

# A Heat Diffusion Method for Mining Web Graphs for Recommendations Using Recommendation Algorithm

S. Aarthi <sup>1</sup> and Mr. S. Sampath <sup>2</sup>

<sup>1</sup> M.Phil Research Scholar, Department of Computer Science, <sup>2</sup> Assistant Professor, Department of Computer Science.

P.K.R Arts College for Women, Gobichettipalayam, India.

## Abstract

*Recommendation techniques have become increasingly essential. The different kinds of recommendations are made on the Web workaday, including images, music, books recommendations, query suggestions, etc. This paper, providing a common framework on mining Web graphs for recommendations using heat diffusion method, first propose a Recommendation algorithm the algorithm aggregates items from these similar customers eliminates items the user has already rated, and recommends the remaining items to the user. Which propagates similarities between different recommendations like image recommendation, the proposed algorithm can be utilized in many recommendation tasks on the World Wide Web, including image recommendations, etc. The observational Analysis on huge datasets shows the promising future of our work.*

**Index Terms**— Recommendation, query suggestion, image recommendation.

## 1 INTRODUCTION

Web information, is to organize and utilize the information effectively and efficiently has become more and more critical. This is especially important for Web related applications since user-generated information is more freestyle and more useful information from these data sources. In order to satisfy the information needs of Web users and improve the user experience in many Web applications [10]. Collaborative Filtering is a technique that automatically predicts the interest of an active user by collecting rating information, from other items. The collaborative filtering is that the active User will prefer those items which other similar users prefer [8].

A general graph recommendation algorithm can solve many recommendation problems on the Web [10] and it is inefficient to design recommendation algorithms for different recommendation tasks. The most of these recommendation problems have some features, where a general framework is needed to recommend the tasks on the Web. The existing methods are refined and require tuning a large number of parameters.

## 2 RELATED WORKS

Recommendation on the web is a general term representing a specific type of information filtering technique that attempts to present information items like queries, movies, books, images, WebPages etc. That is likely of interest to the users. In this section, the review several work related to recommendation, including query suggestion techniques, collaborative filtering, and image recommendation methods.

### 2.1 Posting the opinion

The system can get the different opinions from various users and they may post their comments on the user's web page and those comments are stored in a web usage warehouse and the stored comments may be positive or negative.

### 2.2 Rating Prediction

The ratings of unrated items are estimated based on the available information (typically using known user ratings and possibly also information about item content) using recommendation algorithm. The techniques typically calculate recommendations based directly on the previous user activities (e.g., rating values or transactional data). For each user, ranks all the predicted items according to the predicted rating value ranking the candidate (highly

predicted) items based on their predicted rating value, from lowest to highest as a result choosing less popular items.

### 2.3 Collaborative Filtering

Filtering techniques predict the ratings of dynamic users based on the ratings of their similar users, and item-based approaches predict the ratings of active users based on the computed information of items similar to those chosen by the active user.

### 2.4 Ranking Approach

Ranking items according to the rating variance of neighbors of a particular user for a particular thing there be a list of similar rank approaches that improve recommendation diversity by recommending items other than the ones with topmost predicted rating values to a user.

### 2.5 Query Suggestion

Query Suggestion is a technique widely employed by commercial search engines to provide related queries to users' information need. In this section, it demonstrates how the method can be the similar queries based on the users' information need.

## 3 PROPOSED APPROACH

Recommender systems apply knowledge discovery techniques to the problem of making personalized recommendations on the web. The tremendous growth in the amount of available information and the number of visitors to Web sites in recent years poses some key challenges for recommendation systems for producing high quality recommends, of users and items and achieving high coverage in the face of dataset [2]. In recommender systems the amount of work increases with the number of participants in the web, the technologies are needed that can quickly produce high quality recommendations, even for very large problems, using item-based collaborative filtering techniques. Item-based techniques first examine the user-item matrix to identify relationships between dissimilar items, and then indirectly use these relationships to compute recommendations for users [2].

### 3.1 Recommendation Algorithm

Recommendation algorithm is based on user interest Websites, where they use input about a customer's interests to generate a list of recommended search [6]. Recommendation algorithms provide an effective form of by creating scalable over very large customer bases and product catalogs, requires only sub second processing time to generate online recommendations for all users regardless of the number of ratings.

