International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.5, No.1, January 2016 DOI:10.15693/ijaist/2016.v5i1.26-29

Web Mining Techniques for On-line Social Networks Analysis

Inamdar S. A.

G. N. Shinde*
S.R.T.M.University, Nanded.

Abstract*—This paper studies the issues around using web mining techniques for analysis of on-line social networks. Techniques and concepts of web mining and social networks analysis will be introduced and reviewed along with a discussion about how to use web mining techniques for on-line social networks analysis. Social network analysis deals with the interactions between individuals by considering them as nodes of a network (graph) whereas their relations are mapped as network edges. In addition, this paper sets out a process to use web mining for on-line social networks analysis, which can be treated as a general process in this research area. Discussions of the challenges and future research are also included.

Index terms-Web Mining, Social Networking, Social Networks Analysis, Association Rule, Visualization

I. INTRODUCTION

Web Mining is an important sub-branch of Data Mining. On-line social networking has become a very popularwhich allows users to communicate, interact and share on the World Wide Web [1]. Some on-line social networking websites have even become among the most popular sites on the web for example, *Flickr* for online photo sharing, *Youtube*for online video sharing and *Linkdin*for professional networking. Some portal-based on-line social networking websites such as *Facebook*, *MySpace*, etc integrate multiple functions. Blogs also provides good platforms for users to communicate and share. Thus, on-line social networking is now part of human life.

II. RELATED WORK

A. Social Networks Analysis

A social network is usually formed and constructed by daily and continuous communication between people and therefore includes different relationships, such as the positions, betweenness and closeness among individuals or groups [3]. A social network is a social structure made up of individuals (or organizations) called nodes which are tied (connected) by one or more specific types of interdependency, such as friendship, kinship, common interest, financial exchange, dislike, relationships of beliefs, knowledge or prestige [5].

Visualization is also a hot topic of social network analysis, and it is a suitable technique in this area. Through the visualization of social networks, their character, such as the structure of networks, the distribution of nodes, the links (relationships) between nodes and the clusters and groups in the social networks can be more easily understood [13, 14].

B. Web Mining

Web mining is an application of data mining, the technique of discovering and extracting useful information from large data sets or databases [11].

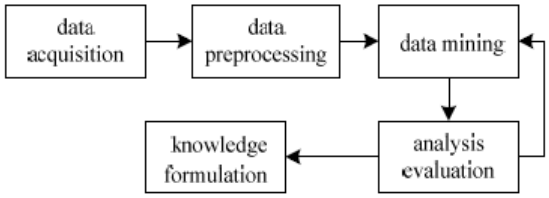

Figure 1. Web mining functions

Web Mining Techniques

The two web mining techniques used for on-line social networks analysis. The two techniques are clustering and association rule mining.

A) Clustering:

Clustering Analysis

Cluster is a group of patterns having some common properties. Patterns can be grouped into clusters by some predefined criterion. The clusters can be formed according to some criterion like distance, angle, curvature, symmetry, connectivity and intensity.

A main function of the clustering analysis is to give the users different classes of webs by clustering retrieval results, which enables them to make a rapid target, to obtain the automatic classification catalogues rapidly.

There are many categorizations of clustering algorithms such as hierarchical-based method, partitioning method, density-based method, grid-based method and model-based method and so forth.

Cluster is a group of patterns having some common properties. Patterns can be grouped into clusters by some predefined criterion. The clusters can be formed according to some criterion like distance, angle, curvature, symmetry, connectivity and intensity.

B) Association rule mining:

This is popular in traditional data mining applications, such as marketing analysis, and it is therefore also called market-basket analysis. Association rules mining has may applications other market basket analysis, including in market, customer segmentation, medicine, electronic commerce, classification, clustering, web mining, bioinformatics, and finance. In social network analysis, association rule mining can help discover the hidden relationships between nodes in a social network or even cross networks.

III. WEB MINING TECHNIQUES FOR ON-LINE SOCIAL NETWORKS ANALYSIS

A .The Three Web Mining Types for On-line Social Networks Analysis

Web content mining, text mining or natural language processing are very useful in social network analysis. For example, web content mining can categorize or classify documents on an on-line social networking website, especially articles on blogs or text forums. Article categorization is usually the first task for many social networks analyses or applications.

