proposed for check-in data. It models check-ins by a language model. The authors also considered the friendship among users in their model. We can divide this model into two modules, namely a historical module when it uses historical preferences of users only, and a social module when it uses only the relationship among users.

A variety of features from different social networks, namely Flickr, Twitter and BlogCatalog, are available in the dataset we collected. We used only a subset of the data in order to demonstrate the effectiveness of making recommendations by considering data from multiple sources. It is possible to add new features in future experiments. The features used in this study and their source social network are explained next:

Flickr groups: The source social network of this feature is Flickr, where users can join different groups depending on their interests. Having the knowledge about historical preferences of users, including group(s) that they have joined before, a recommendation system can predict groups that these users may join in the future.

Flickr contacts: The source social network of this feature is Flickr., where users can connect with other members who they already know or with whom they share similar interests. Knowing the contacts' past preferences can be helpful to recommend new groups to the users, as they most probably share similar interests.

Flickr common contacts: This feature is similar to Flickr contacts. However, this feature uses the common contacts information. Even though two users are not directly connected, their common contacts may indicate that they have similar preferences. This information can be used to give future recommendations to the target users.

Twitter followees: The source social network of this feature is Twitter, where users follow other users who they already know (e.g., friends, family members, etc.) and who they like, admire or support (e.g., political leaders, singers, etc.). Having common followees may indicate that two users are similar to each other, and this can be used to make recommendations.

BlogCatalog followees: Similar to Twitter followees, users in BlogCatalog follow other users if they think the followed person has interests similar to their own interests. Information from other users who have similar interests can be used to give recommendations to the target users.

For all the selected features, except Flickr contacts, we calculated the user-user similarity using the Cosine Similarity measure. For Flickr contacts, the similarity between the target user and his/her contacts is assigned the value 1.0, and for others the value assigned is 0.0.

IV. EVALUATION

In order to evaluate the performance of the methods we used precision@k, recall@k and F1-measure, which are commonly used in the recommendation and search literature. We calculated the measures for an output list that contains k elements. These measures are computed as given in Equations 1, 2 and 3.

$$Precision_k = \frac{tp_k}{tp_k + fp_k} \tag{1}$$

$$Recall_k = \frac{tp_k}{tp_k + fn_k} \tag{2}$$

$$F1 - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

In the above equations, tp refers to true positives, i.e., recommended and actually used items, fp is false positives, i.e., recommended but actually not used items, and fn indicates false negatives, i.e., not recommended but actually used items. Here it is worth mentioning that it is common for recommendation methods to have low precision results as the data is very sparse. For instance, in [33] the authors gave several examples of low precision results, which are in the range [0.030, 0.035], for different datasets.

Besides these metrics, we calculated the hit-rate of the invoked methods. Hit-rate is the ratio of the users who are given at least one true recommendation. The average precison@k value of a method can be high even though it is able to make recommendations just to a few users. For example, assume that we have two different recommendation methods, RM_1 and RM_2, two users u and v, and the output list size is 3. Consider the case where RM_1 gives 2 true recommendations to user u and no true recommendation to user v, and RM_2 gives one true recommendation to each user. Both methods' precision@k will be 0.33, on average. However, RM_2 can give true recommendations to both users; this means we can say that it is better than RM_1. Hit-rate is calculated by Equation 4.

$$HitRate = \frac{\sum_{m \in M} HitRate_m}{|M|} \tag{4}$$

where M is the set of target users, m is one of those users, and $HitRate_m$ is a number whose value is set to 1.0 if the output list contains at least one true recommendation and to 0.0 otherwise.

It is obvious that we need separate training and test sets for the evaluation. Since our objective is to give Flickr group recommendations, we divided the set containing Flickr groups information into two disjoint subsets, one for training and the other for testing. In the original dataset there are 126 Flickr users where each of them is a member of at least one group. From these users, we selected the ones who have at least 5 group memberships; we removed 20% of their memberships from the training set. The ones removed were selected randomly and were used in the testing phase. At the end, we were left with 126 users in the training set and 86 users in the test set. On average, these users are members of 56.008 groups for the training set and 12.628 groups in the test set.

