

Role of Text mining in Selected Social Networking Sites: A review

Mrs. Shivangi S. Patil

KIT'S Institute Of Management Education & Research, Dist. Kolhapur

Dr.R.V.Kulkarni

Chh. Shahu Institute Of Business, Education & Research, Dist. Kolhapur.

Abstract: Text mining is the discovery of interesting knowledge in text documents. Social networks are rich in various kinds of contents such as text and multimedia. It is a challenging issue to find accurate knowledge (or features) in social networking sites to help users to find what they want. Social networks require text mining tools and algorithms for a wide variety of applications such as for Business – connecting with customers, Business – Networking, Marketing, Entertainment etc., The ability to apply text mining algorithms effectively in the context of text data is critical for a wide variety of applications. This paper focuses on role of text mining in Social networking and how information is searched, classified, clustered for the same. We also present avenues for future research in this area.

Keywords: Text mining, social networking

1. INTRODUCTION

“The discovery by computer of new, previously unknown information, by automatically extracting information from a usually large amount of different unstructured textual resources is known as Text mining” According to Marti Hearst, UC Berkeley. “Social Networking” as a phrase to mean the sites such as Facebook, MySpace, Twitter, etc, where people can connect with one another directly with a purpose – to keep in touch with friends. It is fascinating to see that there are so many different things going on in website at one place, and even more interestingly that they seem to all work in harmony. Text mining can be effectively used to extract data from social networking sites.

With help of text mining tool, social media is being embraced by businesses more and more. E.g, Information on facebook can be mined and can be used to deal with advertisement depending upon interest area as well as marketing and selling of products can be done. Innovative ways of utilizing these text mining tools by connecting directly with customers are being found. Not only facebook also Companies are seeing that the best way to conduct themselves online is to speak to their customers directly using these social networking sites like

Twitter being especially useful for this. Text mining can be used to increase the reputation of the Company, gets them all types of reviews especially-positive reviews and shows that they really care about the customer.

It also gives a human impact to the large corporation, meaning that potential Customers will feel much more comfortable coming back. LinkedIn exist for this purpose; for people to network within a business environment. It does for businesses; It could find you a new business deal or could even find you a new job! All this can be done by keyword search, classifying and clustering that is text mining. This paper defines what is role of text mining in social networking sites.

2. LITERATURE REVIEW

Yang, Yunyun; Akers, Lucy; Klose, Thomas [1].A high-level overview of some key text mining and visualization tools is presented in this paper to provides a comparison of text mining capabilities, perceived strengths, potential limitations, applicable data sources, and output of results, as applied to chemical, biological and patent information. In year 2010,Erik Cambria; Robert Speer, Catherine Havasi and Amir Hussain [2] uses dimensionality reduction to infer the polarity of common sense concepts and hence provide a public resource for mining opinions from natural language text at a semantic, rather than just syntactic, level. Hearst, Marti A. [3]

In this paper author first define data mining, information access, and corpus-based computational linguistics, and then discuss the relationship of these to text data mining. The intent behind these contrasts is to draw attention to exciting new kinds of problems for computational linguists. And describe examples of what he considers to be real text data mining efforts and briefly outline recent ideas about how to pursue exploratory data analysis over text.

Zanasi, Alessandro ,2009 mentioned[4] Text mining is the most advanced knowledge management technology which allow intelligence analysts to automatically analyze the content of information rich online data banks, suspected web sites, blogs, emails, chat lines, instant messages and all other digital media detecting links between people and organizations, trends of social and economic actions, topics of interest also if they are “sunk” among terabytes of information. Cohen, K. Bretonnel; Hunter, Lawrence [5] Text mining specialists continue to excel at building system components and designing datasets for evaluation; computational biologists currently appear to be much better at producing useful task definitions.