Web usage mining also plays an important role in social networks analysis. The usage data and user communications on an on-line social networking website can be transformed into relational data for social-networks construction [6, 12].

Web structure mining is the third kind of web mining and it is also useful for extracting and constructing social networks to extract the links from WWW, e-mail or other sources. It can also be used to analyze path length, reachability or to find structural holes, which are very basic and traditional social networks analyses. Web structure mining usually uses graphs and visualized means to represent the data about social networks, enabling the analyst to easily understand and analyze social networks [12].

IV. THE PROCESS OF WEB MINING FOR ON-LINE SOCIAL NETWORKS ANALYSIS

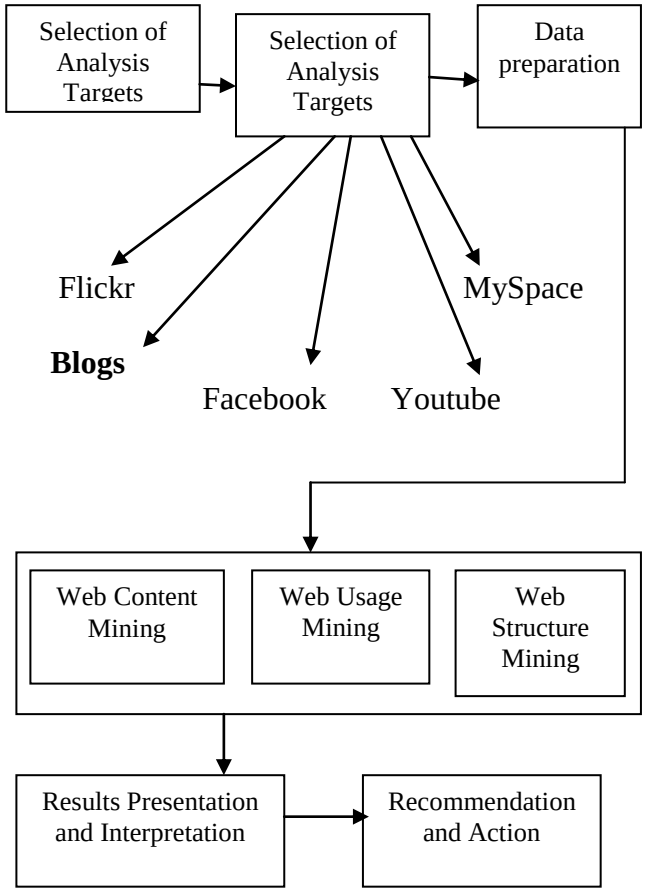

Figure 2. The general process of web mining for on- line social networks analysis

V.ABOUT BLOG

A blog (a truncation of the expression *weblog*) is a discussion or informational site published on the World Wide Web and consisting of discrete entries ("posts") typically displayed in reverse chronological order (the most recent post appears first). Previously blogs were usually the work of a single individual, occasionally of a small group, and often covered a single subject.

There are many different types of blogs, like personal blog, Corporate and organizational blogs, political blogs, health blogs, travel blogs, fashion blogs, project blogs, education blogs, classical music blogs, quizzing entblogs and legal blogs (often referred to as a blawgs) or dreamlogs, How To/Tutorial blogs are becoming increasing popular.

VI. CONCLUSION

This paper studies the application of the concept and techniques of web mining for on-line social networks. Social Network Analysis is the study of social structure. How to use the techniques of web mining for on-line social networks analysis is an interesting topic.

In addition, we will focus on how to apply the web mining techniques to some real on-line social networking websites, such as blogs.

REFERENCES

[1]. Adamic, L. A., and Adar, E. "Friends and Neighbors on the Web"Social Networks, Vol. 25, 2007, pp. 211-230

[2]. Margaret H. Dunham and Sridhar, "Data Mining, Introduction and Advanced Topics", Prenticce Hall, Publication, ISBN 81-7758-785-4, 2003.

[3]. E.Raju, K.Sravanthi, " Analysis of Social Networks Using the Techniques of Web Mining",International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012 ISSN: 2277 128X, www.ijarcsse.com.