We evaluated several different methods with a variety of features from the dataset we collected. In Table I, we present the list of methods together with the used features and their abbreviations. This information will be used in the rest of this paper.

TABLE I: The abbreviations used in this study

Methods	Abbreviation
Collaborative filtering	CF
Multi-objective optimization	MO
Hybrid	HI
Social-historical	SH
Features	Abbreviation
Flickr Groups	FG
Flickr Contacts	FC
Flickr Common Contacts	FCC
Twitter followees	TF
BlogCatalog followees	BCF

Two variables need to be assigned in the experiments. These are neighbors count (N) and the output list size (k). The performance of the methods may differ based on these parameters. We first started with some arbitrary values of these parameters and decided on the method that performs the best. Afterwards, we decided on the best values of N and k using only the selected method. To be fair to the other methods, lastly, we performed the analysis on all the methods using the determined N and k values.

Five (5) is the first arbitrary value assigned to both N and k. We did not want to assign a larger value to N because we only have 126 users (one target and 125 candidate neighbors) where each user is a member of at least one group. We assigned to k the value 5 based on experience because we observed in our daily life that most recommendation systems prefer to present a small number of items as recommendations to their users. The evaluation results of the methods with our initial assigned values of N and k are shown in Figure 2, where the method and features combination are reflected in the form $Method - Feature1\_Feature2\_Feature3$. For example, the combination $HI - FG\_FCC\_TF\_BCF$ refers to the Hybrid method combined with the features Flickr groups, Flickr common contacts, Twitter followees and BlogCatalog followees. Also, instead of giving the hit-rate directly, it has been scaled to its 10% in order to have a better representation of the other metrics.

According to Figure 2, instead of using the previously joined groups, using the common contacts in collaborative filtering performs better for group recommendation. Using the contacts directly did not report good performance, since in the dataset only a limited number of contacts are members of at least one group. Actually, only 5 users who are seen as contacts are actually members of at least one group. Using external social networks only, such as CF_BCF or CF_TF, did not show better performance compared to using the data available in the same network. This can be explained by the fact that people use different web-platforms for different purposes [19] and may behave differently in different social networks. Using the Social-Historical model [6] did not show good performance when the output list size is set to 5. The hybridization of item recommendations showed the best performance when Flickr groups, Flickr common contacts and BlogCatalog followees are used altogether. This indicates that using multiple features from multiple sources can lead to better recommendations than using a single feature. Our multi-objective optimization based recommendation method [22] performed the best when Flickr groups, Flickr common contacts or when additionally Twitter followees are used together. The hit-rate performance of the utilized methods follows the precision results. According to Figure 2, the best performing method is $HI-FG\_FCC\_BCF$; it is actually used to decide on the best values of the two parameters N and k.

