

A Survey On Analyzing Students' Academic Experiences Using Social Media Data

Priya Lande¹, Vipul Dalal²

¹Vidyalankar Institute of Technology, University of Mumbai, Department of Computer Engineering, Mumbai, India

²Vidyalankar Institute of Technology, University of Mumbai, Department of Computer Engineering, Mumbai, India

Abstract: *Students' social media behavior 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.*

Keywords: Education, web text analysis, social-networking, students

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. Students' 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.

2. Historic Development

The study mainly targets the educational experiences of engineering students as engineering is the building block of our technology and future. The application of lean science is directly proportional to academic growth of the engineers. Below we discuss the contributions of various researchers in this field given in the literature, chronologically.

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 [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. In [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.

In [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 into:

- Single-label classification
- Multi-label classification

Single label classification is when each data item can be labeled for a single class only. Multi-label classification is when each data item can be labeled for more than two classes. Multi-label classification has been attracting many studies therefore, in [4] Naïve-Bayes multi-label classifier is explored as it is one of the most popular multi-label classifier. Implementation of this classifier is done by breaking down the multi-label classification problem into multiple single-label classification problems. This can be aggregated into a method called as one versus all or binary relevance. It basically works over the assumption of independent categories and providing training to binary classifier for each label set.

In [5] the hash-tag #engpproblems was most talked about hash-tag for engineering students. Tweets and posts were collected

from this hash-tag and inductive content analysis was performed over the collected data. This led to formation of categories: Heavy-study workload, Lack of social engagement, Sleep-deprivation, Negative emotion, Diversity issues and others. Inductive content analysis is where the manual text analysis is performed. It is one of the most popular methods for qualitative purpose. It was observed that a number of tweets were members of more than one category for example- the tweet “Too many assignments, no time for parties” can be classified into Lack of social engagement category as well as Heavy-study workload category. This leads to multi-label classification and such types of posts and tweets have to be taken special care of.

3. 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.

References

- [1] E. Goffman, *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