

Personalized Recommendation Driven by Information Flow

Xiaodan Song

Department of Electrical Engineering
University of Washington, Box
352500, Seattle, WA 98195 USA
song@ee.washington.edu

Belle L. Tseng

NEC Laboratories America
10080 N. Wolfe Rd, SW3-350
Cupertino, CA 95014 USA
belle@sv.nec-labs.com

Ching-Yung Lin, Ming-Ting Sun

Department of Electrical Engineering
University of Washington, Box
352500, Seattle, WA 98195 USA
{cylin,
sun}@ee.washington.edu

ABSTRACT

We propose that the information access behavior of a group of people can be modeled as an information flow issue, in which people intentionally or unintentionally influence and inspire each other, thus creating an interest in retrieving or getting a specific kind of information or product. Information flow models how information is propagated in a social network. It can be a real social network where interactions between people reside; it can be, moreover, a virtual social network in that people only influence each other unintentionally, for instance, through collaborative filtering. We leverage users' access patterns to model information flow and generate effective personalized recommendations. First, an early adoption based information flow (EABIF) network describes the influential relationships between people. Second, based on the fact that adoption is typically category specific, we propose a topic-sensitive EABIF (TEABIF) network, in which access patterns are clustered with respect to the categories. Once an item has been accessed by early adopters, personalized recommendations are achieved by estimating whom the information will be propagated to with high probabilities. In our experiments with an online document recommendation system, the results demonstrate that the EABIF and the TEABIF can respectively achieve an improved (precision, recall) of (91.0%, 87.1%) and (108.5%, 112.8%) compared to traditional collaborative filtering, given an early adopter exists.

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

General Terms: Algorithms, experimentation

Keywords: Information flow, personalized recommendation, collaborative filtering

1. INTRODUCTION

The quantity of new information a person receives everyday usually goes beyond his limited processing capabilities. Recommender systems can help users to deal with information

overload. In particular, recommender systems based on Collaborative Filtering (CF), which has been proven to be very successful in a variety of domains, try to automate the “word-of-mouth” process [1]. For example, when we think about whether going to see a new movie, we often ask friends with similar movie tastes and then we act based on their recommendations. Many Internet-based CF systems try to automate this process to a world scale. It infers an individual's interests/preferences from those of other people with similar tastes [1]. The system stores the preferences and opinions of thousands of users. When a user needs a recommendation, the system finds users with a similar taste, and uses their opinions to generate a recommendation.

Various CF algorithms have been designed to identify users of similar interests [31]. Nearest Neighborhood [2] is the most typical method that computes similarity scores between all pairs of users, and the predictions for a given user are generated by weighting other users' ratings proportionally to their similarity to the user. A variety of similarity metrics have been tested, including cosine distance [3], correlation [4], or mean-squared difference [5].

Typical recommender environments are dynamic in many aspects. Related works on exploring dynamic features include [7, 27-29]. [27] studies how to learn users' long-term and short-term interest categories in a dynamic environment. [28] proposes a multidimensional approach to present the temporal and other contextual information besides the user and item-rating information. [29] adopts forgetting weights to model evolving user interests in clickstreams by taking into account the age of each user activity/session. In [7], we make recommendations by exploiting dynamic nature from both documents' and users' aspects. In this paper, we propose to handle the dynamic recommendation issue as an information flow problem.

Identifying how information is propagated in a network is important in many applications. A social network plays a fundamental role as a medium for the spread of information, ideas, and influence [10-17]. [17] presents a non-parametric approach by identifying early buyers as those consumers who tend to purchase products earlier than the other consumers, by making use of a graphical representation of purchase data. Their algorithm is applied to Amazon.com, providing insights into consumer behaviors. [11-13] attempt to model influence among consumers and understand how this influence propagates in the networks. The problem they addressed is: suppose that we have data on a potential network of consumers with estimation of the extent to which individuals influence one another, if we are given a new product and a marketing budget, how can we maximize the adoption of the new product through customers? In addition, trust propagation is also one of the active research issues that utilize

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGIR '06, August 6-11, 2006, Seattle, Washington, USA
Copyright 2006 ACM 1-59593-369-7/06/0008...\$5.00

information flow. Modeling the propagation of the trust scores is a fundamental building block in many of today's most successful e-commerce and recommendation systems. [8, 18] try to predict trust and distrust between any two users based on a trust propagation model. Information propagation through blog space was also recently studied [19]. In this paper, we use the terms “information propagation” and “information flow” interchangeably to represent the same concept.

In Rogers’ “Diffusion of Innovation” [6], the adoption curve classifies adopters of innovations into five categories based on the idea that certain individuals are inevitably more open to adoption than others in a social network. The innovation can be an idea, a practice, or an object. The five adopter categories, (1) innovators, (2) early adopters, (3) early majority, (4) late majority, and (5) laggards, follow a standard deviation curve. Innovators adopt the innovation at the beginning (2.5%); early adopters make up for 13.5% a short time later; the early majority is 34%; the late majority is 34%; and finally the laggards make up for 16%. It is also referred to as the Multi-Step Flow Theory or Diffusion of Innovations Theory.

In this paper, instead of posing the typical CF question: *how will a user X like an item Y?* We view the problem from another perspective by asking the question: *given that a user X acts on an item Y who will then be more likely to act on the item Y?* The fact that two users access or purchase the same item sequentially is modeled as an information flow process: *the information is flowing from early adopters to late adopters.* We do not explicitly explore how the early adopter influences the late adopter. There may be existing social network connections between these two users. Or, these two persons may have habits of watching the same TV channel, one in the morning and one in the evening, showing the same advertisement. We consider, as long as there are consistent time-based user access patterns, then the appearance of the early adopter accesses an item indicates his late adopter is very likely to access soon. On the other hand, the early adopter may not be interested in the items that the late adopter accesses. This time-factor was not previously addressed by collaborative filtering recommendation methods.

