8-1-2010

Friend Suggestion and Friend Browsing in Web 2.0 Applications

Michael Chau
School of Business The University of Hong Kong, mchau@business.hku.hk

Friend Suggestion and Friend Browsing in Web 2.0 Applications

Michael Chau
School of Business
The University of Hong Kong
mchau@business.hku.hk

ABSTRACT

Web 2.0 and social network applications have become increasingly popular. It is important for these applications to help users in maintaining their social networks by providing functions on friend suggestion and friend browsing. However, little study in this area has been reported in the literature. This paper proposes the design of two modules for friend suggestion and friend browsing. The first module is based on Hopfield Net spreading activation, while the second module is based on hyperbolic tree and self-organizing map. The proposed evaluation plan is also presented in the paper.

Keywords

Web 2.0; friend suggestion; friend browsing; spreading activation, hyperbolic tree, self-organizing map.

1. INTRODUCTION

Web 2.0 applications have become very popular in recent years. Many Web 2.0 sites, such as Facebook, MySpace, Friendster, and Twitter, allow users to create and maintain their social network online. There has been a significant amount of research in the area of Web 2.0 applications and social network analysis. Many of these studies focus on identifying the important nodes in a network (e.g., Chau & Xu, 2007) or studying the effectiveness and efficiencies of the network (e.g., Cao et al., 2009). In addition, there are some studies that apply data mining on social networks.

Managing one's social network is an important function for users in Web 2.0 sites. This includes finding existing and new friends in the network and browsing the social network. A friend suggestion function can help a user to locate old friends (friends whom the user already knew from other media) and new friends with similar interest. A friend browsing function allows users to visualize their friend circles more clearly. Although these features exist in many social networking websites, it is suggested that there are two major areas that can be improved. First, only some data mining techniques have been applied but there are many other data mining and text mining techniques that have not been tested on social network management. Second, many of the existing studies focused on finding the characteristics of the social network data, but few studies have investigated how the results should be presented to and utilized by users. The proposed study is to address these issues by applying some mining and visualization techniques on social network data and studying how this can help in social network applications and Web 2.0 sites. The study aims to answer the following research questions. First, is it feasible and effective to apply mining algorithms and visualization in friend suggestion and in friend browsing? Second, what are the characteristics of users' behaviors in searching and browsing for friends on the web? This paper proposes to design a module for friend suggestion using Hopfield Network spreading activation and a visualization module for friend browsing based on hyperbolic tree and self-organizing maps. These methods are chosen because of their wide successes in previous text and web retrieval research and their ability to help users in real tasks by network analysis and visualization. In practice, the proposed techniques can be applied in real-world social networking websites to help users in finding friends and maintaining their social networks on the web.

The rest of the paper is organized as follows. In the following section, related work will be reviewed. This is followed by the detailed design of the two modules. Lastly, the evaluation plan and potential implications of the research are discussed.

2. RELATED WORK

Web 2.0 Sites and Social Network Analysis

Web 2.0 sites, such as Facebook, MySpace, and Wikipedia, are becoming increasingly popular in recent years. These Web sites are often characterized by user-generated content and user collaboration. There are various types of social linkages and

interactions among users on these sites. Examples of these social linkages include the friend relationship in social networking sites (e.g., Facebook), the commenting relationship in blogs, and the co-purchase relationship in e-commerce sites (e.g., Amazon), among others. These linkages form the social networks that are of interest to both research studies and commercial applications (Chau & Xu, 2007; Kumar et al., 2005).

Previous research has proposed several methods to analyze relationships and identify communities in online social networks and the characteristics of these networks. For example, it has been suggested that different types of connections between online users have different characteristics and play different roles in facilitating interactions between them (Ali-Hasan & Adamic, 2007) and enable various types of network functions (Adar & Adamic, 2005).

Researchers have also employed various techniques, including graph algorithms and web structure mining methods, to extract communities in websites (Kumar et al., 2005; Fang et al., 2007). It has been reported that these communities often show significant temporal bursts around a specific topic in a period of time (Kumar et al., 2005). Recently, the topology of online social networks has also been studied, where different topological features, such as in-degree distributions, average shortest path lengths, and clustering coefficients of different online social networks, are compared (Shi et al., 2007).