Most recommendation algorithms start by finding a set of customers rated items of the user. The recommendation algorithm aggregates items from these similar customers, and eliminates the items that user has already rated, and recommends the left items to the user [6].

#### 3.1.1 Item to Item Collaborative Filtering

Item-to-item collaborative filtering finds the massive data sets and produces high-quality recommendations in real time.

##### How It Works

Item-to-Item collaborative filtering matches each of the user's purchased and rated items to like items, then aggregates those similar items into a suggestion list. To determine the most related match for a given item by finding items that customers tend to search together. However, many product pairs have no common users, and thus the approach is inefficient in terms of processing time. The following iterative algorithm defines a better approach by calculating the similarity between a single product and all related products:

```

For each item in product catalog, I1
  For each customer C who purchased I1
    For each item I2 purchased by
      Customer C
        Record that a customer purchased
          I1
        And I2
        For each item I2
          Compute the match between I1 and I2
  
```

It's possible to compute the similarity between two items in various ways, but a common method is to use the rated item. The algorithm finds items similar to each of the user's ratings, aggregates those items, and then recommends the most popular items [6]. The estimation is very fast, depending only on the number of items the user purchased or rated.

#### 3.1.2 Image Recommendation

Besides query suggestion another interesting recommendation application on the Web is image recommendation system. It may focus on recommending interesting images to Web users based on users prediction. Normally, these systems ask users to rate some images as they like or dislike, and then suggest images to the users based on the tastes of the users [10].



(a)Seed Image (b)Suggestion 1 (c) Suggestion 2

Fig.1 Examples for Image Recommendation

## 4 HEAT DIFFUSION

Heat diffusion is a natural phenomenon. Always the heat flows from a position with high temperature to low temperature. It has been successfully applied in various domains such as classification and dimensionality reduction problems [9].

In this paper, the model diffusion of innovations as processes of heat diffusion. Actually, the process of people influencing others is very similar to the heat diffusion phenomenon. In a social network, the innovators and early adopters of a product or innovation act as heat sources, and have a very high amount of heat to the margin of this social network, and the laggards adopt this product or innovation.

### 4.1 Diffusion on Undirected Graphs

Consider an undirected graph  $G = (V, E)$ , where  $V$  is the vertex set, and  $V = \{v_1, v_2, \dots, v_n\}$ .  $E = \{(v_i, v_j) \mid \text{there is an edge between } v_i \text{ to } v_j\}$  is the set of all edges. The edge  $(v_i, v_j)$  is considered as a pipe that connects nodes  $v_i$  and  $v_j$ . The value  $f_i(t)$  describes the heat at node  $v_i$  at time  $t$ , beginning from an initial distribution of heat given by  $f_i(0)$  at time zero.  $f(t)$  denotes the vector consisting of  $f_i(t)$ . It constructs the model as follows. Suppose, at time  $t$ , each node  $i$  receives an amount  $M(i, j, t, \Delta t)$  of heat from its neighbor  $j$  during a time period  $\Delta t$ . The heat  $M(i, j, t, \Delta t)$  should be proportional to the time period  $\Delta t$  and the heat difference  $f_j(t) - f_i(t)$ . Moreover, the heat flows from node  $j$  to node  $i$  through the pipe that connects nodes  $i$  and  $j$ . Based on this consideration, assume that  $M(i, j, t, \Delta t) = \alpha(f_j(t) - f_i(t))\Delta t$ , where  $\alpha$  is the thermal conductivity-the heat diffusion constant, the node  $i$  between time  $t+\Delta t$  and time  $t$  will be equal to the sum of the heat that it receives from all its neighbors.