To model information flow, first of all, we construct a weighted directed network, which we call Early Adoption Based Information Flow (EABIF) network, based on pair-wise comparisons to describe how likely each user make adoptions earlier than the other user. Then we propose to model information flow behavior of a node by using ergodic Markov chains, in which all states are reachable from all other states (i.e., all states communicate), so that the convergence is guaranteed. Furthermore, considering that adoption is typically category-specific, we proposed a topic-sensitive early adoption based information flow (TEABIF) network. Personalized recommendation is exploited as an application, which addresses the question of “if one user or multiple users access an item, who else will likely follow these early adopters and access this item next?” In a social network, this process is similar to when one node is triggered, then subsequently which other nodes in the network will also be triggered. Figure 1 illustrates the overall structure of our proposed personalized recommendation driven by information flow scheme.

The major contributions of this paper are as follows:

- 1) Leverage time- and relation-based users adoption patterns for personalized recommendation
- 2) Propose topic-sensitive model based on the observation that adoption is typically category specific

- 3) Propose three information propagation models based on ergodic Markov chains to guarantee convergence

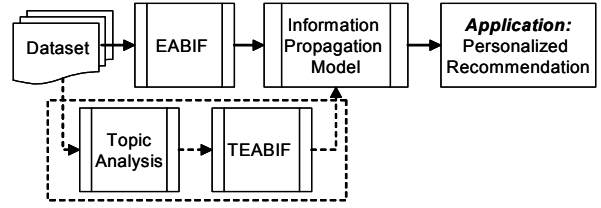


Figure 1: An overview of the personalized recommendation driven by information flow scheme through either EABIF or TEABIF

The rest of the paper is organized as follows. In Section 2, we present our information flow networks including EABIF and TEABIF. In Section 3, we propose the information propagation models. In Section 4, we show the experimental results of the proposed personalized recommendation methods, comparing to those using collaborative filtering. Finally, conclusions and future work are addressed in Section 5.

2. INFORMATION FLOW NETWORK

Market research has shown that users exhibit a variety of different adoption behaviors; specifically, some tend to adopt items earlier than the others. In this section, we describe our information flow networks including EABIF and TEABIF to model users’ adoption patterns.

2.1 Intuition

We first give an intuitive description of how leveraging users’ adoption patterns will help for recommendations. In the example of Figure 2, there are three users in the universe: u , v , and r . From the historical data, we learn that User u and User v have accessed 40 items in common; also, User u and User r also have accessed 40 items in common. Thus the cosine similarities between u and v , and u and r are the same, based on their previous accessing behaviors. When u adopts an item D , then a Collaborative Filtering method based on Cosine Similarity will equally recommend this item D to r and v . Similarly, when r or v accessed an item, this item will also be recommended to u with an equal weight. However, if we compare their adoption timestamps, among 40 items, u accessed 10 of them earlier than v , and 30 times of them later than v . On the other side, u accessed 30 of them earlier than r , and 10 times later than v . These two types of adoption patterns are quite different, which basically tells us that the information “flows” from u to r is more likely than from u to v .

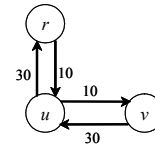


Figure 2: An illustration of how adoption patterns affect the recommendations

Assume we have a social network for the early-late adoption patterns. This network describes how likely the information will flow from one user to another. Imagine an example in Figure 3; User u adopts information in the initial state. We assume this information can be transferred to the other nodes in the network through multiple possible paths. Given the propagation probability

of each path, it is possible to get the final information receiving probabilities of the nodes (as illustrated as the bars in Figure 3.)

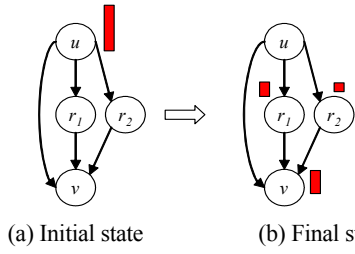


Figure 3: An illustration of information flow

Note that, in the recommendation process, we care more on the rank of the users who are more likely to adopt an item after u , rather than the exact probabilities that these users will adopt. In this paper, we make an assumption that the information flow is similar to the water flow, in which information is not duplicated through propagation. Imposing this assumption seems to be artificial. However, it provides an ergodic Markov chain property that guarantees convergence and simplifies the calculation of final-state probabilities, while keeping the rank of users. Thus, the system can be implemented and executed efficiently.

2.2 Early Adoption based Information Flow Network

In this subsection, first we propose an early adoption matrix to represent the adoption patterns. Then we generate an early adoption based information flow network as a Markov chain for personalized recommendation.

2.2.1 Early Adoption Matrix

Assume we have N users and M items. Given the detailed log information of each user, including the timestamp of each adoption, for each pair of users, we compare their timestamps of adopting the same items, and record how many items one user adopts earlier than the other into an early adoption matrix (Figure 4). Then, we represent this matrix using a weighted directed graph $G(n, l)$, where each node n corresponds to a user and l represents the non-negative weight on each edge. We have an edge with weight $l_{u,v}$ if among the items that u and v accessed in common, $l_{u,v}$ of these items were accessed by u earlier than v .

	User 1	User 2	...	User N
User 1	0	10	...	8
User 2	27	0	...	18
...
User N	0	30	...	0

Figure 4: Adoption matrix: the value in each unit represents how many items user (in rows) accessed earlier than user (in columns)

For ease of description, we represent a missing edge from u to v as an edge with zero weight. An edge (u, v) with $l_{u,v} = 0$ means that among the common items that u and v adopt, u always adopts them later than v . Note that for any u and v , the sum of $l_{u,v}$ and $l_{v,u}$ equals the total number of items adopted by u and v . A large value of $l_{u,v}$ relative to $l_{v,u}$ indicates that in most cases, User u adopts the items earlier than User v .