Hopfield Network Spreading Activation

Hopfield Network is a type of artificial neural networks based on asynchronous parallel processing (Hopfield, 1982). Characterized by a number of locally stable states, Hopfield Network is often used to store and retrieve knowledge (Chau & Chen, 2003; Chau & Chen, 2007). Spreading activation algorithms have been proposed in several information retrieval applications over Hopfield Network (Belew, 1989; Chen & Ng, 1995). In these studies, the items to be retrieved (e.g., keywords, documents, or concepts) are represented by the neurons in the Hopfield Network. The relationships among these entities are calculated based on probabilities or similarities, and are represented by the synaptic weights between the neurons. When a search query is received, the corresponding node will be activated and this activation will spread through the network to retrieve relevant results.

Network Visualization

There exist a large number of studies in the area of visualizing network data. In general, a major issue for consideration in displaying network data is the complexity of the network. Because of the large number of nodes and the edges between them, network data can be too complex for display because of the overcrowded display (Eick, 1996). Visualization techniques often need to focus on only one part of the network. For example, a hyperbolic display allows users to focus on some data of interest while keeping some neighbor data in the contextual display (Lamping et al., 1995). Other techniques such as zooming have also been proposed.

Similarly, clustering and summarization techniques can be applied to the data for better visualization display. For example, Kohonen's self-organizing maps have been applied in visualizing web documents (Kohonen et al., 2000; Chen et al., 2002). Self-organizing maps are useful for mapping high-dimensional data into two-dimensional displays, thus making the visualization more easily understandable by users.

3. PRPOSED APPROACH

Friend Suggestion

The first step to develop a visualization model for social networks is to define a similarity measure $sim(x, y)$ for the relationship between any pair of nodes (people) x and y in the network. Many measures have been used to calculate the relationship between people in social networking websites in previous research, such as the similarity between the two people's profiles or the number of interactions between the two people. Previous research has shown the effectiveness of combining content-based and structure-based features (Chau & Chen, 2008). This study proposes to use three measures for calculating similarity between two people based on profile content, entry content, and linkage relationships. The similarities for profile content and entry content between two people can be calculated by the cosine similarity score between the content based on TFIDF (term frequency multiplied by inverse document frequency) using a vector space model (Salton, 1989). The linkage relationship can be defined based on whether the two people belong to the same community, have subscribed to each other, or have commented on each other's entry (Ali-Hasan and Adamic, 2007; Chau & Xu, 2007; Kumar et al., 2005).

Based on the similarity measures, this study proposes the use of a spreading activation algorithm to suggest relevant friends to users. The reason for using Hopfield Net is because of its previous success in applications searching for similar objects (Chau & Chen, 2007). When the friend suggestion function is used, a user is represented as a node in the network and the

neighbors of the node will be activated and suggested to the user. The activation strength will be based on the similarity score on the edge (Chen & Ng, 1995) and only nodes with an activation score about a threshold will be suggested. Afterwards the neighbors of these friends will also be activated, and the process will continue until the required number of suggested friends is reached or the activation scores of all neighbors are below the threshold. Specifically, the social network of a user is modeled as a Hopfield Network (Hopfield, 1982), where every person p_i can be defined as a neuron i. For suggesting friends to a person p_0, all current friends of this person are first activated. This set of persons at form the starting set of neurons in at time $t = 0$ and the activation score $\mu_i(0)$ for each person p_i can be defined by the similarity measure between p_0 and p_i:

$$\mu_i(0) = sim(p_i, p_0) \quad (1)$$

During the activation process, the activation score of a neuron i at time $t > 0$ is defined as follows:

$$\mu_i(t) = f_s\left(\sum_{i \neq j} sim(p_j, p_i)\mu_j(t-1)\right) \quad (2)$$

The summation term in Equation (2) represents the summation of the similarity scores of between a person p_j and p_i for all persons p_j who have a link with p_i. The function f_s is a slightly modified sigmoidal function that normalizes the summation value into the interval [0,1). When this activation score of a neuron is above a threshold, the neuron should be activated (i.e., fired). All neighbor nodes (i.e., friends) of an activated node will then be added to the network. The network converges when the difference between the average activation scores at time t and t-1 is not significantly different, or when t is larger than a certain limit.

