International Journal of Research in Engineering, Science and Management
Volume-2, Issue-4, April-2019
www.ijresm.com | ISSN (Online): 2581-5792

A Survey of Student Learning from Social Media

Shailendra Pardeshi[1], Rahul Shukla[2], Shrikant Wadekar[3], Mayuri Yeole[4], Divya Patil[5], Nizare Dipak[6]
[1]*Assistant Professor, Department of Information Technology, R. C. Patel Institute of Technology, Shirpur, India*
[2,3,4,5,6]*Student, Department of Information Technology, R. C. Patel Institute of Technology, Shirpur, India*

***Abstract*: Social network analysis (SNA) has been a research focus in multiple disciplines for decades, including sociology, health care, business management, etc. Traditional SNA researches concern more human and social science aspects trying to undermine the real relationship of people and the impacts of these relationships. While online social networks have become popular in recent years, social media analysis, especially from the viewpoint of computer scientists, is usually limited to the aspects of people's behavior on specific websites and thus are considered not necessarily related to the day-today people's behavior and relationships. Conduct research to bridge the gap between social scientists and computer scientists by exploring the multi facet existing social networks in organizations that provide better insights on how people interact with each other in their professional life. Describe a comprehensive study on the challenges and solutions of mining and analyzing existing social networks in enterprise. Several aspects are considered, including system issues, privacy laws, and the economic value of social networks, people's behavior modeling including channel, culture, and social inference. Social network visualization in large-scale organization; and graph query and mining. SNA tool (SmallBlue) that was designed to overcome practical challenges and is based on the data collected in a global organization of more than 400 000 employees in more than 100 countries.**

***Keywords*: Social Network Analysis, Facebook, Network URLs, Machine Learning, Classification**

1. Introduction

Social media sites such as Twitter, Facebook, and YouTube provide great venues for students to share their experiences, vent emotion and stress, and seek social support [1]. On various social media sites, students discuss and share their everyday encounters in an informal and casual manner. Student's digital footprints provide vast amount of implicit knowledge and a whole new perspective for educational researchers and practitioners to understand student's experiences outside the controlled classroom environment. This understanding can inform institutional decision making on interventions for at risk students, improvement of education quality, and thus enhance student recruitment, retention, and success. The abundance of social media data provides opportunities to understand student's experiences, but also raises methodological difficulties in making sense of social media data for educational purposes.

On various social media sites, students discuss and share their everyday encounters in an informal and casual manner. Student's digital footprints provide vast amount of implicit knowledge and a whole new perspective for educational researchers and practitioners to understand student's experiences outside the controlled classroom environment. This understanding can inform institutional decision-making on interventions for at-risk students, improvement of education quality, and thus enhance student recruitment, retention, and success [1]. The abundance of social media data provides opportunities to understand student's experiences, but also raises methodological difficulties in making sense of social media data for educational purposes. Just imagine the sheer data volumes, the diversity of Internet slang, the unpredictability of location and timing of students posting on the web, as well as the complexity of student's experiences. Pure manual analysis cannot deal with the ever-growing scale of data, while pure automatic algorithms usually cannot capture in-depth meaning within the data [2].

Most of these user generated content is textual information. The rapid growth in volume of web texts from major social network sites like Facebook and Twitter drives us to analyze and mine the data through computational techniques. Identifying their sentiments has become an important issue and attracted many attentions. Recently, there have been a number of studies attempting to model/predict real-world events using information from social media networks. Among these, Facebook has attracted additional attention because of the huge surge in its popularity. Perform a large-scale analysis of brand sentiments on Facebook.

Just imagine the share data volumes, the diversity of Internet slang, the unpredictability of location and timing of students posting on the web, as well as the complexity of student's experiences. Pure manual analysis cannot deal with the ever-growing scale of data, while pure automatic algorithms usually cannot capture in-depth meaning within the data Traditionally, educational researchers have been using classroom activities to collect data related to student's learning experiences These methods are usually very time consuming, thus cannot be duplicated or repeated with high frequency [2]. The scale of such studies is also usually limited. In addition, when prompted about their experiences, students need to reflect on what they were thinking and doing sometime in the past, which may have become obscured over time [1].