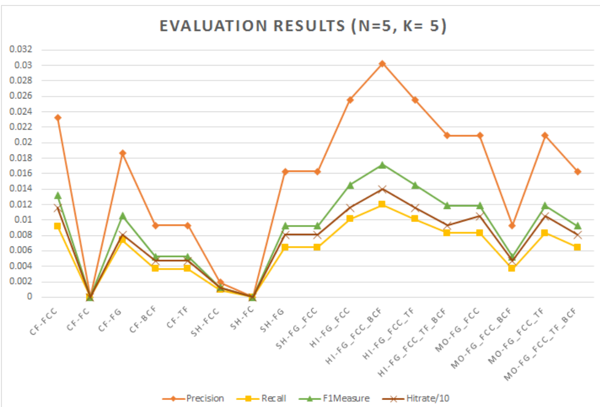

Fig. 2: Evaluation results for N=5 and k=5

We decided on the best value of N by considering the range [1,62] with 1 increment. We stopped at 62 neighbors at most since this is half of the candidate neighbors. We used the method $HI-FG\_FCC\_BCF$ only because it is the best performing method we found in the previous experiment. We kept the value of k set to 5 as in the previous experiment. The evaluation results for different N values are given in Figure 3 which shows that the performance of the algorithm deviates based on neighbors count. If the method uses only few neighbors, it cannot make good recommendations. Similarly, assigning many users as neighbors produces poor performance because they may introduce noise to the methods. Besides depending on the method, choosing many neighbors may cause time and memory problems. For example, in collaborative filtering based algorithms the candidate items are selected by looking at the historical preferences of each neighbor. If there are many neighbors, the number of candidate items will be high, and this will lead to longer computations. Interestingly, the hit-rate performance remains nearly the same for most of the neighbors count, except few spikes. This may indicate that independent of neighbors count, predicting at least one true recommendation is possible, however a tuned number of neighbors can help to increase the number of true predictions for each user. Based on Figure 3, 16 has been determined as the best performing N value for this dataset; this value actually provides the best precision, recall and f1-measure performance.

After deciding on the value of N, the next step is to decide on the value of k, the output list size. Using the method

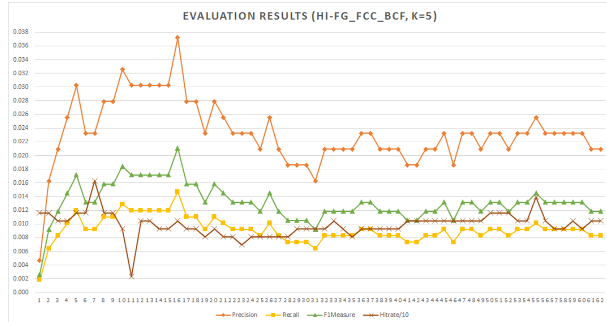

Fig. 3: Evaluation results for different values of N

$HI-FG\_FCC\_BCF$ and setting N to 16, we searched for the best value of k in the range [1,30] with increments of 1. We set the upper bound as 30 items because it is usual for recommendation platforms to present smaller number of recommendations. Also, it is known that in search engines usually one outcome is composed of 15 results and users tend to select the results in the first few reported pages. The evaluation results for the best k value are presented in Figure 4. The results show that for different measures, different values of k perform better. The best value of k for precision is 4, it is 27 or 28 for recall, and it is 12 for F1-measure. As expected, when the value of k increases, precision decreases but recall increases, since larger number of recommendations are presented. The hit-rate performance follows a pattern similar to recall, since it is expected that both of these measures perform better as k increases.

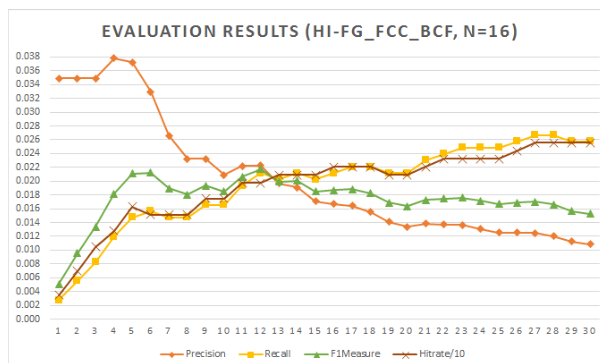

Fig. 4: Evaluation results for different values of k

After deciding on the best values of N and k, we invoked all the methods with these selected values of the two parameters. We discarded the methods that use Flickr contacts, since in Figure 2 we observed that this information does not provide any successful recommendations. Instead of restricting the experiments to a single value of k, we decided to use all values of k that contributed to the best performance for different measures. So we fixed the value of N at 16 and used three values of k, namely 4, 12 and 27. The evaluation results are presented in Figures 5, 6 and 7. When the value of k is 4, which is the value that produced the best precision in the previous experiment, the best performing methods are collaborative filtering using Flickr groups only and the hybrid method using different combinations of Flickr groups, Flickr common contacts, Blog Catalog followees, and Twitter followees. Unlike the results shown in Figure 2, with this setting, collaborative filtering using only Flickr groups performed better. This shows that the number of neighbors and the output list size are important variables and should be tuned carefully. According to the hit-rate performance, the best performing method is the hybrid method using Flickr groups, Flickr common contacts, and Twitter followees. Together with the results of precision, recall and F1-measure, hit-rate results confirm that using data from multiple sources (e.g., social networks) improves the recommendation performance. When we set k to 27, recall increases but precision decreases, as expected. According to Figure 6, the best performing method is based on the social-historical model that used Flickr groups information. In this figure, we observed similar patterns of hit-rate and recall for all the methods. However the hit-rate results of the best method and the worst method do not differ as much as it is the case for recall results. This shows that some of the methods are able to give better recommendations for certain users as k increases. When we set k to 12, which is the value that produced the best F1-measure in the previous experiment, again the best performing method is the social-historical model that used Flickr groups information only. The best hit-rate result was also reported by the same method.