Coussement, Kristof; Van Den Poel, Dirk [6] introduces a methodology to improve complaint handling strategies through an automatic email classification system that distinguishes complaints from non-complaints. As such, complaint handling becomes less time-consuming and more successful. The classification system combines traditional text information with new information about the linguistic style of an email. [7]Authors mentions in the present study classification of proteins by basing on its primary structures.

The sequence of proteins collected in a file. The application of Text mining technique for extracting the features is proposed. An algorithm is also developed which extracts all the n-grams existing in the file of data and produced a learning file. Algorithm supplies three files, Boolean file that is a relation of existence or not existence, frequencies files and occurrences files. The applied forward selection and backward elimination method is a learning file with an accepted features numbers.[8]

In this paper author uses Co-occurrence analysis is a technique applied in text mining and the methodologies and statistical models are used to evaluate the significance of the relationship between entities such as disease names, drug names, and keywords in titles, abstracts or even entire publications. In this paper author presents a method and an evaluation on knowledge discovery of disease-disease relationships for rheumatic diseases.[9] In this paper, author reexamines the impact of reviews on economic outcomes like product sales and see how different factors affect

social outcomes such as their perceived usefulness. Authors approach explores multiple aspects of review text, such as subjectivity levels, various measures of readability and extent of spelling errors to identify important text-based features. In addition, author also examine multiple reviewer-level features such as average usefulness of past reviews and the self-disclosed identity measures of reviewers that are displayed next to a review.[10] In this paper author cuts system out subsequences from time series data. Several representative sequences are extracted from these subsequences by using clustering.

Feature patterns are acquired from these representative sequences. For this purpose, author developed a method that applies weight technique, which is often used in text mining, to time series data. The time series data are classified by using the acquired feature patterns.[11] This paper represents how text mining on relevant text corpora can be used to identify matching ontology terms of two separate ontology’s and to propose new ontology terms for a given term.[12] The aim of this paper is to give an overview of text mining in the contexts of its techniques, application domains and the most challenging issue.

The focus is given on fundamentals methods of text mining which include natural language possessing and information extraction. This paper also gives a short review on domains which have employed text mining. [13] In this paper author says there is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information(knowledge) from the rapidly growing volumes of digital data.[14] In this paper author defines how Text Mining serves as a powerful technique to manage knowledge encapsulated in large document collection.[15] In this paper author mentions Knowledge discovery from textual databases, refers generally to the process of extracting interesting and non-trivial patterns or knowledge from unstructured text documents.

It can be viewed as an extension of data mining or knowledge discovery from (structured) databases. [16] This article discusses various Text data mining: Issues, techniques, and the relationship to information access. [17] Tan, discussed how Integrating neural computation and symbolic

knowledge processing can be done.[18] Teo, C. (1998). learned user profiles for personalized information Dissemination. In proceedings, International Joint Conference on Neural Networks.[19] In 1996, Simoudis, E, provided reality check for data mining and text mining.

3. CONCLUSION

It is observed from above literature reviews that none researcher has used text mining tools to define role of text mining in social networking sites. As per various reviews, text mining is used in biomedical sciences, for virtual security, for untangling text, for understanding customer complaints on email etc., also various tools and algorithms have been built. But, we propose in this paper what role text mining plays in social networking sites and how text can be searched. Using text mining tools and algorithms we can define role played in social networking such as search keywords, classify and cluster text.

It is very necessary to understand customer interest and target those customers for Business – connecting, Business – Networking, Marketing, Entertainment etc. Text mining techniques enables us to discover and use implicit structure of texts (e.g. grammar structure) and integrate some specific natural language processing. Text mining technique can range from simple one to intermediate complexity .Among all document clustering & text summarization is effective method. Also we are going to use text mining tools like IBM Intelligent Miner for Text , Semio Map, InXight LinguistX / ThingFinder to extract data from text.

4. FUTURE RESEARCH.

The overall objective of this research is to find hidden datasets by adopting various text mining tools that will help to define role of text mining in social networking. In future the results can be expanded for finding more roles played by text mining in social networking.

