Exploiting Endorsement Information and Social Influence for Item Recommendation

Cheng-Te Li[1], Shou-De Lin[1], Man-Kwan Shan[2]

[1] Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan

[2] Department of Computer Science, National Chengchi University, Taipei, Taiwan

{d98944005, sdlin}@csie.ntu.edu.tw, mkshan@cs.nccu.edu.tw

ABSTRACT

Social networking services possess two features: (1) capturing the social relationships among people, represented by the social network, and (2) allowing users to express their preferences on different kinds of items (e.g. photo, celebrity, pages) through endorsing buttons, represented by a kind of endorsement bipartite graph. In this work, using such information, we propose a novel recommendation method, which leverages the viral marketing in the social network and the wisdom of crowds from endorsement network. Our recommendation consists of two parts. First, given some query terms describing user's preference, we find a set of targeted influencers who have the maximum activation probability on those nodes related to the query terms in the social network. Second, based on the derived targeted influencers as key experts, we recommend items via the endorsement network. We conduct the experiments on DBLP co-authorship social network with author-reference data as the endorsement network. The results show our method can achieve effective recommendations.

Categories and Subject Descriptors: H.3.3 [**Information Systems**]: Information Search and Retrieval.

General Terms: Algorithms, Performance, Design.

Keywords: Social Network, Endorsement Network, Social Influence, Recommendation.

1. INTRODUCTION

In current social networking services (SNS), like Facebook and Twitter, an important characteristic is allowing users to express their favorites on different kinds of items by *endorsing* them. For example, in Facebook, users can "Like" pages, photos, celebrities, videos, and etc. In Twitter, users utilize "Favorite" or "Retweet" to show their preferences. Such endorsement information between users and items can be regarded as a kind of novel source for recommendation. A recent study on such endorsing data is to recommend items which not only relate to the user-given query of tags but also share a significant number of common endorsers [2]. However, in their proposal, they assume individuals in SNS are independent with each other. The social relationships (e.g. friendships) among users, which capture the knowledge about how people interact among different items, are ignored. Some studies use user-item interactions for recommendation. I. Konstas et al. [1] and I. Guy et al. [4] propose diverse modeling approaches to represent the interactions among users, items, and tags as a unified graph and perform different random walk-based methods to recommend items for specific users. On the other hand, in SNS, social network is a crucial medium for the spreading of influence and information. People who adopt a new idea or buy a new item could be affected by the trusts and decisions of their friends. Kempt et al. [3] investigate the behaviors of information diffusion and the *word-of-mouth* effect. They try to find influential individuals who can trigger maximum cascades of influence in a network, for *viral marketing*.

In this work, we propose a novel recommender system, which use the endorsement information and the influence propagation in a social network. Our central idea is to find topical influencers as the recommenders to recommend items via the endorsement information. Specifically, given query terms describing user's preference, we aim to find some individuals (termed *targeted influencers*) who have maximum influence potentials on some persons related to the query. We regard these targeted influencers as authoritative roles w.r.t. user's interests. And then using the endorsing information, we recommend the items which are commonly and effectively endorsed by the targeted influencers. Our method has two characteristics: (1) recommending items by the targeted influencers can be more explainable and convincing. (2) The social endorsement is a kind of the wisdom of crowds, and thus recommendation combining targeted influencers and such endorsing information can be regarded as crowdsourcing from key players.

Preliminary. First, the input of our recommendation is a set of query terms which represents user's preferences. The set of nodes whose interests satisfy the query terms is denoted by Q. Second, we represent the relationships among users in a SNS as a social network $G = (V, E, P)$, where V is the user set, E is the set of relationships between users, and each edge is associated with an influence probability $p(u, v) \in [0,1]$ as weight. Each node i is associated with a set of labels L_i describing his/her preferences. Third, we formulate the endorsement information as a bipartite graph $H = (V, U, E_H)$, where V is the user set, U is the item set, and E_H is the set of endorsing edges. Note that the set of user nodes in H could contain the user set in G. Fourth, we propose the *Collective Influence* (*CI*) model for query-based influence propagation in a social network. The *CI* model starts from an initial set of active nodes A_0, and proceeds by the following rules in discrete time rounds. In round r, each active node u has a single chance to activate each of its active or inactive neighbors v with the pre-determined influence probability $p(u, v)$. If u successfully activates v and v is inactive, v will become active at round $(r + 1)$. If u successfully activates v and v is active, v will keep active at round $(r + 1)$. This procedure will be terminated until no activations are possible. Fifth, given a source node $s \in V$ as a candidate targeted influencer, and a target node $t \in Q$ $(s \neq t)$, we define the *propagation path* $P_{s\to t} = \langle s = v_1, v_2, \dots, v_m = t\rangle$, where $v_1, v_2, \dots, v_m$ are active, as the highest probabilistic path from source s to target t. We also define the *activation probability* $ap(s, t) = \prod_{i=1}^{m-1} p(v_i, v_{i+1})$ as the chance that s successfully affects t. If s and t are disconnected, $ap(s, t) = 0$. In addition, for the set of nodes with query terms Q and the set of candidate targeted influencers S, we define the *average activation probability* as