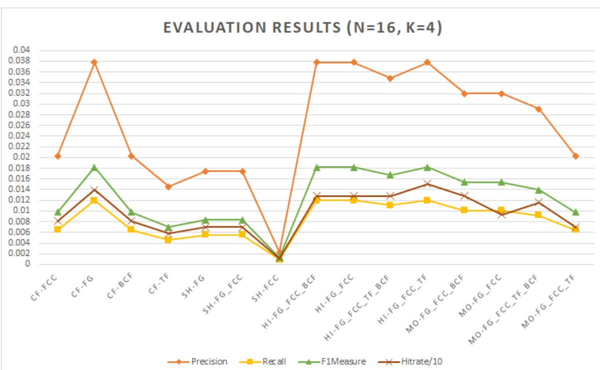

Fig. 5: Evaluation results for N=16 and k=4

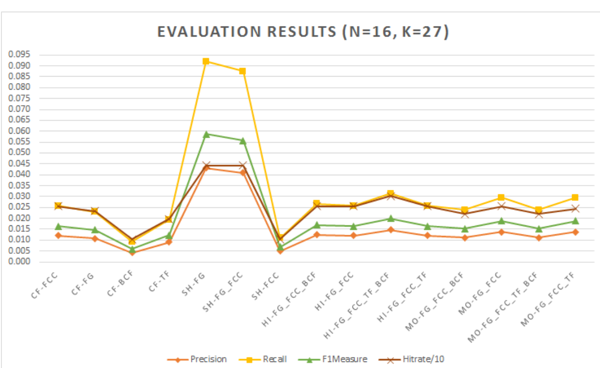

Fig. 6: Evaluation results for N=16 and k=27

In real life, recommendation is a never ending process, such that users always expect new suggestions when they come back to the system. The environment is so dynamic that

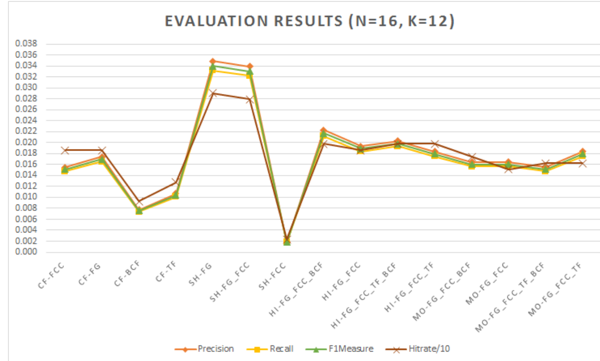

Fig. 7: Evaluation results for N=16 and k=12

same recommendations are not guaranteed to repeat because the process is data analysis driven and the data underlying a recommendation system is subject to change continuously. It is more important to make true recommendations to as many users as possible and present in the limited length output list one or more items that the user is expected to use in the future. Based on our analysis, for shorter lists and when precision and hit-rate are considered more important, the best performing method is the hybridization method that combines information from multiple features from multiple social networks. This way the method can model its users with other aspects which are not obvious for a single social network.

V. Related Work

Recommendation systems recommend items by estimating preferences of the target users [25]. Traditional recommendation systems consider only past preferences/ratings of users, even though many other kinds of information are available in a variety of sources, e.g., social networking platforms. Another trend that can be used for better recommendations is to model users' interests on other domains and use that in the target domain; this is known as cross-domain recommendations. Most cross-domain recommendation methods described in the literature are based on item matches. Alternatively, mapping identities across domains can be useful to figure out how users behave in different domains and use this information in making recommendation.

