Recommendation Framework for Online Social Networks

Przemysław Kazienko[1] and Katarzyna Musiał[1,2]

[1] Wrocław University of Technology, Wyb.Wyspiańskiego 27, 50–370 Wrocław, Poland kazienko@pwr.wroc.pl
[2] Blekinge Institute of Technology, S–372 25 Ronneby, Sweden weaving@wp.pl

The recommendation framework that supports the creation of new interpersonal relationships within the social networks is presented in the paper. It integrates many sources of data in order to generate the relevant personalized recommendations for network members. The unique social filtering techniques and measures of the activity and strength of relationship are encompassed by the framework.

1 Introduction

Recommender systems became an important part of recent web sites; the vast numbers of them are applied to e–commerce. They help people to make decisions what items to buy, which news to read [17] or which movie to watch. Recommender systems are especially useful in environments with information overload since they cope with selection of a small subset of items that appears to fit to the users preferences [2, 12, 15]. Furthermore, these systems enable to maintain the loyalty of the customers and increase the sales [10]. However, not only products or multimedia content can be suggested to users. The new area where recommender systems can be applied are various online communities called social networks that rapidly develop in the web and usually have thousands or even millions of members like Friendster or LinkedIn. The main goal of a recommendation system, in this case, is to help the user to establish new relationships and in consequence to expand the human community.

In general, there are three main approaches to recommendation: collaborative filtering, content–based filtering, and hybrid recommendations [2]. The collaborative filtering technique relies on opinions about items delivered by users. The system recommends products or people that have been positively evaluated by other people, whose ratings and tastes are similar to the preferences of the user who will receive recommendation [2, 5, 17]. There are two main variants of collaborative filtering. The first one is the k–nearest neighbour and the second one is the nearest neighbourhood. In the content–based filtering the items that are recommended to the user are similar to the items that the user had liked previously [13]. The hybrid method combines two previously enumerated approaches [8, 10, 17].

The proposed recommendation framework is supposed to be applied to the social networks that have recently become more and more important element of information society [1, 6]. A social network is the set of the actors (a single person is the node of the network) and ties, called also relationships, that link the nodes [1, 4]. The evolution of the social network depends on the mutual experience, knowledge, relative interpersonal interests, and trust of human beings [3, 14]. The measurements can be collected to investigate the number and the quality of the relationships within the network.

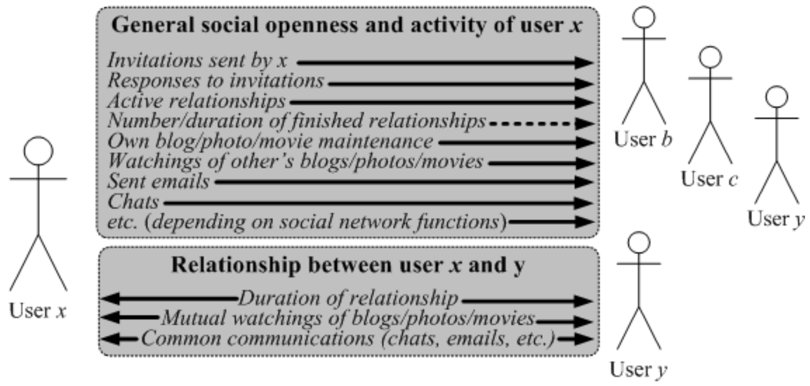

Fig. 1. Social features of relationships in a social network.

2 Problem Description

Recommender systems for social networks differ from typical kinds of recommendation solutions since they suggest rational human beings to other ones rather than inanimate goods. After recommendation selection, one person initializes the relationship with another one, and the latter can respond positively or negatively to the invitation. Such an interaction is impossible with products or content. Moreover, the bond between people is bidirectional in opposite to the relationship between a person and an object. Thus, we would like to find for the current user another person who would also like to react in a favourable way. Things possess no free will and cannot refuse to be sold. For that reason, the recommender systems for social networks need to respect preferences and human limitations of people who would be suggested. In conclusion, the recommendation system should suggest to user x only those network members, who would be potentially good friends or co–workers for user x. In consequence, new relationships of user x can appear in the network.