2.2.2 Transition Probability

Markov Chain-based models have been proven successful in social networks analysis ([20]) and web search ([21-23], [25]). The PageRank algorithm computes the importance scores of web pages through a stochastic ergodic Markov transition matrix which is constructed from all hyperlinks between web pages. In this paper, we model the adoption graph as an ergodic Markov chain with primitive transition probability matrix as PageRank does [23, 25] to guarantee the convergence of the power of the matrix.

To study the early/late adoption relationship between users, regarding an item, we define each user can be either *active* (adopted the item) or *inactive* (have not adopted the item). For *active* mode, we define a consequential mode according to the time they make the adoptions. If User u initiates the adoption behavior, his/her mode is *active*, User v follows u to access the same item, then his/her mode is *active after u* (af_v).

According to the Bayesian rule:

$$P(v = af_u | u = a) = \frac{P(v = af_u, u = a)}{\sum_r P(r = af_u, u = a)} = \frac{l_{u,v}}{\sum_r l_{r,v}} \quad (1)$$

where $P(v = af_u | u = a)$ represents the probability that given u obtains the information, how likely v will adopt the information after u among all the users. Then the normalized probability serves as the transition probability in a Markov chain \mathbf{F} , with $F_{u,v} = P(v = af_u | u = a)$ as the transition probability of information flowing from u to v .

2.2.3 Adjustment

When a Markov chain is ergodic, the stationary distribution of its Markov matrix is guaranteed to exist. To achieve this ergodic Markov chain, the first problem of solely using the adoption patterns to generate the transition probability matrix is that some rows of the matrix contain all zeroes. This occurs whenever one user did not adopt any items earlier than the others. Thus, \mathbf{F} is not stochastic. A matrix \mathbf{F} is stochastic if all $F_{u,v}$ are non-negative and $\sum_v F_{u,v} = 1$. The nodes which do not satisfy $\sum_v F_{u,v} = 1$ are called dangling nodes. Similar to what PageRank does [23,25], we apply the remedy to replace all zero rows, 0^T with $\frac{1}{N}e^T$, where e^T is the row vector of all ones. The revised transition probability matrix called $\bar{\mathbf{F}}$ is

$$\bar{F}_{u,v} = \begin{cases} P(v = af_u | u = a) & \text{if } \sum_v \bar{F}_{u,v} \neq 0 \\ \frac{1}{N} & \text{else} \end{cases} \quad (2)$$

Another adjustment to make a Markov chain ergodic is to consider the transition probability matrix as a primitive matrix:

$$\bar{\bar{\mathbf{F}}} = \alpha \bar{\mathbf{F}} + (1 - \alpha) ee^T / N \quad (3)$$

This convex combination of the stochastic matrix and a stochastic perturbation matrix ee^T insures that $\bar{\bar{\mathbf{F}}}$ is both stochastic and irreducible. Every node is now directly connected to every other node, making the chain irreducible by definition. Although the transition probability may be very small in some cases, it is always nonzero. The irreducibility adjustment also ensures that $\bar{\bar{\mathbf{F}}}$ is

primitive, which implies that when n is large $\bar{F}^{(n)}$ will converge to the stationary distribution. By this definition, \bar{F} is constructed as a primitive stochastic matrix to model the probabilities of information moving from one node to another. In our experiments, we found that the results are insensitive to the value of α and set $\alpha = 0.5$.

We call this generated irreducible Markov chain with \bar{F} as the transition probability matrix is called Early Adoption Based Information Flow (EABIF) Network.

2.3 Topic-Sensitive Early Adoption based Information Flow Network

People have different behavior patterns regarding to different categories or topics; an early adopter of fashion may not be an early adopter of technology. We have utilized people's different social networks regarding different topics through emails in [9] to model and predict human activities. In this paper we will follow the same intuition to generate the topic-sensitive information flow networks for recommendations. Here we consider items as documents for topic analysis. The basic idea is that people may tend to get information of certain topics earlier than others, while for some other topics, they may be not so eager to obtain that. Thus to get the more consistent early/late behavior patterns, we cluster documents into latent topics and then build the networks based on how likely one user accesses the documents regarding to a certain topic earlier or later than the other users.

We adopted Latent Dirichlet Allocation [26] to generate the topics. Latent Dirichlet Allocation is a well-defined probabilistic generative model, which can be generalized easily to new documents. And, the number of parameters does not grow with the size of the training corpus. Thus, it is not prone to overfitting. Given T topics (T is determined by using cross-validation), the probability of the i th word in a given document is formularized as:

$$P(w_i) = \sum_{j=1}^T P(w_i | z_i = j) P(z_i = j) \quad (4)$$

where z_i is a latent variable indicating the topic from which the i th word was drawn, and $P(w_i | z_i = j)$ is the probability of the word w_i under the j th topic. $P(z_i = j)$ gives the probability of choosing a word from topics j in the current document, which varies across different documents. Gibbs sampling [30] is applied to estimate $\phi_j^{(w)}$, the probability of using word w in topic j , and $\theta_j^{(d)}$, the probability of topic j in document d [16].