This is formulated as:

$$\frac{f_i(t+\Delta t) - f_i(t)}{\Delta t} = \alpha \sum_{j: (v_j, v_i) \in E} (f_j(t) - f_i(t)), \quad (1)$$

Where  $E$  is the set of edges, to find a closed form solution to Eq. (1), express it in a matrix form:

$$\frac{f(t + \Delta t) - f(t)}{\Delta t} = \alpha(H - D)f(t), \quad (2)$$

Where

$$H_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \text{ or } (v_j, v_i) \in E, \\ 0, & i=j \\ 0, & \text{otherwise} \end{cases}, \quad (3)$$

And

$$D_{ij} = \begin{cases} d(v_i), & i=j, \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Where,  $d(v_i)$  is the degree of node  $v_i$ . From the definition, the matrix  $D$  is a diagonal matrix. In order to generate a more generalized representation, normalize all the entries in matrices  $H$  and  $D$  by the degree of each node. The matrices  $H$  and  $D$  can be modified to

$$H_{ij} = \begin{cases} 1/d(v_i), & (v_i, v_j) \in E, \\ 0, & i=j \\ 0, & \text{otherwise} \end{cases}, \quad (5)$$

And

$$D_{ij} = \begin{cases} 1, & i=j, \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

In the limit  $\Delta t \rightarrow 0$ , this becomes

$$\frac{d}{dt}f(t) = \alpha(H - D)f(t) \quad (7)$$

Solving this derived equation, as:

$$f(t) = e^{\alpha(H-D)t}f(0), \quad (8)$$

where  $d(v)$  denotes the degree of the node  $v$ , and  $e^{\alpha(H-D)t}$  could be extended as:

$$e^{\alpha(H-D)t} = 1 + \alpha(H-D)t + \frac{\alpha^2}{2!}(H-D)^2 t^2 + \frac{\alpha^3}{3!}(H-D)^3 t^3 + \dots, \quad (9)$$

The matrix  $e^{\alpha(H-D)t}$  is called the diffusion kernel in the sense that the heat diffusion process continues infinitely many times from the initial heat diffusion.

### 4.2 Diffusion on Directed Graphs

The above heat diffusion model must be modified to fit the situation where the links between Web pages are directed. On one Web page, when the page-maker creates a link (a; b) to another page b, he actually forces the energy flow, for example, people's click-through activities, to that page, and so there is added energy imposed on the link. As a result, heat flows in a one-way manner, only from a to b, not from b to a. Based on such consideration, we modified the heat diffusion model on an undirected graph as follows.

Consider a directed graph  $G = \{V, E, W\}$ , where  $V$  is the vertex set, and  $V = \{v_1, v_2, v_n\}$ .  $W = \{w_{ij} \mid \text{where } w_{ij} \text{ is the probability that edge } (v_i, v_j) \text{ exists}\}$  or the weight that is associated to this edge.  $E = \{(v_i, v_j) \mid \text{there is an edge from } v_i \text{ to } v_j \text{ and } w_{ij} > 0\}$  is the set of all edges. On a directed graph  $G(V,E)$ , in the pipe  $(v_i, v_j)$ , heat flows only from  $v_i$  to  $v_j$ . Suppose at time  $t$ , each node  $v_i$  receives  $RH = RH(i, j, t, \Delta t)$  amount of heat from  $v_j$  during a period of  $\Delta t$ , make three assumptions: (1)  $RH$  should be proportional to the time period  $\Delta t$ ; (2)  $RH$  should be proportional to the heat at node  $v_j$ ; and (3)  $RH$  is zero if there is no link from  $v_j$  to  $v_i$ . As a result,  $v_i$  will receive  $\sum_j: (v_j, v_i) \in E \sigma_j f_j(t) \Delta t$  amount of heat from all its neighbors that point to it. At the same time, node  $v_i$  diffuses  $DH(i, t, \Delta t)$  amount of heat to its subsequent nodes.

Assume that: (1) The heat  $DH(i, t, \Delta t)$  should be proportional to the time period  $\Delta t$ ; (2) here  $DH(i, t, \Delta t)$  should be relative to the heat at node  $v_i$ ; (3) Each node has the same ability to diffuse heat; (4) The heat  $DH(i, t, \Delta t)$  should be proportional to the weight assigned between node  $v_i$  and its subsequent nodes.

