

A Survey on Student's Academic Experiences Using Social Media Data

Sushmita Kumari¹, Sonal Rai², Dr. Shiv Kumar³

M.Tech Scholar, Department of CSE, Lakshmi Narain College of Technology & Excellence, Bhopal (M.P), India¹

Assistant Professor, Department of CSE, Lakshmi Narain College of Technology & Excellence, Bhopal (M.P), India²

Professor & Head, Department of CSE, Lakshmi Narain College of Technology & Excellence, Bhopal (M.P), India³

Abstract: Student's social media behaviour reveals about their day-to-day life. Students post their experiences on social networking sites be it personal or academic. Analysis of these posts, however, is not an easy task. Pure manual analysis is not fruitful as data increases at a rapid rate. There is a workflow developed by the survey topic that assimilates the qualitative analysis and data mining techniques. The focus is primarily on engineering students' posts so their problems can be analyzed. This uses the approach of multi-label classification, which is due to building of categories among students' problems-heavy study load, sleep deprivation, lack of social engagement, etc. This enables social media to shed light on students' academic experiences.

Keyword: Education, computers and education, web-text analysis, Social Networking.

1. Introduction

The online social life of people defines the complete picture of their life experiences. There is a lot that can be analyzed and explored beneath this social media data of the people. Student's learning experiences can be studied from their online posts. But, the problem arises when these posts are counted on and on and it becomes a tedious job to study them. Hence, pure manual analysis cannot alone result out into productivity.

It needs use of data mining techniques for the analysis. Students' online conversations reveal aspects of their experiences that are not easily observed in formal classroom settings. Traversal of engineering students' informal conversations to trace their root problems hence is essential[1].

1.1 Overview of Data Mining and Social Media

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to availability of large amount of data and imminent need for turning such data into useful information and knowledge. Data mining is the process of digging through data and looking meaningful trends and patterns. The information and knowledge gained can be

used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration Data mining can be viewed as a result of the natural evolution of information technology. Data mining is iterative process.

Data cleaning: It is a process of removing noise and inconsistent data.

Data integration: In this step data from multiple sources are combined.

Data selection: In this step data relevant for mining task is selected.

Data transformation: In this step data will be transformed into form that is appropriate for mining.

Data mining: In this step some intelligent methods are applied for extracting data patterns.

Pattern evaluation: In this step truly interesting patterns representing knowledge based on some interestingness measure are identified.

Knowledge presentation: In this step visualization and knowledge representation techniques are used to present the mined knowledge to the user.

A social network can be defined as a set of people, organization or other social entities connected by set of social relationship such as friendship, co-working or information exchange. Social network analysis focuses on the analysis of the pattern of relationships among people, organization, states and such social entities. In this paper a survey of work done in the field of social network analysis is done and this paper also concentrates on the future trends in research on social network analysis.[9]



Figure 1: Social Media[2]

1.2 Social Media Mining Purpose

Student's informal conversations on social media (e.g. Twitter, Facebook) shed light into their educational experiences opinions, feelings, and concerns about the learning process. Data from such un-instrumented environments can provide valuable knowledge to inform student learning. Analysing such data, however, can be challenging. The complexity of students' experiences reflected from social media content requires human interpretation. However, the growing scale of data demands automatic data analysis techniques. In this paper, developed a workflow to integrate both qualitative analysis and large-scale data mining techniques To demonstrate a workflow of social media data sense-making for educational purposes, integrating both qualitative analysis and large-scale data mining techniques To explore engineering students' informal conversations on Twitter, in order to understand issues and problems students encounter in their learning experiences. There are so many social media sites available in market such as (Face book, twitter).there are 1 million active user on Face book and 506 million user on twitter .daily user send 500 tweets and upload 55 million photos. So, it is necessary to mining this large amount of data to avoid complexity.

There are many algorithms available in data mining such as:

1. Naive Bayes
2. Maximum entropy
3. Support vector machine
4. Logistic regression

2. Literature Survey

The study mainly targets the educational experiences of engineering students as engineering is the building block of our technology and future. The application of leant science is directly proportional to academic growth of the engineers. Below discuss the contributions of various researchers.

From Goffman in [1] has a theory called as the Goffman's theory of social performance, which relates to the students as they are liberal and open to a platform on which they are comfortable unlike surveys conducted in closed classrooms, group discussions, forums conducted on the topic of debate. Their opinions are not transparent. Hence it can be concluded that many social media websites available for people to stay connected and share their experiences is a platform where students find themselves comfortable to share their experiences.

From support.twitter.com. [2] it can profoundly be said that for analyzing students' posts considering twitter is a good option. Twitter uses the concept of hash-tag. The symbol # is a hash-tag that consists of all the related keywords and related content of the topic specified in the hash-tag.Tweet classification is the prime reason for the choice of twitter for the study.

From D.Davidov,O.Tsur,and A.Rappoport [3] a framework is proposed for supervised sentiment classification based on twitter data. It avails 15 smileys and 50 hash-tags as training labels for various sentiments. It cuts the requirement of manual intervention into labeling. It allows discerning and classifying multiple types of sentiments of short text. It also assesses multiple types of features like punctuation, patterns, words and n-gram for drawing out sentiments. There are two more methods that provide for automated recognition of sentiment types that overlap and deal within their interdependencies.