After the topics are generated, each document is clustered into a cluster j with a probability $\theta_j^{(d)}$ with $\sum_{j=1}^T \theta_j^{(d)} = 1$. Then for each cluster (topic), a topic adoption matrix as well as a weighted directed graph are generated for all users. If User u accesses a document earlier than User v , and this document is clustered into Topic j with a probability of $\theta_j^{(d)}$, then this early adoption will contribute $\theta_j^{(d)}$ to the unit (u, v) in the topic adoption matrix as well as the weight from u to v in the weighted graph. Then these T weighted networks, with each one corresponding to a topic are used to generate Markov chains with a primitive stochastic matrix as the transition probability matrix \bar{F}_i is constructed for each topic i . These

Markov chains are called Topic-Sensitive Early Adoption Based Information Flow Network (TEABIF).

3. INFORMATION PROPAGATION MODELS

After we construct the information flow networks, the current problem is: how to model the information propagation through the network?

We first give an intuitive description of an information propagation step (Figure 5). In our information flow networks, the pairwise adoption patterns between users are known. By one step calculation, people with highest transition probability from the current user will be ranked as the most likely follower among all the users. However, if we consider how information flows, it is possible that User u obtains the information and spreads it to User v directly, or the information goes through User r_1 to User v , or User r_2 , User r_3 to User v , or other paths. Different from the traditional collaborative filtering technique which is solely based on pairwise similarity, our information flow driven recommendation scheme utilizes how information propagates through multiple steps and paths to rank how likely users adopt the information.

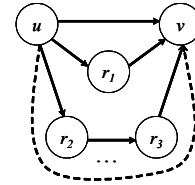


Figure 5: Information flows through multiple paths

In the following we analyze how information flows through multiple steps and multiple paths.

1. The transition probability of the information flowing from u to v among all the users in one step is $P(v = af_u | u = a)$, which is the transition probability from u to v , in our information flow networks.
2. The transition probability of the information flowing from u to v through two steps is $\sum_r P(r = af_u | u = a) P(v = af_r | r = a)$.
3. The transition probability of more steps has similar forms.

Assume p_k being a distribution with $\sum_k p_k = 1$ to be used for

describing how likely each type of paths contributes to the overall information flow. Then the overall information flow (if) probability that the information will propagate from u to v among all users is equal to:

$$P_{if}(v = af_u | u = a) = p_1 \cdot P(v = af_u | u = a) + p_2 \cdot \sum_r P(r = af_u | u = a) P(v = af_r | r = a) + \dots + p_{K-1} \cdot \sum_{r_1 \dots r_{K-1}} P(r_1 = af_u | u = a) \dots P(v = af_{r_{K-1}} | r_{K-1} = a) \quad (5)$$

where K is the maximum value of propagation steps.

In this paper, we propose three ways to calculate this probability.

1. Summation of various propagation steps

By equally considering various propagation steps, i.e., p_k is a uniform distribution, Eq. (5) can be calculated as:

$$\mathbf{F}_{if(m)} = \left(\bar{\bar{\mathbf{F}}} + \bar{\bar{\mathbf{F}}}^{(2)} + \dots + \bar{\bar{\mathbf{F}}}^{(m)} \right) / m \quad (6)$$

where m is the number of propagation steps. When $m = 1$, $\mathbf{F}_{if(m)} = \bar{\bar{\mathbf{F}}}$ is a special case of Eq. (6) which only considers the pairwise relationship.

2. Direct summation for $m = N - 1$, and $N \rightarrow \infty$

Considering p_k being a uniform distribution and the maximum propagation steps in the network with N users being $(N - 1)$, Eq. (5) is equal to

$$\mathbf{F}_{if(N-1)} = \left(\bar{\bar{\mathbf{F}}} + \bar{\bar{\mathbf{F}}}^{(2)} + \dots + \bar{\bar{\mathbf{F}}}^{(N-1)} \right) / (N - 1) \quad (7)$$

which can be calculated as

$$\mathbf{F}_{if(N-1)} = \frac{\bar{\bar{\mathbf{F}}} \cdot (\mathbf{I} - \bar{\bar{\mathbf{F}}}^{(N-1)})}{\mathbf{I} - \bar{\bar{\mathbf{F}}}} \cdot \frac{1}{N - 1} \quad (8)$$

where \mathbf{I} is a $N \times N$ unit matrix.

Assume the eigenvalues of matrix $\bar{\bar{\mathbf{F}}}$ be $\lambda_1, \lambda_2, \dots, \lambda_N$,

$$\bar{\bar{\mathbf{F}}}^{(N)} = [\mathbf{U}\mathbf{A}\mathbf{U}^{-1}]^{(N)} = \mathbf{U}\mathbf{A}^{(N)}\mathbf{U}^{-1} = \mathbf{U} \begin{bmatrix} \lambda_1^{(N)} & & & \\ & \lambda_2^{(N)} & & \\ & & \dots & \\ & & & \lambda_N^{(N)} \end{bmatrix} \mathbf{U}^{-1} \quad (9)$$

As we mentioned in section 2.2.3, $\bar{\bar{\mathbf{F}}}$ is a primitive stochastic matrix, which has only one eigenvalue on the unit circle, all other eigenvalues have modulus strictly less than one [24]: $1 = \lambda_1 > \lambda_2 > \dots > \lambda_N$. When the number of users in the network (N) is large (> 100), $\lambda_i^{(N)} \approx 0$ for these $\lambda_i < 1$. This means that the power method applied to a primitive stochastic matrix $\bar{\bar{\mathbf{F}}}$ is guaranteed to converge to the unique dominant eigenvector—the stationary distribution. As a result, there is no issue with convergence of the ranking vector, and any positive probability vector can be used to start the iterative process. Further, the convergence rate of the method is determined by the magnitude of the subdominant eigenvalue of the transition rate matrix. Thus, we have

$$\bar{\bar{\mathbf{F}}}^{(N)} = \mathbf{U} \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \dots & \\ & & & 0 \end{bmatrix} \mathbf{U}^{-1} \quad (10)$$

After $\bar{\bar{\mathbf{F}}}^{(N)}$ is calculated by Eq. (10) and insert into Eq. (9), $\mathbf{F}_{if(N-1)}$ is obtained. When $N \rightarrow \infty$, $\mathbf{F}_{if(N-1)} \rightarrow \bar{\bar{\mathbf{F}}}^{(N)}$.