The emerging fields of learning analytic and Educational Data Mining (EDM) have focused on analyzing structured data obtained from Course Management Systems (CMS), classroom technology usage, or controlled online learning environments to inform educational decision making. However, to the best of our knowledge, there is no research found to directly mine and analyze student posted content from uncontrolled spaces on the social web with the clear goal of understanding student's learning experiences [3].

The Research of goals of this study are: To demonstrate a work flow of social media data sense-making for educational purposes, integrating both qualitative analysis and large scale data mining techniques as illustrated. To explore engineering student's informal conversations on Facebook, in order to understand issues and problems students encounter in their learning experiences. Chose to focus on engineering student's posts on Facebook out problems.

2. Related work

Most of the Internet users are using the SNS's for online interactions. Normally, the users share opinions, facts or issues based on their topic of interest without being at the same place and same time. There are a lot of tools for opinion mining and sentiment analysis such as analysis of customer's product reviews, personality of SNS's users and educational purpose. Sentiments are analyzed after all the opinions in comments or postings are extracted.

By analyzing people's sentiment, the emotions of the public toward a particular issue can be observed, experimented and quantified. The complexities in conveyed texts cause an insufficiency to abide in the existing sentiment analysis studies which identify user behaviors as well as their state of minds. Based on previous research, various methods are implemented in sentiment analysis which is done either manually, semi-automatically or fully automatically. In addition, sentiment analysis consists of several processes which include extraction and classification.

A. Extraction and classification

There are various methods used in previous research for data extraction where methods in sentiment analysis have transformed moderately to automatic computer-based system. Manual approach time-consuming in conducting the surveys for data collection. Unlike automated systems, there are no/less limitations in collecting and analyzing data [7]. Some of the automated tools perform filtration immediately after the contents are extracted. For example, a research done on Facebook extracts relevant and eliminates irrelevant contents from posts by using information retrieval techniques then filters the data after the extraction. The filtration includes emoticons replacements, upper and lower casing, and removal of stop words, repeated words as well as punctuations [2], [9]. In order to classify the texts into emotions, various classifiers or methods are used after performing the filtration of data [6]. In, sentiment analysis is done by using Sentiment Identification Algorithm which are Compositional Semantic Rule, Numeric Sentiment Identification, and Bag-of-Word and Rule-based. All these algorithms are used in Machine Learning Model which involves several classifiers such as Decision Tree (J48), Random Forest, Logistic Regression and Neural Network.

There are also other classifiers being used by other researchers in their works such as Naïve Bayes, Support Vector Machine (SVM), Sequential Mining Optimization (SMO) and Maximum Entropy. Even though the approach is used to increase the accuracy of sentiment classification, the manual Lexicon based approach is more efficient for data analysis than the automated classification methods because it is flexible, especially for multilingual texts. Mostly, automated classification methods are used for English texts and as for the other texts such as Malay, Thai, Chinese, German and Spanish used either manual or semi-automated approach for classifications.

3. System architecture

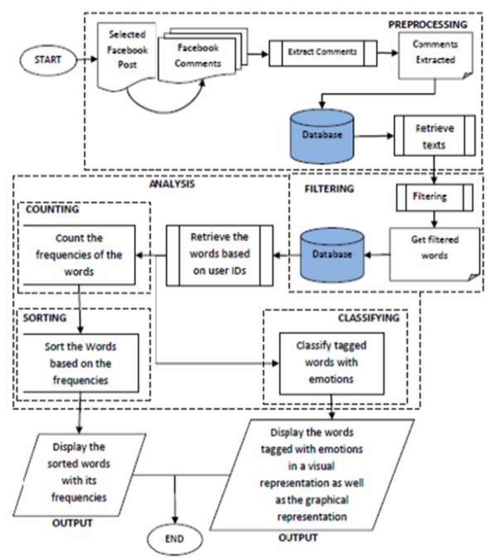

Fig. 1. System architecture

To classify and filter the emotions of the comments posted on the facebook following procedure is followed, that is