On the other hand, the owners of the online social network system can execute their own policy. Hence, their first goal can be to build the widest social network, with as many connections between individuals as possible. Yet another possible purpose would be to create the network in which the relationships reflect the strong similarity between people and in consequence, the network consists of many close groups. This evolution of the social network can be stimulated by different kinds of recommendation. In both cases, the aim is to achieve the community in which the connections between human beings are permanent. The common purpose of the framework proposed below is to enable the adjustment of recommendations to the profile of the particular user as well as to the general policy of the network.

3 Recommendation Framework for Social Network

3.1 Relationships between People

Before building the recommendation framework, let us point the main sources of data useful for the recommendation. In the social networks, not only the typical information about a particular user like their interest, demographic data, etc. can be considered, but also their activities and some measures of relationships with other users, especially those related to the process of initialization of bounds and some further, successive activities. The social statement of the user x in the network consists of two data sets: general, aggregated openness and activities features of this user in relation to all others, also in the past, and measures of the relationship between user x and other members of the network (Fig. 2). The most significant elements in the first set are the user's willingness for initialization of relationships and responding to invitations from others. The most important issue within the second set is frequency and intensity with which all relationships are maintained.

3.2 User Profile

All the data related to the characteristic of the individual user is called a user profile and consists of two main parts: static and dynamic. The users themselves deliver the former by filling in the special forms while the latter is monitored and gathered by the system. Based on the analysis of the existing online communities we split the user profile into a set of components which can be easily extended in the future. Each of the components consists of several separate attributes (Fig. 3). The *preferred* component is the direct hint for the system about people sought by the user. The *search* component includes the data about all searches made by the user within the network. The two most complex components are *activity* and *relationship*. The first one measures the activity of the user within the community. The relationship profile describes the number and duration of the users relationships and some other features that characterize them.

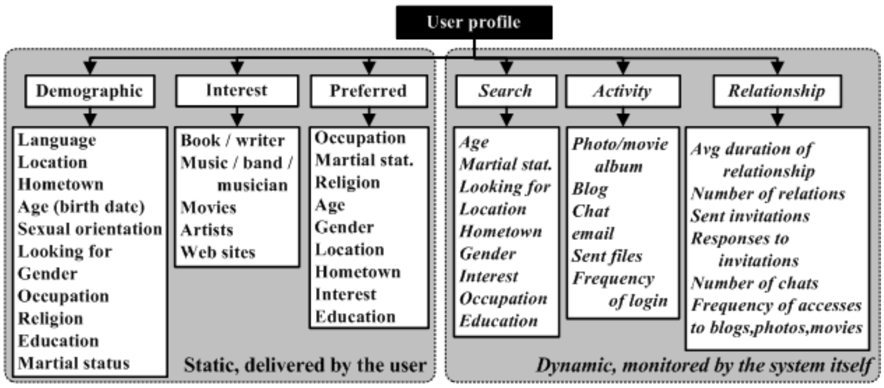

Fig. 2. User profile components. *Location* - the place of residence, *Hometown* - the place of origin, *Looking for* - the purpose of inviting new friends, e.g. friendship.

3.3 Recommendation Process

Based on the gathered information we have built the recommendation framework that supports the creation of the recommendation for the social network. The main goal of the system is to provide the most relevant recommendations to users. Moreover, by combining the several, different sources of data, the method facilitates a new user to join the network and satisfy their expectations.

The overall view of the recommendation process for the social network is presented in Fig. 4. Firstly, knowledge about the users is gathered. Next, the system calculates final similarity function $r(x \Rightarrow y)$ for each pair of users (see below). The final similarity consists of four elements: the direct similarity derived directly from users static attributes, the complementary of relationship initialization, the general activity measure of a candidate for recommendation (user y), and the strength of relationships maintained by user y. The complementary of relationship matches the will of initiation of relationships for user x with the willingness of user y to respond the invitation. Calculation of $r(x \Rightarrow y)$ is periodically repeated due to possible changes in source data according to the network policy. In the next step, static list L of the users that match to particular person x is created when user x logs into the system. Based on user profiles, especially their static components, all users are clustered in separate groups using any of clustering algorithms [16]. Next, depending on the strategy of the evolution and aim of the network, the connectivity social filtering is utilized to promote the creation of connections either within the same group or between the groups. The people from the same group will be recommended if the strong groups are supported, while cross–group recommendations are created to flatten the social network.