$$R(Q, S) = \left(\sum_{t\in Q} max_{s\in S}\{ap(s, t)\}\right)/|Q|.$$

Problem Statement. Given (a) a social network G, (b) a set of nodes with query terms: Q, (c) the endorsement bipartite network $H = (V, U, E_H)$, and (d) the number of targeted influencers k_s and the number of recommended items k_r, using the proposed Collective Influence model, we aim to (1) find a set of k_s targeted influencers S in G such that the objective function of average activation probability $R(Q, S)$ is maximized. (2) Recommend a ranked list of k_r items in H based on the set of targeted influencers S.

2. THE PROPOSED METHOD

Finding Targeted Influencers. This part contains two stages: (1) grouping nodes based on query terms, and (2) a greedy method to maximize the objective function for finding targeted influencers. We

SIGIR 2011, July 24 – 28, 2011, Beijing, China.

first assume nodes with query terms form several components in the social network. We identify such small groups, where each group contains only nodes with query terms, and obtain the corresponding induced subgraphs $\{G'_1, G'_2, \dots, G'_m\}$, where m is the number of groups. This grouping will reduce the graph space and allow the search to be more efficiently. And then our goal is to distribute the set k_s nodes S as targeted influencers over these m subgraphs such that the determination of k_s nodes can maximize the average activation probability $R(Q,S)$. We first explain our objective $R(Q,S)$ satisfies the submodularity property, and then use such property to devise a greedy algorithm to maximize $R(Q,S)$.

First, R is *nondecreasing* $R(Q, X \cup \{v\}) \geq R(Q,X)$: adding a node to a set will not decrease R. Second, R is *submodular* (i.e., satisfies the diminishing returns) $R(Q, X \cup \{v\}) - R(Q,X) \geq R(Q, Y \cup \{v\}) - R(Q,Y)$: the *marginal gain* of adding a node v to a set X is higher than or equal to the marginal gain of adding v to X's superset Y ($X \subseteq Y$) for any nodes $v \in V \setminus Y$. That is, let $R(Q,\emptyset) = 0$ in the initial state, in t^{th} round, a targeted influencer v is determined and added into the set S_t, the marginal gain of the average activation probability $R(Q, S_t \cup \{v\}) - R(Q,S_t)$ is less than or equal to that of the (t-1)th round $R(Q, S_{t-1} \cup \{u\}) - R(Q, S_{t-1})$, where $S_t = S_{t-1} \cup \{u\}$.

To find the set of targeted influencers S, we use the submodularity to devise a greedy algorithm. The algorithm starts from initializing the set $S_0 = \emptyset$, and then iteratively adds a node s into S_t in the following t^{th} round (t>0) such that s maximizes the marginal gain

$$S_t = argmax_{s \in V \setminus S_{t-1}} \{R(Q, S_{t-1} \cup \{s\}) - R(Q, S_{t-1})\}.$$

The greedy algorithm will continue to add nodes into S_t until t=k_s, where we reach the budget of selecting targeted influencers.

Item Recommendation. Based on the derived targeted influencers, we use the *Random Walk with Restart* algorithm (*RWR*) [5] in the endorsement bipartite graph for the item recommendation. RWR is capable of predicting the preference or affinity of items for some indicated users. We set the targeted influencers as the designated restarting nodes and apply RWR to compute the proximity scores for all items in the bipartite graph. The restarting probability is set to 0.25 in this work. A ranked list with k_r items of highest scores will be returned as the recommended items.

3. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed method, we compile the DBLP bibliography data to a connected co-authorship social network containing 5,095 nodes and 11,800 edges in some premier conferences, including SIGIR, WWW, CIKM, WSDM, JCDL, KDD, ICDM, SDM, PAKDD, PKDD, VLDB, SIGMOD, and ICDE. The edges are constructed if two authors co-work more than 3 papers. The set of labels L_i associated with each author i are the textual terms occurring in at least 3 paper titles that he/she ever participated. And the probabilistic weights on edges are determined by $p(u,v) = |L_u \cap L_v| / \min\{|L_u|, |L_v|\}$. For the endorsement network, it consists of authors and papers as nodes. If an author v ever referenced to a paper u, we connect v to u. in the endorsing graph. Totally there are 18,357 nodes and 147,392 papers. Note that we only include the papers occur in mentioned conferences. We also compile two sets of query terms from two areas: information retrieval (IR-Query) and social network (SN-Query). For each query set, we manually find 20 authors as testing influencers, where each author must publish at least 5 related papers and serve as PC in at least 5 proceedings of our used conferences. The details about the queries are shown in Table 1.