As a result, node  $v_i$  will diffuse  $\propto w_{ij} f_i(t) \Delta t / \sum_{k:(i,k) \in E} w_{ik}$  amount of heat to each of its subsequent nodes  $v_j$ , and each  $v_j$  should receive  $\propto w_{ij} f_i(t) \Delta t / \sum_{k:(i,k) \in E} w_{ik}$  amount of heat from node  $v_i$ . Therefore  $\sigma_j = \alpha w_{ji} / \sum_{k:(i,k) \in E} w_{jk}$ . In the case that the out degree of node  $v_i$  equals zero, assume that this node will not diffuse the heat. The heat deviation at node  $v_i$  between time  $t+\Delta t$  and  $t$  will be equal to the sum of the heat receives. It is defined as

$$\frac{f_i(t+\Delta t) - f_i(t)}{\Delta t} = \alpha (-T_i f_i(t) + \sum_{j:(v_j, v_i) \in E} \frac{w_{ji}}{\sum_{k:(j,k) \in E} w_{jk}} f_j(t)), \quad (10)$$

where  $T_i$  is a flag to identify whether node  $v_i$  has any outlinks.

Solving it, obtain  $f(1)e^{\alpha(H - D)}f(0)$ , (11)

Where

$$H_{ij} = \begin{cases} w_{ji} / \sum_{k:(j,k) \in E} w_{jk}, & (v_i, v_j) \in E, \\ 0, & i=j \\ 0, & \text{otherwise} \end{cases}, \quad (12)$$

And

$$D_{ij} = \begin{cases} T_i & i = j \\ 0, & \text{otherwise} \end{cases}. \quad (13)$$

## 5 SYSTEM ARCHITECTURE

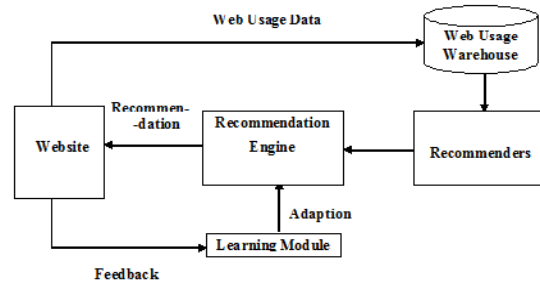


Fig.2 System Architecture

Fig.2 is the System Architecture which shows the recommendation websites to the users according to their interests.

### Website

Based on the user’s query, the recommendation engine will recommend some websites to the user. So the user can easily view the latest and the top rated websites.

### Recommendation Engine

Recommendation Engine contains the history of information of the users who fetched the previous recommended websites and their ratings. So based on the pervious users history this engine will recommend websites or pages to the new user.

### Learning Module

Learning module gets the feedback from the websites which is used by the users and updates the user’s information and their rating for particular websites. This Learning module is mainly focused on the new user, because the new user may like some other websites so the ratings from the new user are updated periodically.

### Recommenders

Recommenders collect the information from the data warehouse and list the top rated websites to the user. This recommenders use Ranking and Collaborative Filtering Techniques to filter the top rated websites.

### Web Usage Warehouse

Web Usage Warehouse collects the web usage data from website which is visited by the users and saves the all details of the user’s information. If a user likes the same websites as the previous user, this data warehouse will retrieve the information and recommends the saved information.

## 6 FUTURE WORKS

### A. Search Results Improvement

To use heat values in query suggestion. These values not only can be used in query suggestions. The Top Web sites given the queries “HQPictures”, “Flicker”, “Photobucket” and “Photospace”, For example, the ranking order for “HQPictures” is different from all of the results retrieved by commercial search engines the search results can be greatly improved since they are the representations of the implicit votes of all the search users.

Query	Rank	Websites	Heat Values
Google	1	www.HQPictures.com	0.6237
	2	www.flicker.com	0.1051
	3	www.Photobucket.com	0.0633
	4	www.photospace.com	0.0141
Query	Rank	Websites	Heat Values
Yahoo	1	www.shutterstock.com	0.5337
	2	www.fineartamerica.com	0.2000
	3	www.imagehousing.com	0.0837
	4	www.freedigitalphotos.com	0.0313