In addition, some research revealed that the number of stable relationships that one human being can maintain is about 150. This number called Dunbar

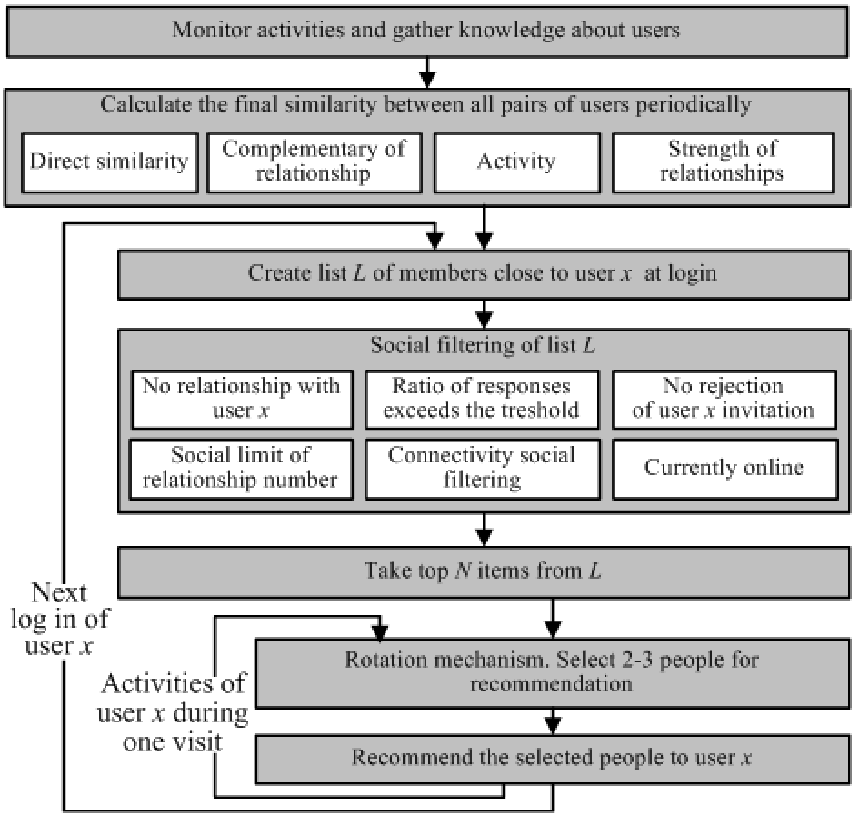

Fig. 3. The process of recommendation for user x in a social network

Number [7] defines the social limit of relationships number within the proposed network. For that reason, the system should not recommend members who have already had more than 150 relationships.

Additionally, the system, within its social filter, promotes people who are currently online to enable the possibly quick response to the invitation. Due to efficiency limits, only top N most suitable users are fixed for recommendation to user x during their stay in the network. Finally, only 2–3 selected people are suggested to person x. The rotary mechanism is used, to prevent the same people to be recommended to user x all the time.

3.4 Final Similarity between Users

The goal of the recommendation process is to find out whether person y ought to be recommended to person x. It is achieved by using the final similarity function $r(x \Rightarrow y)$:

$$r(x \Rightarrow y) = \alpha \cdot s(x, y) + \beta \cdot c(x, y) + \gamma \cdot a(y) + \delta \cdot sr(y) \qquad (1)$$

where: $s(x, y)$ – the direct similarity between user x and y that is derived from the comparison of all attributes from *demographic*, *interest*, *preferred*, and *search* user profile components; $c(x, y)$ – the complementary of relationship initiation function describes the social behaviour of the users in the context of established relationships; $a(y)$ – the activity of user y is calculated basing on the information included in *activity* component; $sr(y)$ – the strength of the relationship between person y and all other members; it is derived from *relationship* user profile component; $\alpha, \beta, \gamma, \delta$ – importance coefficients with values from the range $[0, 1]$.

Coefficients are used to simulate and adjust the evolution of the social network. For example, if α is low and β is high then the goal of the network is to build the wide network regardless the mutual direct similarity of users. The values of these factors tightly depend on the social policy.