From A. Go, R. Bhayani, and L. Huang [4] the study proposed a method to bring out the sentiment into positive and negative, this also aids in cutting down the requirement of manual efforts for the same job. The primary idea is to use tweets along with emoticons for distant supervision learning. This method works better when used in combination with machine learning algorithm. They give high accuracy of the result for sentiment classified when using this method. The classification algorithm takes into account the number of classes .in the given document. Classification can result investigation on such a gathered sample taken from tweets which are related to engineering students' belongs to college life. The proposed system finds engineering students bump into problems are as follows: lack of social engagement, heavy study load, and sleep dispossession. Considering the outcomes to be brought the proposed system will implement of multi labeled classifiers algorithm to identify and organize tweets which will be reflecting students' problems.

3. Problem Identification

A lot of research work has been done in the field. This study explores the previously instrumented space on Twitter in order to understand engineering student's experiences, integrating both qualitative methods and

large-scale data mining techniques. In our study, through a qualitative content analysis, found that engineering students are largely struggling with the heavy study load, and are not able to manage it successfully. Heavy study load leads to many consequences including lack of social engagement, sleep problems, and other psychological and physical health problems. Many students feel engineering is boring and hard, which leads to lack of motivation to study and negative emotions.

Diversity issues also reveal culture conflicts and culture stereotypes existing among engineering students. Building on top of the qualitative insights, implemented and evaluated a multi-label classifier to detect engineering student problems from Purdue University. This detector can be applied as a monitoring mechanism to identify at-risk students at a specific university in the long run without repeating the manual work frequently. Our work is only the first step towards revealing actionable insights from student-generated content on social media in order to improve education quality.

4. Block Diagram

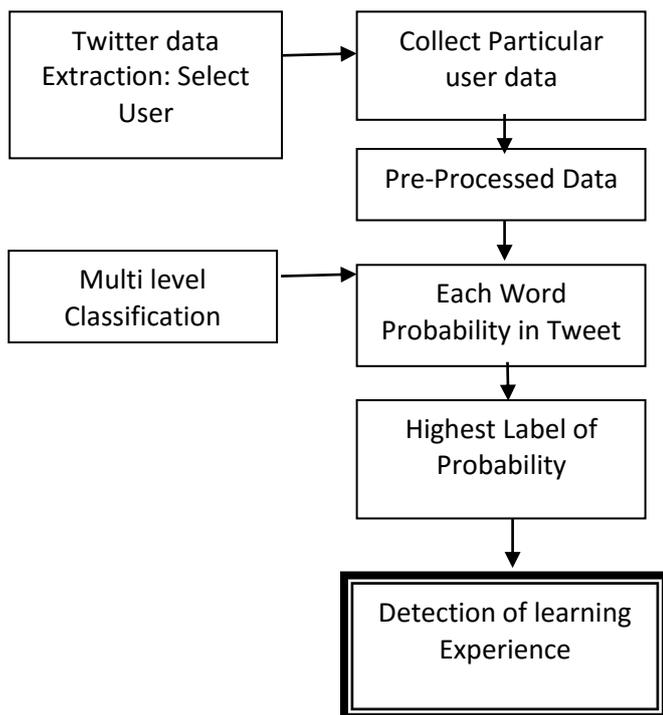


Figure 2: Flow chart of proposed work

In the above Block Diagram system that used for the supervised learning technique for the organization of classified the twitter tweets and such a Foremost it has to be employed a healthy known text arrangement technique called Naive Bayes (N.B.)

5. Conclusion

Looking at the various contributions made by researchers it can be concluded that analyzing the students' problems from their online social image gives transparent result. It gives a classification of their problems into specific labels that helps for easy recognition of the problems and also the severity of any particular problem. There is a no. of tweets that fall into more than one category hence to classify them in a single category Naïve Bayes multi-label classifier is used that gives a high accuracy result. Many a times some tweet may get unidentified and get classified into the "others" category, this category is a long trail. Hence some steps should be essentially taken for maximum tweets going unidentified. Also it should be ensured that students are comfortable with this approach as their privacy is encroached as their informal conversations are accessed.

Reference

- [1] E. Goff man, *The Presentation of Self in Everyday Life*. Lightning Source Inc, 1959
- [2] <http://support.twitter.com>.
- [3] D. Davidov, O. Tsur, and A. Rappoport, "Enhanced sentiment learning using twitter hashtags and smileys," in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 2010, pp. 241-249
- [4] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N Project Report*, Stanford, pp. 1-12, 2009.
- [5] Xin Chen, Mihaela Vorvoreanu, and Krishna Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences", 2013
- [6] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," *Educause Review*, vol. 46, no. 5, pp. 30-32, 2011.
- [7] M. Rost, L. Barkhuus, H. Cramer, and B. Brown, "Representation and communication: challenges in interpreting large social media datasets," in *Proceedings of the 2013 conference on Computer supported cooperative work*, 2013, pp. 357-362.
- [8] M. Clark, S. Sheppard, C. Atman, L. Fleming, R. Miller, R. Stevens, R. Streveler, and K. Smith, "Academic pathways study: Processes and realities," in *Proceedings of the American Society for Engineering Education Annual Conference and Exposition 2008*.
- [9] *Data Mining Techniques*, Arjun K Pujari, Edition 3.