[4]. Cooley, R., Mobasher, B. and Srivastave, J. "Web Mining: Information and Pattern Discovery on the World Wide Web" In Proceedings of the 9th IEEE International Conference on Tool with Artificial Intelligence, 1997, pp. 558-567, Newport Beach, CA, USA

[5]. Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P. and Bhattacharjee, B. "Measurement and Analysis of Online Social Networks" In Proceedings of 2007 Internet Measurement Conference, October 24-26, 2007, San Diego, California, USA, pp. 29-42

[6]. Fu, F., Liu, L., Wang, L. "Empirical Analysis of Online Social Networks in the age of Web 2.0" Physica A, Vol. 387, 2008, pp. 675-684.

[7]. ODUKOYA, O.H, ADEROUNMU, G.A.AND ADAGUNODO, E.R., "An Improved Data Clustering Algorithm for Mining Web Documents", IEEE, 2010.

[8]. Jin, Y. Z., Matsuo, Y., and Ishizuka, M. "Extracting Social Networks among Various Entities On the Web" In Proceedings of the Fourth European Semantic Web Conference, 2007

[9]. Chu-Hui Lee, Yu-Hsiang Fu, "Web Usage Mining based on Clustering of Browsing Features", Eighth International Conference on Intelligent Systems Design and Applications, IEEE Computer Society, IEEE, 2008.

[10]. R. Etemadi , N. Moghaddam, "An Approach in Web Content ining for Clustering Web Pages", Digital Information Management (ICDIM), Fifth International Conference, pp.279-284, IEEE, 2010.

[11]. Hand, D., Mannila, H., and Smyth, P. "Principles of Data Mining",MIT Press, Cambridge, MA, 2001

[12].Nowson, S., and Oberlander, J. "Identifying More Bloggers" In Proceedings of ICWSM 2007, Boulder, Colorado, USA.

[13]. Heer, J., and Boyd, D. "Vizster: Visualizing Online Social Network" In Proceedings of 2005 IEEE Symposium, October 23-25, Minneapolis, MN USA, pp.32-39, 2005.

[14]. Wasserman, S., and Faust, K. "Social Network Analysis: Method and Applications" Cambridge University Press, Great Britain, 2003.

Authors Profile

G. N. Shinde is Principal in Indira Gandhi College, Nanded, Maharashtra, INDIA. He has received "Ideal State Teacher Award" from Government of Maharashtra, India for 2008-09 and "Best Principal Award" for 2009-2010 from S.R.T.M. University, Nanded, Maharashtra. He has received M. Sc. & Ph.D. degree from Dr. B.A. M. University, Aurangabad. He has awarded BenjonjiJalnawala award for securing highest marks at B.Sc. He is Dean, faculty of Science of S. R. T. M. University, Nanded. Thirteen research scholars were awarded Ph.D. degree under his guidance. He has published more than 75 papers in the International Journals and presented more than 50 papers in International Conferences. He was the Chairperson for F-9 session of "International Conference on Computational and Experimental Science & Engineering" (ICCES2008) which was held at Honolulu, U.S.A & Development and Application of Web 2.0 Technology for Education Purpose session of "International Conference on Multimedia & ICT in Education (m-ICT2009)", April 24, 2009, (Hall 1), Lisbon (Portugal) and "Wave Propagation and Wave Interaction with media" Progress In Electromagnetic Research Symposium (PIERS), March 25, 2010, Session 3A4 (Room D) Xian, China. In his account one book is published, which is reference book for different courses. He is also member of different academic & professional bodies such as IAENG (Hon Kong), ANAS (Jordan).He is in reviewer panel for different Journals such as IEEE (Transactions on Neural Networks), International Journal of Physical Sciences (U.S.A.), Journal of Electromagnetic Waves and Applications (JEMWA, U.S.A.).His abroad Visit includes U.S.A., Thailand, Portugal, Germany, Swisszerland, Italy, Vatican City, Monaco, France, Maldives, SriLanka, U. K., Scotland, China and New Zealand. He was Chairman of Grievances Committee and member of Management Council & Senate of S.R.T.M. University, Nanded, INDIA. His research interest includes Filters, Wireless Sensor Network System, Image processing and Multimedia analysis and retrieval system.

Dr. Inamdar S.A. worked as Lecturer at School of Computational Sciences S.R.T.M. University, Nanded. She has received B.Sc., MCA. & Ph.D. degree from S. R. T. M. University, Nanded, and M.Phil. degree from Y.C.M.O.University, Nashik. She has published 5 international papers and 2 national. Her research interest include web mining, software engineering, multimedia systems.