- First of all, select the facebook post on which we have to comment further.
- Then attach the comments as per our requirements. This steps come under the preprocessing phase of the whole process. Later, extract the comments thoroughly and as the comments get extracted store them in databases.
- Then retrieve the texts the databases in filtering stage. As the words are filtered store them in database.
- Again retrieve them from the database, but this time on the basis of their user ID's and count the

frequencies of the words.

- Now it's time to sort the words on the basis of their respective frequencies. As the sorting phase completes the words are either displayed in sorted way with their frequencies as the output or,
- They are classified again with emotions and then this tagged words with emotions are displayed in a visual representation as well as the graphical representation as output.
- Here, the words are classified in two ways SVM (Support Vector Machine) and Naïve Bayes classifiers.
- At last the process ends and the comments are classified.

4. Software environment

Java technology is both a programming language and a platform. The Java programming language is a high-level language that can be characterized by all of the following buzzwords:

- Simple
- Architecture neutral
- Object oriented
- Portable
- Distributed
- High performance
- Interpreted
- Multi-threaded
- Robust
- Dynamic
- Secure

With most programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is unusual in that a program is both compiled and interpreted. With the compiler, first you translate a program into an intermediate language called Java byte codes the platform independent codes interpreted by the interpreter on the Java platform. The interpreter parses and runs each Java byte code instruction on the computer. Compilation happens just once; interpretation occurs each time the program is executed.

Virtual Machine (Java VM). Every Java interpreter, whether it's a development tool or a Web browser that can run applets, is an implementation of the Java VM. Java byte codes help make "write once, run anywhere" possible. You can compile your program into byte codes on any platform that has a Java compiler. The byte codes can then be run on any implementation of the Java VM. That means that as long as a computer has a Java VM, the same program written in the Java programming language can run on Windows 2000, a Solaris workstation, or on an iMac.

A. The java platform

A platform is the hardware or software environment in which a program runs. Already mentioned some of the most popular platforms like Windows 2000, Linux, Solaris, and Mac OS. Most platforms can be described as a combination of the operating system and hardware. The Java platform differs from most other platforms in that it's a software-only platform that runs on top of other hardware-based platforms.

The Java platform has two components:

- The Java Virtual Machine (Java VM)
- The Java Application Programming Interface (Java API)

Introduced to the Java VM. It's the base for the Java platform and is ported onto various hardware-based platforms. The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The Java API is grouped into libraries of related classes and interfaces, these libraries are known as packages.

5. Naive Bayes classifier

Naive Bayes classifiers are a collection of classification algorithms based on Baye's Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. It is a classification technique based on Bayes Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods. In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

6. Support Vector Machine (SVM) classifier

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper plane. "Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate.

7. Applications

- Can be used for analysis and visualization of any kind if social networking portal.
- Social Network Visualization.
- With some modifications it can be used for other domains like, ecommerce or any other similar kind of portal.

8. Conclusion

Discussed various challenges and solutions for conducting SNA (Social Network Analysis) in enterprise. Multi-modality aspects of people relationships, including social aspect, financial aspect, and human property aspect. System challenges such as large-scale graph mining and large-scale network visualization. Focused on the fundamental research and system issues.

9. Future scope

It would be interesting to test current system on different algorithm for results. Large size dataset can be tested. Real time dataset can be tested. In the future, the context information in a document is important. And conditional random fields are a good model to capture it. Besides these, building a comprehensive and domain specific sentiment words will help sentiment identification. In addition, the major and minor entities in a comparative sentence will influence sentiment decision. Capture contextual information. Include active learning to improve learning.

