Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 6, June 2016, pg.76 – 82

A Survey on Analyzing Social Media Data for Understanding Students' Learning Experiences

Anurag D. Kshirsagar
Research Student
Department of IT
Sinhgad College of Engineering, Pune
Maharashtra, India
anurag.kshirsagar@yahoo.com

Dr. Kalpana S. Thakre
Associate Professor
Department of IT
Sinhgad College of Engineering, Pune
Maharashtra, India
ksthakare.scoe@sinhgad.edu

ABSTRACT

Now a days, Social media has become a very popular medium for communicating, expressing our feelings and sharing information with our friends. As a result, Students' social media behavior reveals opinion about their day-to-day life. Social media lets the creation and interactions of user created content. Facebook, Twitter etc. are very popular social media sites. Students informally discuss on social media about their educational mindset, experience and worry about the learning procedure. We are focusing on students such educational experiences i. e. their opinion and concerns about the learning process. To report student problem, the information from such uninstrumented environments presents valuable data. Examination of such data from social media is not an easy task. Solving a problem of students educational experiences reflected from social media requires human interaction. Researchers were collected the engineering student's Twitter posts to know the problem and troubles in their educational practices. There is a workflow which provides both qualitative investigation and large-scale data mining scheme. A qualitative analysis was conducted on an engineering student sample by collecting his tweets. It reveals that engineering students experience problems like lack of social meeting, sleep deficiency, and heavy learning load. Considering this outcome, a multi-label classification algorithm can be use to categorize the tweets presenting the student's problems. This paper provides a survey on traditional ways to analyze social media data for understanding students' learning experiences.

Keywords— Education, computers and education, social networking, web text analysis.

I. INTRODUCTION

The increase in the enhancement of internet enabled devices such as personal computers and mobile devices enables the social networking sites to become very popular in recent years. A social network is a collection of nodes and edges which has interactions/relationships in among themselves. The nodes are the actors, and the edges has the relationships or interactions between these actors. Students share their views, joy, emotions, stress and seek social support using sites such as Facebook, Twitter, and YouTube. There are various social media sites using which students discuss and share their everyday opinions in an informal and casual manner. Digital views by students reveals a vast amount of implied knowledge. Educational researchers and practitioners get a novel perspective to understand students' problems outside the controlled classroom environment[1]. By enabling this, there can be
improvement in institutional decision-making, improvement of educational quality, and also enhancing student factors such as retention, success and recruitment.

The social media data on large scale provides opportunities to understand students' experiences but also creates methodological challenges to use social media data for educational purposes. To collect data regarding students learning experiences, educational researchers have been using traditional methods such as interviews, focus groups, classroom activities and surveys. Usually, these methods are very time consuming, therefore cannot be repeated or duplicated with a high frequency. When prompting about their experiences, students need to consider what they were thinking and doing sometime in the past, which may have become abstruse over time.

The researchers are trying to integrate both qualitative analysis and large-scale data mining techniques to show a process flow of social media data sense-making for educational purposes. Also the researchers goal is to explore engineering students informal conversations on Twitter, to understand the issues and problems students face in their learning experiences.

Data mining is the process of extracting useful and novel information (which is interesting and interpretable) from data. Since, many years data mining is used widely, to examine large volumes of data such as census data, supermarket scanner data and airline passenger records. Application areas of data mining are Scientific areas, Government sectors and Business areas. Data mining in each application area serves a different purpose. For example, the business objective of data mining is to raise a profit (tangible) which is measured in terms of accounts of money, customer loyalty and number of customers [2].
The field of educational data mining and learning analytics has focused on analyzing structured data obtained from classroom technology, course management systems, usage to inform educational decision making. With the clear objective of getting students learning experiences, there is no research found directly on mining and analysis of content posted by students using uncontrolled spaces on the social web.

Rest of the sections are arranged as follows: the survey of literature is provided in Section II and Section III concludes this survey paper.

II. LITERATURE SURVEY

This literature survey is done by considering to different kind of database and provides mining of Social Media Data for Understanding Students' Learning Experiences.