Since values of all four components are from the range $[0, 1]$, the value of final similarity function $r(x \Rightarrow y)$ belongs to the range $[0, 4]$.

3.5 Direct Similarity

The direct similarity $s(x, y)$ between two users x and y is the function that compares the set of attributes that characterize these members of the network:

$$s(x, y) = \frac{w_1 \cdot f_1(x, y) + w_2 \cdot f_2(x, y) + \ldots + w_n \cdot f_n(x, y)}{w_1^{\max} \cdot f_1^{\max}(x, y) + w_2^{\max} \cdot f_2^{\max}(x, y) + \ldots + w_n^{\max} \cdot f_n^{\max}(x, y)} \tag{2}$$

where: n – the number of attributes; $w_1, \ldots, w_n$ — the weights assigned to the attributes; $f_1(x, y), \ldots, f_n(x, y)$ — the similarity functions between x and y with respect to attribute $1, \ldots, n$, respectively; $w^{\max}$ – the maximum weight that can be assigned to the attribute; $f_1^{\max}(x, y), \ldots, f_n^{\max}(x, y)$ – the maximum values of the functions $f_1(x, y), \ldots, f_n(x, y)$, respectively.

All attributes can be divided into several classes such as categorical, which are ordered or disordered, and continuous ones. The former ones can be either single-valued like *Gender*, *Education*, *Location* or multiple-valued, e.g. *Interest* or *Language*. The values of these attributes can be one or several from the set of potential values. For categorical attributes their similarity function is usually a binary value that denotes whether any of attribute values is the same for both users, e.g. it is enough that users have at least one language in common. However, for some other categorical attributes like *Interest* we can assume that the more values two users share, the better. For continuous attributes e.g. *Age*, the simple, normalized inverse of difference may be utilized: $f_{age}(10, 10) = 1$, and $f_{age}(1, 100) = 0$, where the max. $age = 100$ and min. $age = 1$, see also the concept in formula (5).

To each of the attribute the appropriate weight, which expresses the usefulness of this attribute for recommendation, is assigned. These weights are adjusted to the specific network using the set of special, non-overlapping rules. A rule can concern the whole attributes and/or only their specific values. Each weight w_i for attribute i (or only for its specific values) is derived from two elements: the static system weight and the rule depended weight:

$$w_i = w_i^{rule} \cdot w_i^{sys} \tag{3}$$

where: w_i – the weight assigned to attribute i that results from the rule set; w_i^{sys} – the system static weight. If no rule related to attribute i exists, then $w_i^{rule}(x, y) = 1$. In the particular case, the whole attribute can be excluded by assigning $w_i^{rule} = 0$. Let us consider the following rule: if the person searches someone for business purpose, then the gender is not an important attribute and $w_{gender}^{rule} = 0$. However, if the person looks for serious relationship, the gender probably becomes crucial and then $w_{gender}^{rule} = 1$.

Additionally, both personal and system weights can be adjusted basing on the user's feedback, i.e. the contribution of the specific feature in the recommendation that user has selected increases the weights for this feature while all the others are slightly reduced [10].

3.6 Complementary of Relationship Initiation

The complementary of relationship initiation matches the willingness of initiation of relationships for user x with the will of user y to respond the invitation.

It enables to compare the frequency of sending the invitation with the probability of response to invitations. All network members are classified into three classes (levels) of initiators: low–medium–high, which are equivalent to $0 - 0.5 - 1$, using the appropriate thresholds applied to the average number of set invitations per month within last year. 0 means that a user sent invitations to others rather rarely or not at all. Value of 1 denotes a very open–minded person who wants to extent their contacts very quickly. In the same way the members are divided into three similar classes of responders: $0 - 0.5 - 1$, based on the percentage of positive replays to requests from others.

Thus, every user has two measures assigned: initialization and response level. The complementary of relationship initiation $c(x, y)$ is the function that combines both these measures and it is calculated as the absolute difference between both of them, e.g. if user x belongs to class 1 as initiator and user y to class 0 as responder then $c(x, y) = |1 - 0| = 1$. It means that we would try to initialize some relationships with resistant members (y) with the help of very open–minded people (x). Similarly, if user x is the bad initiator (0) and user y is the very good responder (1) then $c(x, y) = |0 - 1| = 1$, and we prevent these rare invitation of x from being rejected by recommended y.