3. Exponential weighted summation

Usually short paths are more important for information propagation because their direct relationship to the users. Thus, we consider assigning less importance to longer paths:

$$\mathbf{F}_{if(\text{exp})} = \left((\beta \bar{\bar{\mathbf{F}}}) + \frac{1}{2!} (\beta \bar{\bar{\mathbf{F}}})^2 + \dots + \frac{1}{(N-1)!} (\beta \bar{\bar{\mathbf{F}}})^{(N-1)} + \dots \right) \cdot \exp(-\beta) \quad (11)$$

where $\frac{1}{N!}$ assigns less importance to longer paths, β is a parameter to control the effects of transitivity, and $\exp(-\beta)$ is a normalization factor. The larger β is, the longer paths are considered important. When $\beta \rightarrow \infty$, $\mathbf{F}_{if(\text{exp})} \rightarrow \bar{\bar{\mathbf{F}}}^{(N)}$.

As we know,

$$\exp(\beta \bar{\bar{\mathbf{F}}}) = \mathbf{I} + (\beta \bar{\bar{\mathbf{F}}}) + \frac{1}{2!} (\beta \bar{\bar{\mathbf{F}}})^2 + \dots + \frac{1}{(n-2)!} (\beta \bar{\bar{\mathbf{F}}})^{(n-2)} + \dots \quad (12)$$

Therefore, we have $\mathbf{F}_{if(\text{exp})} = \left(\exp(\beta \bar{\bar{\mathbf{F}}}) - \mathbf{I} \right) \cdot \exp(-\beta)$. By using eigenvalue composition, $\exp(\beta \bar{\bar{\mathbf{F}}})$ can be calculated by

$$\exp(\beta \bar{\bar{\mathbf{F}}}) = \mathbf{U} \begin{bmatrix} \exp(\beta \cdot \lambda_1) & & & \\ & \exp(\beta \cdot \lambda_2) & & \\ & & \dots & \\ & & & \exp(\beta \cdot \lambda_N) \end{bmatrix} \mathbf{U}^{-1} \quad (13)$$

4. EXPERIMENTS

4.1 Experimental setup

We describe our proposed scheme using a dataset collected from NEC's "EigyoRyoku 21" (denoted as ER and stands for Sales-Force in Japanese) system. The ER system is a knowledgebase to support sales staffs with registered documents that include articles, slides, etc. We collected a sixteen-month period of clickstream log files from April 2004 to July 2005 covering more than 30,000 users and 30,000 documents. Nine user actions are identified: {"Login", "Register_Feedback", "Preview", "Abstract", "Document Download", "Search", "Register", "Update", "Delete"}. The clickstream log is partitioned into sessions that start with "login" and then sequences of user actions. The timestamps of users' actions as well as disclosure of the documents are included in the dataset. In this paper, we assume if a user access (including "Preview", "Abstract" or "Download") a document, he or she has adopted it.

The dataset is divided into two sets: the training set and the test set. The data from April 2004 to April 2005 serve as the training data, and the data from May to July 2005 serve as the test data. Among more than 30,000 users, approximately 20,000 of them only have no more than 10 adoption behaviors during the 16 months. That demonstrates the sparseness of the dataset. To exclude casual users who had very few activities, we selected 1170 active users who adopted more than 50 documents in the training period; and more than 10 documents in the test period in our experiments. In total, there are 23894 documents involved in this selected dataset. 201,750 adoption actions were recorded. The average number of adoption actions per user is 172, and the average number of actions per document is 8.

In the experiments, we simulate the situation of adoption behaviors. There are 586 documents disclosed during the test period from May to July 2005. The mean value of the number of users who adopted these 586 documents during May to July 2005 is 18. These documents were first adopted by one or multiple users – early

adopters. Then we predict who else will most likely adopt the documents following these early adopters. Hence, instead of directly making document recommendations to users, we predict the potential adopters of each document. This strategy is suitable for online document/product pushing service or advertisement – recommending the items to potential customers according to the behaviors of early adopters. It can also be transferred to the traditional recommendation scenario by estimating how likely one user will be interested in the documents and recommending those top-ranked documents to him or her.

To evaluate the accuracy of predictions, we measure the *precision* and *recall* of the recommendations. Precision represents the ability of the system to withhold non-relevant users, which is measured as the proportion of recommended users who adopt the documents in the test period. Recall represents the ability of the algorithm to present all relevant users, which is measured as the proportion of the recommended users over all the users who adopted the documents in the test period. The values we report here are the respective average of the precisions and recalls for these 586 documents.

In our experiments, we compare the performance of the algorithms as following:

1. Collaborative Filtering based on Cosine Similarity (baseline)
2. Early adoption based information flow network with information propagation models
3. Topic sensitive information flow network with information propagation models

Specifically, for information propagation models, we will demonstrate the performances of propagation steps (m) from 1 to 5, direct summation, and exponential weighted summation.

4.2 Results

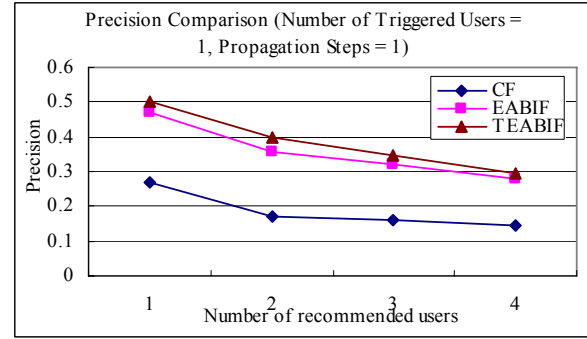
In this subsection, we will first demonstrate the recommendation quality of different algorithms when the propagation step is one, and then discuss the performances of various information propagation models.