Friend Browsing

Based on the linkage among the nodes, this study proposes two visualization methods to present the network to users, namely Social Tree and Social Map. The Social Tree is a visualization tool for exploring the social network using a tree metaphor, while the Social Map is a two-dimensional display for query clusters generated based on artificial neural networks. The Social Tree will be developed as a tool to help users visualize the social network using a tree metaphor. This study proposes to build the Social Tree using both hierarchical list and hyperbolic tree. The hyperbolic tree was developed by Xerox PARC (Lamping & Rao, 1994; Lamping et al., 1995). A hyperbolic tree radiates towards the edge of a circular space and is good at visualizing large networks of data. An example of a hyperbolic tree created by InXight Software Inc. is shown in Figure 1. Previous research has shown that while users are more familiar with the hierarchical display (e.g., the Microsoft Windows Explorer), the hyperbolic tree can allow users to focus on a particular area while keeping the overall context (Lamping & Rao, 1994; Xiang et al., 2005). A combination of the two on the same display will be implemented, which has been shown to be effective (Xiang et al., 2005). Because there is no restriction on the depth of level that the users can explore, the tool will allow users to search for related friends easily as well as to browse the social network space.

The second approach to assist user browsing is a two-dimensional display. A self-organizing map (SOM) approach (Kohonen, 1995; Kohonen et al., 2000) is used to cluster and visualize the social network. A sample SOM display is shown in Figure 2 (Chau et al., 2001). The SOM technique automatically categorizes friends into different clusters based on their similarity. The similarity scores can be calculated in the same way as described earlier in this paper. The algorithm produces a visual map consisting of different regions where each region contains similar people. Regions that are similar are located close to each other (Lin et al., 2000). A number of systems utilizing this technique have been built for various types of data and achieved great success, but there is little work on applying SOM in social network visualization. Based on its success in previous work (Zhu et al., 2010), it is believed that such a multi-layered, two-dimensional will be very useful for friend browsing in social network applications.

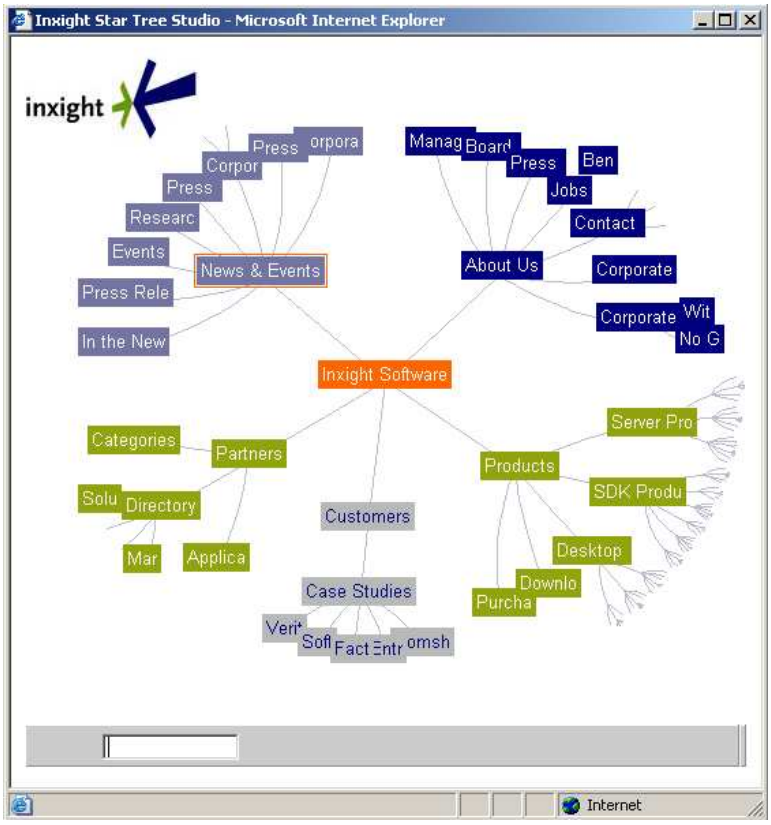

Figure 1. A hyperbolic tree for visualization (InXight Software, Inc.)