REFERENCES

[1]. Yang, Yunyun; Akers, Lucy; Klose, Thomas; Barcelon Yang, Cynthia (2008). "Text mining and visualization tools – Impressions of emerging capabilities". *World Patent Information*

[2]. Erik Cambria; Robert Speer, Catherine Havasi and Amir Hussain (2010). "SenticNet: a Publicly Available Semantic Resource for Opinion Mining".

[3]. Hearst, Marti A. (1999). "Untangling text data mining". *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. pp. 3–10.

[4]. Zanasi, Alessandro (2009). "Virtual Weapons for Real Wars: Text Mining for National Security". *Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08*. Advances in Soft Computing.

[5]. Cohen, K. Bretonnel; Hunter, Lawrence (2008). "Getting Started in Text Mining". *PLoS Computational Biology* .

[6]. Coussement, Kristof; Van Den Poel, Dirk (2008). "Improving customer complaint management by automatic email classification using linguistic style features as predictors".

[7].Textmining, feature selection and datamining for proteins classification Mhamdi, F. ; Elloumi, M. ; Rakotomalala, Ricco (2004)

Information and Communication Technologies: From Theory to Applications, 2004. Proceedings. 2004 .

[8].Disease-Disease Relationships for Rheumatic Diseases: Web-Based Biomedical Textmining an Knowledge Discovery to Assist Medical Decision Making Holzinger, A. ; Simonic, K.-M. ; Yildirim, P. Computer Software and Applications Conference (COMPSAC), 2012 IEEE 36th Annual

[9].Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics Ghose, A. ; Ipeirotis,P.G. Volume: 23 , Issue: 10 Digital Object Identifier: 10.1109/TKDE.2010.188 Publication Year: 2011 , Page(s): 1498 – 1512

[10].Classification system for time series data based on feature pattern extraction Sugimura, H. Matsumoto, K. Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on Digital Object Identifier: 10.1109/ICSMC.2011.6083844 Publication Year: 2011 , Page(s): 1340 – 1345.

[11].A Corpus-Driven Approach for Design, Evolution and Alignment of Ontologies Wachter, T. ; Wobst, A. ; Schroeder, M. ;Tan,H.;Lambrix,P. Simulation Conference, 2006. WSC 06. Proceedings of the Winter Digital Object Identifier: 10.1109/WSC.2006.322932 Publication Year: 2006 , Page(s): 1595 – 1602

[12].Techniques, Applications and Challenging Issue in Text Mining Shaidah Jusoh and Hejab M. Alfawareh.(2002)

[13].Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From data mining to knowledge discovery: An Overview. In *Advances in Knowledge Discovery and Data Mining*, U.Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds., MIT Press,Cambridge, Mass., 1-36.

[14].Knowledge Management: A Text Mining Approach[1995]Ronen Feldman (ronen@instinct-soft.com),* Moshe Fresko,*Haym Hirsh,** Yonatan Aumann,* Orly Liphstat,*Yonatan Schler,* Martin Rajman,

- [15]. Feldman, R. & Dagan, I. (1995) Knowledge discovery in textual databases (KDT). In proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, August 20-21, AAAI Press, 112-117.
- [16]. Hearst, M. A. (1997) Text data mining: Issues, techniques, and the relationship to information access. Presentation notes for UW/MS workshop on data mining, July 1997.
- [17]. Tan, A.-H. (1997). Cascade ARTMAP: Integrating neural computation and symbolic knowledge processing. IEEE Transactions on Neural Networks 8(2), 237-250.
- [18]. Tan, A.-H. & Teo, C. (1998). Learning user profiles for personalized information dissemination. In proceedings, International Joint Conference on Neural Networks (IJCNN'98), Alaska, 183-188.
- [19]. Simoudis, E. (1996). Reality check for data mining. IEEE Expert 11(5).