References

[1] Smallblue Network Visualization Demo Videos. [Online]. Available: http://smallblue. research.ibm.com/demos/
[2] Visual Exploration on Mapping Complex Networks. [Online]. Available: http://www. visualcomplexity.com/vc/
[3] E. Airoldi, D. Blei, S. Fienberg, and E. Xing, BMixed membership stochastic blockmodels, J. Mach. Learn. Res., vol. 9, pp. 1981–2014, 2008.
[4] L. Akoglu, M. McGlohon, and C. Faloutsos, BOddball: Spotting anomalies in weighted graphs, in Proc. Pacific-Asia Conf. Knowl. Disc Data Mining, 2010, pp. 410–421.
[5] I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, k-core decompositions: A tool for the visualization of large scale networks. [Online]. Available: http://arxiv.org/abs/cs. NI/0504107.
[6] R. Burt, Structural Holes: The Social Structure of Competition. Cambridge, MA: Harvard Univ. Press, 1992.
[7] R. Burt, B Structural holes and good ideas, Amer. J. Sociol., vol. 110, no. 2, pp. 349–399, 2004.
[8] W. Cui, H. Zhou, H. Qu, P. C. Wong, and X. Li, BGeometry-based edge clustering for graph visualization, IEEE Trans. Inf. Visual. Comput. Graph., vol. 14, no. 6, pp. 1277–1284, Nov. 2008.
[9] R. I. M. Dunbar, B Neocortex size as a constraint on group size in primates, J. Human Evol., vol. 22, no. 6, pp. 469–493, 1992.
[10] K. Ehrlich, C.-Y. Lin, and V. Griffiths-Fisher, B Searching for experts in the enterprise: Combining text and social network analysis, in Proc. ACM Conf. Supporting Group Work, 2007, pp. 11
[11] L. Fleischer, M. Goemans, V. Mirrokni, and M. Sviridenko, B. Tight approximation algorithms for maximum general assignment problems, in Proc. ACM/SIAM Symp. Discrete Algorithm, 2006, pp. 611–620.
[12] J. F. Gantz, C. Chute, A. Manfrediz, S. Minton, D. Reinsel, D. Schlichting, and A. Toncheva, BThe diverse and exploding digital universe, White Paper. International Data Corporation, Framingham, MA, Mar. 2008.
[Online]. Available: www.emc.com/collateral/analystreports/ diverse-exploding-digitaluniverse.pdf.
[13] M. Ghoniem, J.-D. Fekete, and P. Castagliola, BA comparison of the readability of graphs using node-link and matrix-based representations, in Proc. InfoVis, 2004, pp. 17–24.
[14] B. Golub and M. O. Jackson, B, "Using selection bias to explain the observed structure of Internet diffusions," Proc. Nat. Acad. Sci., vol. 107, no. 24, p. 10833, 2010.
[15] P. O. Hoyer, BNon-negative matrix factorization with sparseness constraints, J. Mach. Learn. Res., vol. 5, pp. 1457–1469, 2004.
[16] Information Commissioner's Office (ICO), Improving user interest inference from social neighbors, U.K., 2012.
[17] U. Kang, H. Tong, J. Sun, C.-Y. Lin, and C. Faloutsos, BGBASE: A scalable and general graph management system, in Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining, 2011, pp. 1091–1099.
[18] U. Kang, C. E. Tsourakakis, A. P. Appel, C. Faloutsos, and J. Leskovec, BRadius plots for mining tera-byte scale graphs: Algorithms, patterns, and observations, in Proc. SIAM Int. Conf. Data Mining, 2010, pp. 548–558.
[19] U. Kang, C. E. Tsourakakis, and C. Faloutsos, Bpegasus: A peta-scale graph mining system implementation and observations, in Proc. Int. Conf. Data Mining, 2009, pp. 229–238.
[20] D. D. Lee and H. Sebastian Seung, B, "Algorithms for non-negative matrix factorization," in NIPS'00 Proceedings of the 13th International Conference on Neural Information Processing System, pp. 535-541, 2000.