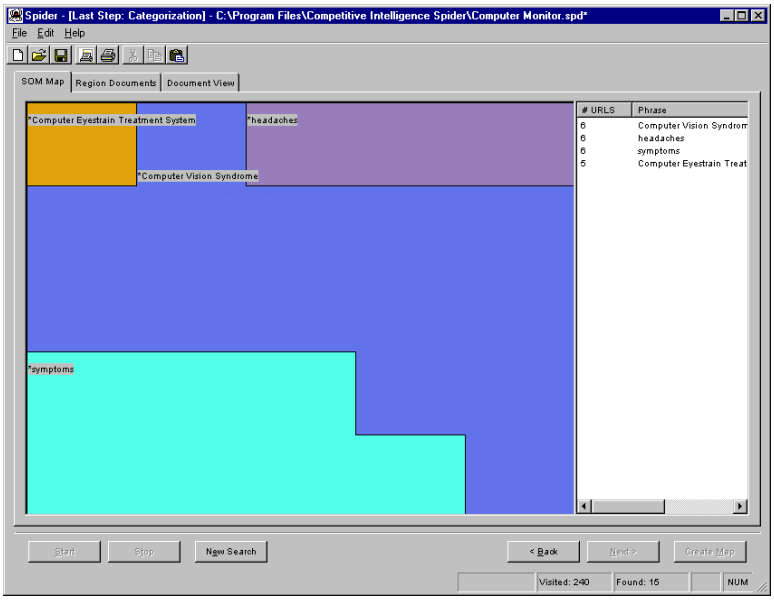

Figure 2. A self-organizing map display of clusters (Chau et al., 2001)

4. PROPOSED EVALUATION AND DISCUSSIONS

The two proposed modules will be evaluated for their effectiveness and efficiencies. First, the friend suggestion module will be evaluated in user studies. Subjects will be recruited to use the friend suggestion system in an online social network website and evaluate the friend list suggested by the proposed algorithm against the list suggested using a simple benchmark

heuristics (e.g., a list of people who are the "friends of friends" of the subject. The performance of the proposed method will be measured by precision and recall rates of the results. The speed of the proposed algorithm will also be measured and evaluated.

For the visualization module, subjects will be recruited to perform various tasks using the two proposed methods Social Tree and Social Map and compare it against a simple list display. The accuracy and time achieved by the subjects in finding their desired friends will be recorded and evaluated. User satisfaction towards the proposed visualization will also be recorded through questionnaires.

The proposed research will have the following potential contributions. First, the study evaluates the feasibility and performance of applying spreading activation in friend suggestion and two visualization techniques in friend browsing. The study will also provide a better understanding of users' behaviors in searching and browsing for friends on the web. In practice, the proposed techniques can be applied in real-world social networking websites to help users in finding friends and maintaining their social networks on the web.

ACKNOWLEDGMENT

The project has been supported in part by the HKU Seed Funding for Basic Research (#10208140 and #10400397). The author thanks the track chairs, mini-track chairs, and anonymous reviewers for their constructive comments on the paper.