4.2.1 Recommendation Quality

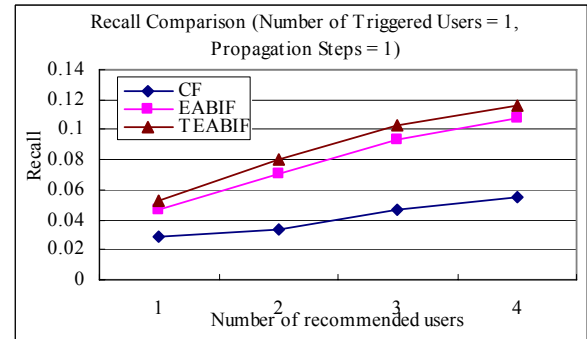
In each experiment, we trigger an user, i.e., change his or her mode from *inactive* to *active*, who are the earliest to access a particular document, and then predict who else will also access this document by comparing the information propagation probabilities.

Figure 6 (a) and (b) compare the average precision and recall of Collaborative Filtering based on Cosine Similarity (labeled by CF), early adoption based information flow network (labeled by EABIF) and topic-sensitive early adoption based information flow network (labeled by TEABIF). The number of propagation steps equals to 1, in which only the pairwise information flow is considered. The recommendations are achieved by ranking the probabilities in the row which corresponds to how information flows from this triggered user to the others in the matrix $\bar{\mathbf{F}}$. Comparing to CF, on average, EABIF is 91.0% better and TEABIF is 108.5% better in terms of precisions. Comparing to EABIF, TEABIF improves 19.2% further. In terms of recalls, on average, EABIF is 87.1% better and TEABIF is 112.8% better than CF. Comparing to EABIF, TEABIF improves 29.5% further.

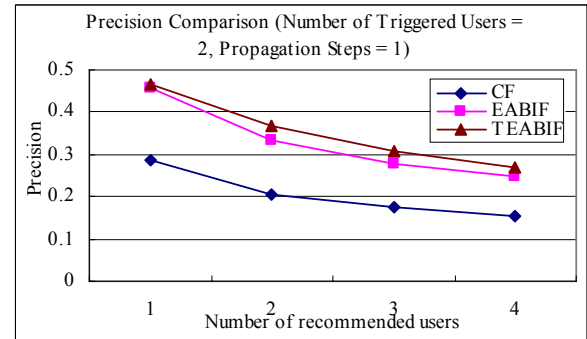
In the second experiment, we trigger users who are among the earliest two to access a particular document, and then predict whom else will also likely access this document. Note the



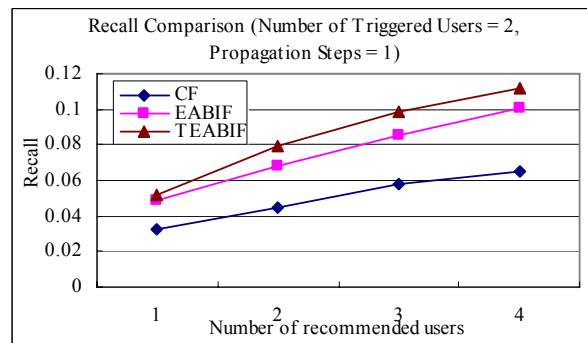
(a)



(b)



(c)



(d)

Figure 6: Average Precision and Recall of EABIF and TEABIF compared to the Collaborative Filtering

precisions and recalls cannot be directly compared to the results in the first experiment because the ground truth is different in these situations. For instance, there are totally ten users accessing one document in the test period, if we trigger one user, the other nine users are ground truth for recommendation. However, if we trigger two users, the other eight users are the ground truth.

By the “word-of-mouth” effects, each user’s tendency to become active increases monotonically as more neighbors become active. The probability of one user will access the document among all the users is the summation of the according probabilities in the rows which are corresponding to how information flows from the triggered users to this user in the matrix \bar{F} . Then the system recommends by ranking the probabilities of all users.

When two users serve as triggers (Figure 6(c,d)), we obtain the similar performance in the first experiment. On average, EABIF is 62.2% better than CF, and TEABIF is 72.4% better than CF in terms of precision. Comparing to EABIF, TEABIF improves 16.4% further. Comparing to CF, on average, EABIF is 50.2% better, TEABIF is 66.8% better in terms of recall. Comparing to EABIF, TEABIF improves 33.1% further.

4.2.2 Propagation Performance

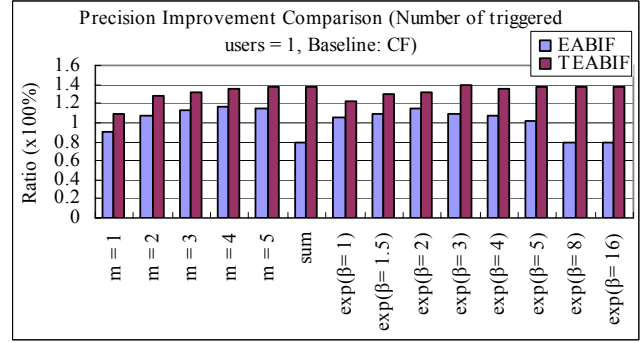
We further compare the performance of different propagation models, including changing the number of propagation steps from 1 to 5 (labeled as $m=1,2,\dots,5$), direct summation, and exponential weighted summation when $\beta=1,1.5,2,\dots,5,8$, and 16.