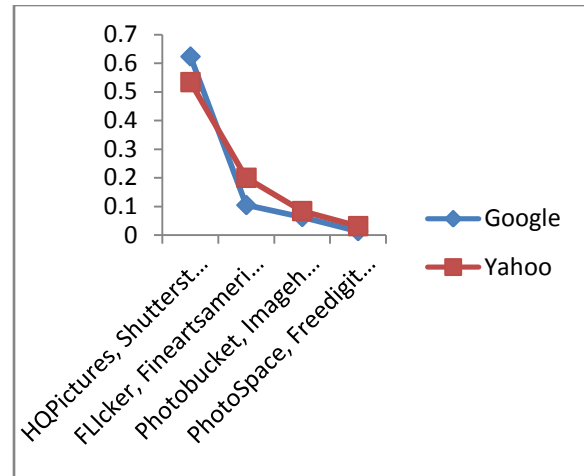
Table 1: Search Results Improvement

### B. Social Recommendation

Since our model is quite general, can apply it to more applications and complicated graphs such as Social Recommendation problem, the volatile growth of Web applications, Social recommendation, produces suggestions by incorporating users' social network information, is becoming to be an essential feature for the next generation of Web applications.

## 6 EXPERIMENTAL ANALYSIS

The experimental analysis on large data sets shows the promising future of our work that is time-consuming and inefficient to design recommendation algorithm for different recommendation tasks. It satisfies the information needs of web user to improve the user experience in many web pages. It provide an effective form of by creating scalable over very large customer bases and product catalogs, requires only sub second processing time to generate online recommendations for all users regardless of the number of ratings. The observational analysis on several large scale Web data sources shows the promising future of this approach.



## 8 CONCLUSION AND FUTUREWORK

In this paper, a Recommendation algorithm, aggregates items from these similar customers, eliminates items the user has already rated, and recommends the remaining items to the user which propagates similarities between different recommendations then illustrate how to generalize different recommendation problems into a framework. This is a general framework which can basically be adapted to most of the Web graphs for the recommended tasks, such as query suggestions, image recommendations, etc. The generated recommends are related to the inputs. The observational analysis on several large scale Web data sources shows the promising future of this approach.

## 9 REFERENCES

- [1] A.S. Das, M. Datar, A. Garg, and S. Rajaram, “Google News Personalization: Scalable Online Collaborative Filtering,” WWW ’07: Proc. 16th Int’l Conf. World Wide Web, pp. 271-280, 2007.
- [2] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, “Item-Based Collaborative Filtering Recommendation Algorithms,” WWW ’01: Proc. 10th Int’l Conf. World Wide Web, pp. 285-295, 2001.
- [3] B. Velez, R. Weiss, M.A. Sheldon, and D.K. Gifford, “Fast and Effective Query Refinement,” ACM SIGIR Forum, vol.31 (SI) 6-15, 1997.
- [4] B. Zhang, H. Li, Y. Liu, L. Ji, W. Xi, W. Fan, Z. Chen, and W.-Y. Ma, “Improving Web Search Results Using Affinity Graph,” SIGIR ’05: Proc. 30th Ann. Int’l ACM



- SIGIR Conf. Research and Development in Information Retrieval, pp. 504-511, 2005.
- [5] B.J. Jansen, A. Spink, J. Bateman, and T. Saracevic, "Real Life Information Retrieval: A Study of User Queries on the Web," ACM SIGIR Forum, vol. 32, no. 1, pp. 5-17, 1998.
- [6] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, Jan./Feb. 2003.
- [7] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma, "Query Expansion by Mining User Logs," IEEE Trans. Knowledge Data Eng., vol. 15, no. 4, pp. 829-839, July/Aug. 2003.
- [8] H. Ma, I. King, and M.R. Lyu, "Effective Missing Data Prediction for Collaborative Filtering," SIGIR '07: Proc. 30<sup>th</sup> Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 39-46, 2007.
- [9] H. Yang, I. King, and M.R. Lyu, "DiffusionRank: A Possible Penicillin for Web Spamming," SIGIR '07: Proc. 30<sup>th</sup> Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 431-438, 2007.
- [10] Hao Ma, Irwin King and Michael R. Lyu, "Mining Web Graphs for Recommendations", Vol 24, No.6, June, 2012.
- [11] J.D. Lafferty and G. Lebanon, "Diffusion Kernels on Statistical Manifolds," J. Machine Learning Research, vol. 6, pp. 129-163, 2005.
- [12] J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl, "Evaluating Collaborative Filtering Recommender Systems," ACM Trans. Information Systems, vol. 22, no. 1, pp. 5-53, 2004.
- [13] Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: The Adaptive Web, pp. 291-324. Springer Berlin / Heidelberg (2007).