During estimating the response level, time is an important factor. Hence, we introduced the "grey period". Invitations sent within this period are not included in calculation of percentage of responses. Moreover, the frequency of invitation takes into account the month: last month has more influence on the final class than the old ones.

3.7 Activity of the User

To define the activity of a user, e.g. their updates of photo/movies/blogs and the assessment whether their behaviours are frequent while some others are not ought to be done. Moreover, we need to respect the general activity context in the network, i.e. all other members who are the best and the worst in their update activities. In addition, the user can be more active in one period and less in another one and the oldest periods should have the least influence on the final measure of an activity. The activity function $a(y)$ of user y is defined as follows:

$$a(y) = \frac{a_1(y) + a_2(y) + \ldots + a_n(y)}{a_1^{\max}(y) + a_2^{\max}(y) + \ldots + a_n^{\max}(y)} \quad (4)$$

where: $a_1(y), \ldots, a_n(y)$ – component activity functions that describe the frequency of the particular activity (e.g. video updates, maintenance of the photo album or frequency of login into the system) in the context of the rest members of the community; $a_1^{max}(y), \ldots, a_n^{max}(y)$ – the maximum values of the functions $a_1(y), \ldots, a_n(y)$; n – the number of attributes.

Each of the functions $a_1(y), \ldots, a_n(y)$ is calculated in the same way. Let us consider for example the frequency of updating the photo album. We find the people in the community who update their photo albums most frequently and most rarely. Additionally, we introduce a time factor. Thus, the component activity function $a_1(y)$ related to photo updates, for user y is:

$$a_1(y) = \frac{1}{1} \cdot \frac{y_1 - l_1}{m_1 - l_1} + \frac{1}{2} \cdot \frac{y_2 - l_2}{m_2 - l_2} + \ldots + \frac{1}{n} \cdot \frac{y_n - l_n}{m_n - l_n} \quad (5)$$

where: n – the fixed number of months that we consider; $y_1, \ldots, y_n$ – the number of updates of a photo album made by person y in the first,$\ldots$, n–th period (usually month), respectively; $l_1, \ldots, l_n$ – the number of updates of a photo album in the first,$\ldots$, n–th period, , respectively, made by the least active person in this area; $m_1, \ldots, m_n$ – the number of maximum updates of a photo album in the first,$\ldots$, n–th period, respectively. If for any period n, $m_n = l_n$, we assume that $\frac{1}{1} \cdot \frac{y_1 - l_1}{m_1 - l_1} = 1$.

This function counts how many times the member updates a photo album in a given period in comparison to the most and least active users within this area. Moreover, the last period has the highest influence on the final value of $a_1(y)$ due to 1/1 factor, while the last period is the least significant.

Note that the value of $a_1(y)$ can exceed 1 and for that reason we need to use maximum values in the denominator of (4).

All elements that are in the activity profile are dynamic because they change over the time. The calculation of activities of users should be repeated for all network members regularly, i.e. after each period (month).

3.8 Strength of the Relationship

The strength of relationship $sr(y)$ is the measure that indicates how firm the relationships between user y and all the others are in the social network:

$$sr(y) = \frac{sr_1(z_1, y) + sr_2(z_2, y) + \ldots + sr_n(z_n, y)}{max_{sum}} \qquad (6)$$

where: n – the number of the relationships maintained by user y; $sr_1(z_1, y), \ldots, sr_n(z_n, y)$ – the partial function that denotes the strength of the relationship between user y and user $z_1, \ldots, z_n$, respectively; max_{sum} - the highest sum of $sr_1(z_1, y) + \ldots + sr_n(z_n, y)$ among all users y_j, used for normalization.

$sr_i(z_i, y)$ is calculated based on the number of sent emails to user z_i by user y, the number of received emails by user y from the user z_i, the number of the mutual readings and comments on their blogs, the number of common chats in specified time, e.g. per month, the frequency of mutual accesses to photo albums, movies, etc. This range of elements can vary between systems and tightly depends on the functionality that the specific system provides.