REFERENCES

1. Adar, E. and L.A. Adamic (2005). "Tracking Information Epidemics in Blogspace," in *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, Compiegne, France.
2. Ali-Hasan, N.F., and Adamic, L.A. (2007). "Expressing Social Relationships on the Blog through Links and Comments," *International Conference on Weblogs and Social Media*, Boulder, Colorado, 2007.
3. Belew, R. K. (1989). "Adaptive Information Retrieval: Using a Connectionist representation to Retrieve and Learn about Documents," in *Proceedings of the 12th ACM-SIGIR Conference*, Cambridge, MA, June 1989.
4. Cao, J., Knotts, T., Xu., J. and Chau, M. (2009). "Word of Mouth Marketing through Online Social Networks," in *Proceedings of the Americas Conference on Information Systems*, San Francisco, California, USA, August 6-9, 2009.
5. Chau, M. and Chen, H. (2003). "Comparison of Three Vertical Search Spiders," *IEEE Computer*, 36(5), 56-62.
6. Chau, M. and Chen, H. (2007). "Incorporating Web Analysis into Neural Networks: An Example in Hopfield Net Searching," *IEEE Transactions on Systems, Man, and Cybernetics (Part C)*, 37(3), 352-358.
7. Chau, M. and Chen, H. (2008). "A Machine Learning Approach to Web Page Filtering Using Content and Structure Analysis," *Decision Support Systems*, 44(2), 482-494.
8. Chau, M. and Xu, J. (2007). "Mining Communities and Their Relationships in Blogs: A Study of Online Hate Groups," *International Journal of Human-Computer Studies*, 65(1), 57-70, 2007.
9. Chau, M., Zeng, D., and Chen, H. (2001). "Personalized Spiders for Web Search and Analysis," in *Proceedings of the First ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'01)*, Roanoke, Virginia, USA, June 24-28, 2001, 79-87.
10. Chen, H., Chau, M., and Zeng, D. (2002). "CI Spider: A Tool for Competitive Intelligence on the Web," *Decision Support Systems*, 34(1), 1-17.
11. Chen, H. and Ng, T. (1995). "An Algorithmic Approach to Concept Exploration in a Large Knowledge Network (Automatic Thesaurus Consultation): Symbolic Brand-and Bound Search vs. Connectionist Hopfield Net Activation," *Journal of the American Society for Information Science*, 46(5), pp. 348-369.
12. Eick, S. T. (1996). "Aspects of Network Visualization," *IEEE Computer Graphics and Applications*, 16(2), 69-72.
13. Fang, X., Sheng, O. R. L., and Chau, M. (2007). "ServiceFinder: A Method Towards Enhancing Service Portals," *ACM Transactions on Information Systems*, 25(4), 2007.
14. Hopfield, J. J. (1982). "Neural Network and Physical Systems with Collective Computational Abilities," *Proceedings of the National Academy of Science,* USA, 1982, 79(4), pp. 2554-2558.
15. Kohonen, T. (1995). *Self-organizing Maps*, Springer-Verlag, Berlin.
16. Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Honkela, J., Paatero, V., and Saarela, A. (2000). "Self Organization of a Massive Document Collection," *IEEE Transactions on Neural Networks*, 11(3), 574-585.
17. Kumar, R., Novak, J., Raghavan, P., and Tomkins, A. (2005). "On the Bursty Evolution of Blogspace," *World Wide Web: Internet and Web Information Systems* 8, 2005, pp. 159-178.
18. Lamping, J. and Rao, R. (1994). "Laying out and Visualizing Large Trees Using a Hyperbolic Space," *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '94),* pp. 13-14, Marina del Rey, California, November 1994.

19. Lamping, J., Rao, R., and Pirolli, P. (1995). "A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies," *Proceedings of the ACM CHI '95 Conference on Human Factors in Computing Systems*, Denver, Colorado, May 1995.
20. Lin, C., Chen, H. and Nunamaker, J. F. (2000). "Verifying the Proximity Hypothesis for Self-Organizing Maps," *Journal of Management Information Systems*, 16(3), 57-70, 2000.
21. Salton, G. (1989). *Automatic Text Processing*, Reading, MA: Addison-Wesley.
22. Shi, X., Tseng, B., and Adamic, L.A. (2007). "Looking at the Blogosphere Topology through Different Lenses," *International Conference on Weblogs and Social Media*, Boulder, Colorado, 2007.
23. Xiang, Y., Chau, M., Atabakhsh, H., and Chen, H. (2005). "Visualizing Criminal Relationships: Comparison of a Hyperbolic Tree and a Hierarchical List," *Decision Support Systems*, 41(1), 69-83, 2005.
24. Zhu, B., Watts, S., and Chen, H. (2010). "Visualizing Social Network Concepts," *Decision Support Systems*, 2010, forthcoming.