Some of the recent works described in the literature combine historical preferences of users with location, social network, and time information. It is shown that users tend to visit locations periodically [3] and their behavior differs depending on the hour of the day (daytime vs. night) and day of the week (weekdays vs. weekend)([3], [4], [16], [20]). The works described in [30], [16], [7], [5], [31] and [26] use this observation to give time-aware recommendations. Besides temporal information, location and social network information are used by many recommendation methods. Based on a conducted analysis, it is stated in [12] that recommendations should be given by users who are living in the same or similar regions and recommended items should be close to the target user. LARS [12] [35] [29] [2] are some examples of systems that use location to improve the performance of their recommendation process. The work described in [15] stated that the quality of recommendations can be improved by using friendship relationships among users. The literature, e.g., [6], SoCo [15], [28] and [18] includes some examples that use social relations in the recommendation process. In order to combine multiple criteria in the recommendation process, multi-objective optimization methods can also be used, e.g., [11], [21] and [22].

Research efforts which addressed cross domain recommendations mostly focused on giving recommendations across domains, e.g., [27], [23], [34], [10], [8] and [17]. These cross-domain recommendation systems focus on item matches and do not consider users' identities or assume different categories, such as books and movies, as different domains and use data from a single source. One of the first initiatives that focused on cross-domain recommendations is described in [27]. In this work, information from users was directly collected by asking them to give some category names and rating them. They analyzed the data at group and individual levels. As a result, they showed that using multiple information sources for recommendations is promising. In [23] the authors used a Bayesian hierarchical approach based on Latent Dirichlet Allocation (LDA) to model users' interests and objects' topics. The produced model was used to find out the correlation between objects. The correlations are then used to give recommendations to the target users based on their interests. The work described in [34] focused on giving browser-oriented recommendations across the web-sites by browsing information of the users. Even though this idea might look initially similar to ours, it uses browser history of the users which may not be always available. The work described in [10] used textual information of items to map them across domains. Then the related information is used for the recommendations. The work described in [8] modeled the user-item-domain relationship with the assumption that users behave similarly across domains. They tested their approach on books and movies datasets obtained from Amazon. The authors of [17] modeled users' preferences separately on each domain and based on types of item, and then combined them with the help of factorization machines. The work described in [13] aimed to identify user and item mapping across the rating matrices and to use the mapping in the recommendation process. Their method does not need to know any user or item mapping beforehand. They assumed similar rating behavior of users on both domains, and there were some overlapping users/items. They evaluated their method on a synthetic dataset and Yahoo! Music dataset.

There are several works in the literature which aimed to connect identities across social networks, namely identity resolution, such as [19], [32], [9] and [24]. These algorithms mainly focus on mapping users across domains, but not on their preferences or interactions with the related social network, i.e., they do not make any recommendation. The work described in [19] searches and matches users across online social networks. For matching purposes, they used several different attributes of the users, such as age, gender, location, country and name. The authors of [32] mapped individuals across social media sites. Their method (MOBIUS) first identifies users' unique behavior patterns, such as using similar names or typing patterns, then constructs features based on the captured behavior, and lastly uses machine learning to identify users. The work described in [9] improved profile attribute based identity resolutions by additionally using content and network features. They applied their method to map users in Facebook to users in Twitter. They concluded that using different attributes provides distinct aspects of the identity of the users, and helps to improve the performance of the identity resolution process. Finally, [24] proposed a semi-supervised manifold alignment method, called Manifold Alignment on Hypergraph (MAH), to map users across social networks. They used social structures, but the authors stated that it is possible to boost the performance by integrating the names of users.