The i–th partial function $sr_1(z_1, y)$ between person x and y_i is:

$$sr_i(z_i, y) = w_1 \cdot h_1(z_i, y) + w_2 \cdot h_2(z_i, y) + \ldots + w_n \cdot h_n(z_i, y) \qquad (7)$$

where: n – the number of criteria: readings, emails, chats, etc.; $w_1, \ldots, w_n$ – the weight assigned to criterion 1,$\ldots, n$, respectively; $h_1(z_i, y), \ldots, h_n(z_i, y)$ – the functions that describe the specific criteria, e.g. a total number of mutual emails, a number of chats, readings or comments of blogs, etc.

Due to many changes in source data - changing $h_n(z_i, y)$, similarly to activity $a(y)$, also values of functions $sr(y)$ have to be recalculated periodically.

4 Conclusions and Future Work

The proposed framework facilitates the creation of the recommendation within a new application domain: in social networks, where one person is suggested to another one. Besides the typical demographic matching of network members, the unique social elements were introduced. They make use of behaviours and activities of users as well as their common interaction and relationship quality. The social filtering, which is one of the parts of the recommendation process, includes some social elements of the network like the limit of the stable relationships that one person can maintain or connectivity social filtering mechanism that enables to create either close or distributed human communities. The interpersonal similarity encompasses demographic, activity and the strength of relationship components and ensures the appropriate balance between people who are able to initialize new relationships and those who willingly respond to invitations.

The future work will focus on the extension of recommendation by means of the new mechanism that supports the renewal of the declined relationships, which previously were strong but now the information flow between its members is infinitesimal. Besides, the feedback from the used recommendation will be also considered [10]. as well as the usage of the concept of social capital [11]. The monitoring of the user feedback would also help to evaluate the effectiveness of the system.

References

1. Adamic L, Adar E (2003) Friends and Neighbors on the Web. Social Networks 25(3) 211-230
2. Adomavicius G, Tuzhilin A (2005) Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, IEEE Transactions on Knowledge and Data Engineering 17(6) 734-749
3. Golbeck J (2005) Computing and Applying Trust in Web-Based Social Networks. Ph.D. Thesis, University of Maryland http://www.cafepress.com/trustnet.20473616
4. Golbeck J, Hendler J (2004) Accuracy of Metrics for Inferring Trust and Reputation in Semantic Web-Based Social Networks. EKAW 2004, LNCS 3257 116-131
5. Ha S (2002) Helping Online Customers Decide through Web Personalization, IEEE Intelligent Systems 17(6) 34-43
6. Hanneman R, Riddle M (2005) Introduction to social network methods. Online textbook, http://faculty.ucr.edu/ hanneman/nettext/
7. Hill R, Dunbar R (2003) Social Network Size In Humans, Human Nature 14(1) 53-72
8. Kazienko P, Kiewra M (2004) Personalized Recommendation of Web Pages. Chapter 10, Nguyen T (ed.) Intelligent Technologies for Inconsistent Knowledge Processing. Advanced Knowledge International, Adelaide, Australia 163-183
9. Kazienko P, Kiewra M (2003) ROSA - Multi-agent System for Web Services Personalization. AWIC'03, LNAI 2663, Springer Verlag 297-306
10. Kazienko P, Kołodziejski P (2005) WindOwls Adaptive System for the Integration of Recommendation Methods in E-commerce, AWIC'05, LNAI 3528, Springer Verlag, 218-224
11. Kazienko P, Musiał K (2006) Social Capital in Online Social Networks
12. McDonald D (2003) Ubiquitous recommendation systems, IEEE Computer 36(10) 111-112
13. Mooney R, Roy L (2000) Content-based book recommending using learning for text categorization, ACM Conf. on Digital Libraries 195-204
14. Palau J, Montaner M, Lpez B, de la Rosa J (2004) Collaboration Analysis in Recommender Systems Using Social Networks, CIA 2004, LNCS 3191 137-151
15. Perguini S, Goncalves M, Fox E (2004) Recommender systems research: A Connection-Centric Survey, J. of Intelligent Information Systems 23(2) 107-143
16. Tan P, Steinbach M, Kumar V (2006) Introduction to Data Mining, Chapter 8 and 9. Addison-Wesley, 487-649
17. Terveen L, Hill W, Amento B, McDonald D, Creter J (1997) PHOAKS: A system for sharing recommendations, Communications of the ACM 40(3) 59-62