1) **Mining Student Learning Experiences**

2) **Social Media Database**

1) Mining Student Learning Experiences

a. eMUSE

Elvira Popescuet et al [7] has proposed a system using eMUSE. eMUSE uses local database to store and retrieve students' actions with each tool. Also, it provides a summary of each student's activity such as evolution over time, peers comparison, graphical visualization, as well as complete data from a database. Following the instructor-defined criteria eMUSE calculate a score based on the recorded student activity. Also, it provide basic administrative services such as authentication, enroll students to the course, edit profile etc. The creation of eMuse involves the step of selecting the suitable web 2.0 tools to be integrated into the system for meeting the essential requirements. Mashups were used to integrate the web 2.0 tools into the platform. For creating new web application a mashup reveals a combination of data and/or functionalities from two or more external resources.

b. Visualization Technique

Joris Klerkx, Katrien Verbert and Erik Duval [3] introduced visualization technique. Learning can be enhanced using visualization technique. Information visualization is a powerful way of making sense of the data that has emerged from various research areas such as, computer science, psychology, human computer interaction etc. The application areas of information visualization field are digital libraries, financial data analysis, data mining, scientific research etc. Using visualization techniques in learning is usual. Maps and drawings also will also use them for thousands of years. This requires high analysis on how more novel visualization techniques can be useful for enhancing various activities during the learning process: discovering and understanding educational resources, evaluate learners' progress, cooperation between learners and teachers, and building learning experience. Mathematics is the other learning domain where also visualizations play an important role for enabling students to see the unseen in data. It is difficult to find any mathematics textbook that explains mathematical like the Pythagorus theorem without using visualization techniques. In 1980, Presmeg provided a research review based on visualization in learning and teaching mathematics. Computer Supported Collaborative Learning (CSCL) reveals that, learning is not only subject of accepting fixed things, but it is the ongoing, dynamic, evolving result of complex interactions basically taking place within communities of people. Therefore, visualization of a social network is very useful to make people aware of their social context and to enable them to explore context.

c. Educational data mining and Learning Analytics

In "Mining Social Media Data for Understanding Students' Learning Experiences" Xin Chen, Mihaela Vorvoreanu, and Krishna Madhavan [1] proposed a system using Educational data mining and Learning Analytics are the emerging data driven educational approaches. To understand students and their learning environments in order to inform institutional decision-making, these approaches analyze data generated in educational settings. Broadly, there are four types of relationship mining: sequential pattern mining, causal data mining, association rule mining and correlation mining [1]. The methods useful in understanding student learning in various collaborative settings are EDM and LA. Collaborative learning behaviors have been analyzed to determine the behaviors having characteristic of more successful groups, more successful learners, in multiple surroundings, also includes computer-mediated discussions, online collaboration using software development tools, and interactive table top collaboration.

d. Human-centered

Julie S Hui, Elizabeth M Gerber and Steven P Dow [4], explained Human-centered design involves interaction of users. Interacting with users within industry can be challenging and fostering these interactions in a classroom setting is not so easy. This explores the use of crowd-based design activities as a means to support student-user interactions online. Industry and

government focus on new ways to prepare students for careers in innovation. Hence, there is a growing demand for human-centered design instruction. Instructors teach students the weightage of authentic user interactions as users can provide a good understanding of real-world needs, help generate advantageous and creative solutions, and provide useful feedback. Generally, designers use in-person research methods such as interviews and contextual inquiry to interact with users. But the challenges such as locating users and arranging meetings, can bound the opportunities for such interactions. The above methods provide a rich understanding of users, these methods takes weeks or months to execute. Designers in industry and academia explores the value of soliciting design feedback and ideas online, like testing first impressions of web-pages through an online usability tool and using crowd funding platforms[4].

e. Text based mood classification

Thin Nguyen et al [5] introduced text based mood classification is one another method in this context. Mood is a means of conveying a state of the mind such as being happy, sad or angry. It is a strong form of sentiment expression. Social media texts are rich in sentiment. In the blogosphere, mood classification can filter search results and provides detailed patterns of how bloggers behave and relate to one another and to ascertain the mental health of communities. However, text-based mood analysis raises additional challenges across standard text classification and clustering. It is expected to construct a feature set for blogosphere that works with no supervised feature training to classify mood, because the blogosphere is vast and is continuing to grow.