The first experiment is computing the average activation probability $R(Q,S)$ for the proposed method of finding targeted influencers S (k_s is varied from 1 to 20.). We compare our method to the random selection as baseline and three heuristic methods: degree, closeness, and betweenness centralities, which aim to find different kinds of important nodes in a social network. The experimental results are shown in Figure 1. We can find our method outperforms the other five heuristics for both IR-Query and SN-Query.

Table 1. Query terms and testing authors for IR and SN queries.

	IR-Query	SN-Query
Query terms	retrieval, text, ranking	social, network(s), mining
Testing Authors	Wei-Ying Ma, Tat-Seng Chua, W. Bruce Croft, Chris H. Q. Ding, Susan T. Dumais, James Allan, Ji-Rong Wen, Charles L. A. Clarke, Chris Buckley, Jian-Yun Nie, and et al. (totally 20)	Jure Leskovec, Ee-Peng Lim, Evimaria Terzi, Jimeng Sun, Jon M. Kleinberg, Jie Tang, Hari Sundaram, Lise Getoor, Yutaka Matsuo, Tina Eliassi-Rad, and et al. (totally 20)

The second experiment is to test whether the top-k recommended papers are referenced by compiled productive authors. We define an accuracy measure *author hit rate* $HitRate_k = AuthorHit_k / N_{test}$, where $AuthorHit_k$ is the number of authors referencing to any of top-k returned papers and N_{test} (=20) is the total testing authors. We compare our method with the approach of pure Random Walk with Restart [5] considering all nodes with query terms as the restarting ones. We set k_s=10 and k_r=10. The experimental results are shown in Figure 2. For IR-query, we can find ours outperforms the RandomWalk approach as the k_r increases. This implies ours provides influential recommendations in IR domain. For SN-query, we find the effectiveness of ours is more distinguished over all k returned. After studying the recommended papers (that is not shown here for space limit), we find that ours precisely suggest social-networked papers while pure RandomWalk is prone to return papers that mix the areas of social network and sensor network mining.

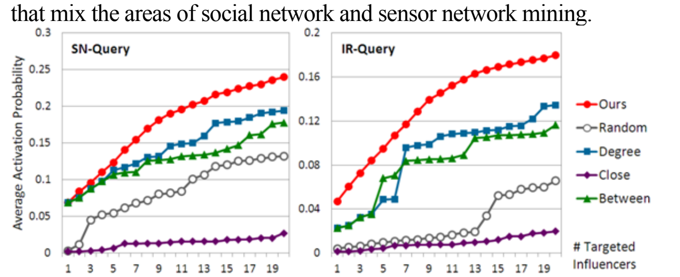

Figure 1. Average activation probability for SN-Query and IR-Query.

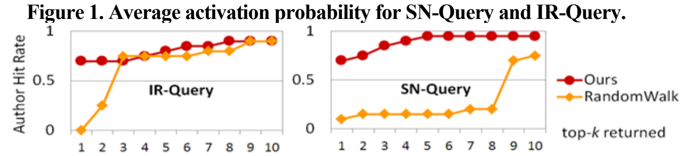

Figure 2. Author hit rate for IR-query and SN-Query.

4. CONCLUSION

We develop a novel recommender, which exploits the information of social and endorsement networks. The central idea is to leverage the concepts of viral marketing and the wisdom of crowds for item recommendation. We devise a greedy method which maximizes a submodular objective function to find the targeted influencers who can effectively activate nodes with query terms in the social network. And then we use the random walk mechanism to recommend items based on the targeted influencers in the endorsement network. Evaluations on academic data show promising results.

5. REFERENCES

[1] I. Guy, N. Zwerdling, L. Ronen, D. Carmel, and E. Uziel. Social Media Recommendation based on People and Tags. In *ACM SIGIR* 2010.

[2] T. Lappas and D. Gunopulos. Interactive Recommendations in Social Endorsement Networks. In *ACM Recommender System* 2010.

[3] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the Spread of Influence through a Social Network. In *ACM SIGKDD* 2003.

[4] I. Konstas, V. Stathopoulos, and J. M. Jose. On Social Networks and Collaborative Recommendation. In *ACM SIGIR* 2009.

[5] H. Tong, C. Faloutsos, and J. Y. Pan. Fast Random Walk with Restart and Its Applications. In *IEEE ICDM* 2006.