Figure 7 compares the precision and recall improvement comparing to CF, i.e., $\frac{Performance_{IF} - Performance_{CF}}{Performance_{CF}}$ when the

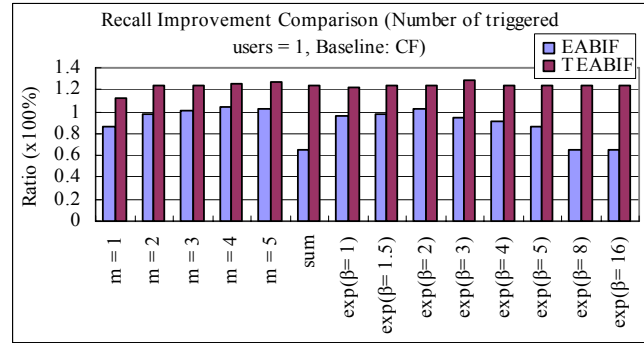
number of triggered users is one (Figure 7(a,b)) or two (Figure 7(c,d)). Among all propagation models, when the number of triggered users is one (Figure 7(a,b)), TEABIF with exponential weighted summation ($\beta=3$) achieves the best performance and improves 108.5% on precision and 116.9% on recall comparing to CF. Using the same propagation model (exponential weighted summation ($\beta=3$)), EABIF improves 91.0% in terms of precision, and 95.5% in terms of recall comparing to CF. Comparing to the number of propagation steps equals one, EABIF improves 25.3%, and TEABIF improves 28.3% in terms of precision; EABIF improves 17.1%, and TEABIF improves 9.8% in terms of recall.

When the number of triggered users equals to two, Figure 7(c,d) compares the precision and recall improvement comparing to CF. Among all the propagation models, TEABIF with exponential summation with $\beta=1$ achieves the best precision. Comparing to CF, EABIF improves 69.1%, and TEABIF improves 81.1% in terms of precision; EABIF improves 54.5%, and TEABIF improves 69.4% in terms of recall. Comparing to the number of propagation steps equals one, EABIF improves 11.1%, and TEABIF improves 12.0% on precision; and EABIF improves 8.6%, TEABIF improves 5.4% on recall.

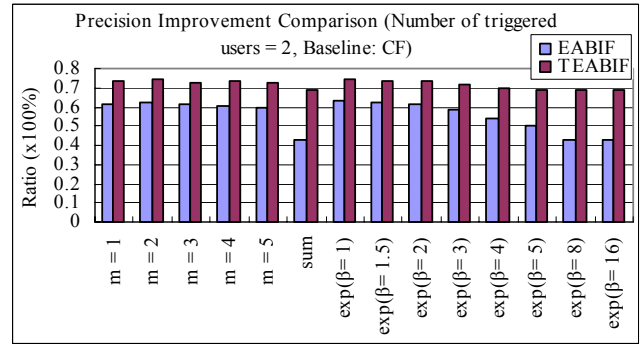
In Figure 7, EABIF with direct summation achieves the worst performance among all the propagation models we tested. However, the performance which TEABIF with direct summation achieves is much better than EABIF with direct summation even though it is not among the best performances. This phenomenon verifies again that users’ adoption patterns are topic sensitive. An



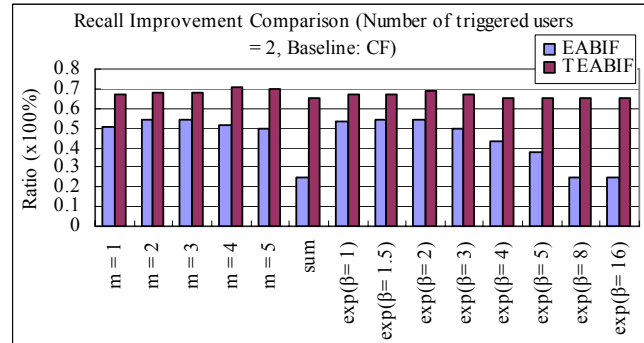
(a)



(b)



(c)



(d)

Figure 7: Precision and Recall improvement (relative to CF) comparison when the number of propagation steps change from $m=1\sim 5$, the direct summation (as “sum”), and the exponential weighted summation with the weighting parameter $\beta=1,1.5,2,\dots,5,8$, and 16.

early adopter of one topic is not necessary to be the early adopter of another topic. Therefore, long-path propagation is not as effective for the topic-independent model – EABIF, as that of TEABIF which considers early adoption patterns regarding the same topic. The direct summation and exponential weighted summation when $\beta \geq 8$ have achieved the stationary distribution of the Markov chain.

Among all propagation models we have tested in Figure 7, TEABIF with $m = 4\sim 5$ achieves similar performance to the best performance, which can serve as a simpler approximation to the more complex models.

The direct summation and the exponential summation with large β of all the paths with different lengths converge to the stationary distribution of the Markov chain. Long-path propagation is not as effective for the topic-independent model – EABIF, as that of TEABIF which considers early adoption patterns within the same topic. Overall, short-path propagations play more important roles than long-path propagations because the direct relationship among the users is more reliable.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we model user's adoption patterns as an information flow network for a recommendation system. By comparing the timestamps when users access documents, an early adoption based information flow (EABIF) network is proposed. Furthermore, observing that adoption is typically category-specific, we propose a topic-sensitive early adoption based information propagation (TEABIF) network, in which users' adoption patterns are clustered regarding the topics of the documents they accessed. Three information propagation models are proposed to predict how specific information will propagate through the network, given known triggers. Thus we proposed a novel perspective that a recommendation problem can be addressed as: once an item is adopted by one or more users, who else in the network will also likely to adopt it.