VI. Conclusion and Future Work

Recommendation systems are very commonly used by today's web-based platforms; such as social networks, review web-sites, e-commerce web-sites, etc. The main target is to serve and make recommendations to interested users. Each of these platforms base its recommendations on the local information captured by the Website; only this information is used to model their users [14]. However, by restricting the analysis merely to locally captured information these platforms may miss some vital information about individuals who are using their services. It is more beneficial and rewarding to consider information from multiple sources because it is known that people tend to use different web-platforms for different purposes, e.g., LinkedIn for professional and Facebook for social connections [19]. In other words, to have more complete information about each user, it is essential to consider integrated information from multiple social networks [32].

In this work, we combined information collected from multiple different social networking platforms to create integrated model of individuals and to give recommendations to them. To the best of our knowledge, this is the first work aiming to use integrated information from multiple social networking platforms in the recommendation process. We integrated information from multiple social networks to model "people" not only "users". For this purpose, we collected a dataset that contains information collected from BlogCatalog, Twitter and Flickr web-sites. This dataset contains users who have accounts in all the three websites and their preferences/interactions in each website. We used this dataset to give recommendations to the target people. We implemented several different types of recommendation methodologies to observe their performance. We compared the performance of these recommendation methodologies while using single versus multiple features from a single versus multiple sources. The conducted experiments showed that using multiple features from multiple sources improved the recommendation performance.

As future work, we want to integrate identity resolution methods into our work and produce an end-to-end recommendation system. We also want to use more features to be captured from other social networking platforms not covered in this paper. Finally, we want to try some other recommendation methods to observe their effectiveness while using a multi-source dataset .

Acknowledgment

This work is supported by TUBITAK-BIDEB 2214/A program.

REFERENCES

[1] R. D. Burke, "Hybrid recommender systems: Survey and experiments," *User Model. User-Adapt. Interact.*, vol. 12, no. 4, pp. 331–370, 2002.

[2] C. Cheng, H. Yang, M. R. Lyu, and I. King, "Where you like to go next: Successive point-of-interest recommendation," in *IJCAI 2013, 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, 2013.

[3] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services." in *ICWSM*, L. A. Adamic, R. A. Baeza-Yates, and S. Counts, Eds. The AAAI Press, 2011.

[4] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '11. New York, NY, USA: ACM, 2011, pp. 1082–1090.

[5] H. Gao, J. Tang, X. Hu, and H. Liu, "Exploring temporal effects for location recommendation on location-based social networks," in *7th ACM conference on Recommender systems*. ACM, 2013, pp. 93–100.

[6] H. Gao, J. Tang, and H. Liu, "Exploring social-historical ties on location-based social networks," in *Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012*, 2012.

[7] B. Hu, M. Jamali, and M. Ester, "Spatio-temporal topic modeling in mobile social media for location recommendation," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 1073–1078.

[8] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu, "Personalized recommendation via cross-domain triadic factorization," in *22Nd International Conference on World Wide Web*, ser. WWW '13. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2013, pp. 595–606.

[9] P. Jain, P. Kumaraguru, and A. Joshi, "@i seek 'fb.me': Identifying users across multiple online social networks," in *22Nd International Conference on World Wide Web Companion*, ser. WWW '13 Companion. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2013, pp. 1259–1268.

[10] A. Kumar, N. Kumar, M. Hussain, S. Chaudhury, and S. Agarwal, "Semantic clustering-based cross-domain recommendation," in *Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on*, Dec 2014, pp. 137–141.

[11] H. Lee and W. Teng, "Incorporating multi-criteria ratings in recommendation systems," in *IEEE International Conference on Information Reuse and Integration, IRI 2007, 13-15 August 2007, Las Vegas, Nevada, USA*, 2007, pp. 273–278.

[12] J. J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel, "LARS: A location-aware recommender system," in *IEEE 28th International Conference on Data Engineering (ICDE 2012), Washington, DC, USA (Arlington, Virginia), 1-5 April, 2012*, 2012, pp. 450–461.

[13] C.-Y. Li and S.-D. Lin, "Matching users and items across domains to improve the recommendation quality," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '14. New York, NY, USA: ACM, 2014, pp. 801–810.

[14] H. Liu and P. Maes, "Interestmap: Harvesting social network profiles for recommendations," *Beyond Personalization 2005*, p. 54, 2005.