In our experiments, comparing to traditional Collaborative Filtering, EABIF with one-step propagation prediction improves 91.0% and 87.1%, and TEABIF with one-step propagation prediction improves 108.5% and 112.8% on precision and recall respectively. Furthermore, comparing to the one-step propagation prediction, EABIF with information propagation models improves 25.3% and 17.1%, and TEABIF with information propagation models improves 28.3% and 9.8% further on precision and recall respectively. When more users serve as the triggers, similar performance is achieved. In conclusion, the proposed algorithms demonstrate effective recommendation results.

Future work includes extending the model to other domains where ratings are discrete and also comparing with the item-item recommenders. We will also apply the information flow models for modeling topic propagation in blogspace.

6. ACKNOWLEDGEMENT

We would like to thank Koji Hino and Yusuf Akca for collecting and preprocessing the data. This work was supported by funds from NEC Laboratories America.

7. REFERENCES

[1] G. Adomavicius and A. Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *IEEE Trans. Knowl. Data Eng.* 17(6): 734-749, 2005.

[2] J. Schafer, J. A. Konstan and J. Riedl, E-commerce Recommendation Applications, *Journal of Data Mining and Knowledge Discovery*, 2001.

[3] G. Linden, B. Smith, and J. York, Amazon.com recommendations: item-to-item collaborative filtering, *IEEE Internet Computing*, 7(1), 2003.

[4] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, GroupLens: An open architecture for collaborative filtering of netnews, In *Proc. of the ACM Conf. on Computer Supported Cooperative Work*, 175-186, 1994.

[5] U. Shardanand, and P. Maes, Social Information Filtering: Algorithms for Automating "Word of Mouth", In *Proc. of CHI*, 1995.

[6] E. M. Rogers, *Diffusion of Innovations*, The Free Press: New York, 1995.

[7] X. Song, C.-Y. Lin, B. L. Tseng and M.-T. Sun, Modeling Evolutionary and Relational Behaviors for Community-based Dynamic Recommendation, In *Proc. of the SIAM Intl. Conf. on Data Mining*, 2006

[8] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, Propagation of Trust and Distrust, In *Proc. of the Intl. World Wide Web Conf.*, 2004.

[9] X. Song, C.-Y. Lin, B. L. Tseng and M.-T. Sun, Modeling and Predicting Personal Information Dissemination Behavior, In *Proc. of the ACM SIGKDD*, 2005.

[10] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, Mining Newsgroup Using Networks Arising From Social Behavior, In *Proc. of the Intl. World Wide Web Conf.*, 2003.

[11] M. Richardson and P. Domingos, Mining Knowledge-Sharing Sites for Viral Marketing, In *Proc. of the ACM SIGKDD*, 2002.

[12] P. Domingos and M. Richardson, Mining the Network Value of Customers, In *Proc. of the ACM SIGKDD*, 2001.

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

[14] V. Mahajan and E. Muller, When Is It Worthwhile Targeting the Majority Instead of the Innovators in a New Product Launch?, *Journal of Marketing Research*, 35, pp. 488-95, Nov 1998.

[15] V. Mahajan, E. Muller and R. K. Srivastava, Determination of Adopter Categories by Using Innovation Diffusion Models, *Journal of Marketing Research*, 27, 37-50, 1990.

[16] T. Valente, *Network Models of the Diffusion of Innovations*, Hampton Press, 1995.

[17] P. Rusmevichientong, S. Zhu and D. Selinger, Identifying Early Buyers from Purchase Data, In *Proc. of the ACM SIGKDD*, 2004

[18] J. O'Donovan and B. Smyth, Trust in recommender systems. In *Proc. of the 10th Intl. Conf. on Intelligent User Interfaces*, 167-174, 2005.

[19] D. Gruhl, R. Guha, D. Liben-Nowell and A. Tomkins, Information diffusion through blogspace, In *Proc. of the Intl. World Wide Web Conf.*, 2004

[20] J. Scott, *Social Network Analysis: A Handbook*. Sage Publications, London, 2000.

[21] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, *Journal of the ACM*, 46(5):604-632, 1999.

[22] P. Baldi, P. Frasconi and P. Smyth, *Modeling the Internet and the Web: Probabilistic Methods and Algorithms*, John Wiley and Sons, 2003.

[23] S. Brin and L. Page, The anatomy of a large-scale hypertextual web search engine, *Computer Networks*, 30(1-7):107-117, 1998.

[24] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, Philadelphia, PA: SIAM, 2000.

[25] A. N. Langville and C. D. Meyer, Deeper inside PageRank, *Internet Mathematics*, 1(3):335-400, 2004.

[26] D. Blei, A. Ng, and M. Jordan, Latent Dirichlet allocation, *Journal of Machine Learning Research*, 3:993-1022, Jan 2003.

[27] D. Widyantoro, T. Ioerger and J. Yen, Learning User Interest Dynamics with a Three-Descriptor Representation, *Journal of the American Society for Information Science and Technology*, 52(3): 212-225, 2001.

[28] O. Nasraoui, C. Cardona, C. Rojas and F. Gonzalez, Mining Evolving User Profiles in Noisy Web Clickstream Data with a Scalable Immune System Clustering Algorithm, in *KDD Workshop on Web mining as a Premise to Effective and Intelligent Web Applications*, 2003.

[29] G. Adomavicius, R. Sankaranarayanan, S. Sen and A. Tuzhilin, Incorporating contextual information in recommender systems using a multidimensional approach, *ACM Transactions on Information Systems*, 32(1): 103-145, 2005.

[30] T. Griffiths and M. Steyvers, Finding Scientific Topics, In *Proc. of the National Academy of Sciences*, 5228-5235, 2004.

[31] J. S. Breese, D. Heckerman and C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, In *Proc. of the Conf. on Uncertainty in Artificial Intelligence*, 43-52, 1998.