[15] X. Liu and K. Aberer, "Soco: a social network aided context-aware recommender system," in *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*, 2013, pp. 781–802.

[16] X. Liu, Y. Liu, K. Aberer, and C. Miao, "Personalized point-of-interest recommendation by mining users' preference transition," in *22Nd ACM International Conference on Conference on Information Knowledge Management*, ser. CIKM '13. New York, NY, USA: ACM, 2013, pp. 733–738.

[17] B. Loni, Y. Shi, M. Larson, and A. Hanjalic, "Cross-domain collaborative filtering with factorization machines," in *Advances in Information Retrieval - 36th European Conference on IRResearch, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings*, 2014, pp. 656–661.

[18] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011*, 2011, pp. 287–296.

[19] M. Motoyama and G. Varghese, "I seek you: Searching and matching individuals in social networks," in *Eleventh International Workshop on Web Information and Data Management*, ser. WIDM '09. New York, NY, USA: ACM, 2009, pp. 67–75.

[20] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "Mining user mobility features for next place prediction in location-based services," in *2012 IEEE 12th International Conference on Data Mining*, ser. ICDM '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 1038–1043.

[21] F. Ortega, J. L. Sánchez, J. Bobadilla, and A. Gutiérrez, "Improving collaborative filtering-based recommender systems results using pareto dominance," *Inf. Sci.*, vol. 239, pp. 50–61, 2013.

[22] M. G. Ozsoy, F. Polat, and R. Alhajj, "Multi-objective optimization based location and social network aware recommendation," in *10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom 2014, Miami, Florida, USA, October 22-25, 2014*, 2014, pp. 233–242.

[23] S. Tan, J. Bu, X. Qin, C. Chen, and D. Cai, "Cross domain recommendation based on multi-type media fusion," *Neurocomput.*, vol. 127, pp. 124–134, Mar. 2014.

[24] S. Tan, Z. Guan, D. Cai, X. Qin, J. Bu, and C. Chen, "Mapping users across networks by manifold alignment on hypergraph," in *Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.*, 2014, pp. 159–165.

[25] M. Tavakolifard and K. C. Almeroth, "Social computing: an intersection of recommender systems, trust/reputation systems, and social networks," *IEEE Network*, vol. 26, no. 4, pp. 53–58, 2012.

[26] C.-Y. Tsai and B.-H. Lai, "A location-item-time sequential pattern mining algorithm for route recommendation," *Knowledge-Based Systems*, vol. 73, pp. 97–110, 2015.

[27] P. Winoto and T. Tang, "If you like the devil wears prada the book, will you also enjoy the devil wears prada the movie? a study of cross-domain recommendations," *New Generation Computing*, vol. 26, no. 3, pp. 209–225, 2008.

[28] M. Ye, P. Yin, and W.-C. Lee, "Location recommendation for location-based social networks," in *18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. New York, NY, USA: ACM, 2010, pp. 458–461.

[29] M. Ye, P. Yin, W. Lee, and D. L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July 25-29, 2011*, 2011, pp. 325–334.

[30] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Time-aware point-of-interest recommendation," in *36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '13. New York, NY, USA: ACM, 2013, pp. 363–372.

[31] Q. Yuan, G. Cong, and A. Sun, "Graph-based point-of-interest recommendation with geographical and temporal influences," in *23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 2014, pp. 659–668.

[32] R. Zafarani and H. Liu, "Connecting users across social media sites: A behavioral-modeling approach," in *19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '13. New York, NY, USA: ACM, 2013, pp. 41–49.

[33] J. Zhang and C. Chow, "Ticrec: A probabilistic framework to utilize temporal influence correlations for time-aware location recommendations (to be published)," *Services Computing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[34] Y. Zhang, "Browser-oriented universal cross-site recommendation and explanation based on user browsing logs," in *8th ACM Conference on Recommender Systems*, ser. RecSys '14. New York, NY, USA: ACM, 2014, pp. 433–436.

[35] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Collaborative location and activity recommendations with GPS history data," in *19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, 2010, pp. 1029–1038.