f. Twitter data mining

Abdullah Gok et al [9] introduce Twitter data mining, which is very helpful for analyzing and generating specific knowledge. To generate a specific knowledge with respect to researcher's subjective domain, they tend to analyze Twitter contents. Twitter data analysis methods usually involves qualitative content analysis, network analysis linguistic analysis and some simplistic methods like word clouds and histograms. Based on inductive content analysis classification model is applied and validated on dataset. Popular classification algorithms are naïve Bayes, maximum entropy, decision tree, boosting and support vector machine, logistic regression. Depending on the number of classes in the classification algorithms, there are binary and multi-class classification approaches. Binary classification involves only two classes, while multiclass classification consists more than two classes. These binary classification and multiclass classification approaches comes under single-label classification systems. It means each data point can only fall into one class where all classes are mutually exclusive [1]. Many of the existing studies based on tweet classification are either binary classification based on relevant and irrelevant content, or multi-class classification based on generic classes. Sentiment analysis is one of the very popular three-class classification technique. Three classes involves positive, negative, or neutral emotions/opinions. Mining customer opinions about products or companies using their reviews or online posts is a sentiment analysis. It has wide exploration in marketing and customer relationship management. Mining sentiment from texts can be done by many methods. To find what student problems a tweet indicates is a more complex task than determining the sentiment of a tweet even for a human judge [6]. Therefore, it needs a qualitative analysis and is not possible to do in a fully unsupervised way. Multilabel classification, however, at the same time, allows each data point to fall into several classes.

Twitter users make use of some special symbols to convey certain meaning. like # is used to indicate a hashtag, RT indicates re-tweet. Sometimes tweeter users repeat letters in words so as to emphasize the words and to show the impact of same word, for example huuuuungryyy, sooo muuchh and monndayyy[1]. Besides, common stopwords such as a, an, and, of, she, it, non letter symbols, and punctuation also bring noise to the text.

The other useful references related to the context of analyzing social media data are [8], [9], [10], [11], [12], [13], [14], [15], [16].

By studying the above mentioned strategies, we are using different and fruitful approach to analyze social media data for understanding students' learning experiences to solve their educational problems.

2) Social Media Database

There are above a billion users of social media network worldwide, many of whom are recurrently active and can be coupled by means of their smartphone's and tablets. Social media truly has become a main communication network in the daily lives of people around the world creating intractable database. Social media now represents the top and biggest source of consumer data which publishes hundreds and thousands of posts about a company's products or services every day.

Social Media Database is mainly used to describe the exponential growth as well as the obtainability of structured and unstructured data, good management of big data can actually lead to great visions which allow relationships to be found in terms of defining business trends, quality of research, link legal citations and regulate concurrent roadway web traffic conditions. And to retrieve the data from social media database various methods are introduced like KDD and Clustering [17].

Social media has blasted as a class of online discourse where people generate content, share, bookmark and network at exceptional rate. Examples include Facebook, MySpace, Digg, Twitter and JISC listservs on the academic side which is stored in social media database. And yet, the content that is produced from these websites remains largely untapped. This social media data can be used to predict real-world outcomes [18] such as, prediction of movies Box Office collection [19], understanding student learning experiences, to understand student behaviour [20].

Title	Author/Year	Pros	Title
Providing collaborative learning support with Social Media in a integrated environment	2014	Store and retrieve students action with each tool.	Providing collaborative learning support with Social Media in a integrated environment
Enhancing Learning with Visualization technique	2014	Used in digital libraries, finan-cial data analysis	Enhancing Learning with Visualization technique
Mining Social Media Data for Understanding Students' Learning	2014	Provide sequential pattern minig, casual data minig, associate rule minig	Mining Social Media Data for Understanding Students' Learning
Crowd based design activities helping students connect with users online	2014	Provide interaction between users	Crowd based design activities helping students connect with users online
Mood sensing from social media tests and its applications	2013	Help sand provide conveying a state of mind	Mood sensing from social media tests and its applications
Automatic identification of locative expressions from social media text: A comparative analysis	2010	Provide twitter content analysis and twitter data classification	Automatic identification of locative expressions from social media text: A comparative analysis

III. CONCLUSION

Inside the eMUSE platform Students have the opportunity to use the pedagogically valuable tools in collaboration with their peers. In this context, eMUSE is quite similar to Personal Learning Environments, like MUPPLE. In data-driven modeling, course development and enhancement is based on data driven analysis of student problems and of the target skill the course is meant to produce; it is not based on trainer self-reflection as found in original instructor centered models. Visualizations also play a vital role in more learning domains like mathematics where they allow students to see the concealed data. EDM and LA methods have similarly been useful in understanding students learning in various collaborative settings. Social media texts are rich in sentiment and this describes various fundamental issues related to mood sensing through these texts and novel applications of this information. Web mining is an important and functional complement to present methods, as well as offering novel insights not easily obtained from other inconspicuous sources.

ACKNOWLEDGEMENT

We would like to thank the publishers and researchers for making their resources available. We thank to our principal and head of the department for providing the required infrastructure and moral support. We are thankful to our guide for providing their timely and valuable guidance for us. Finally we would like to convey our gratitude to friends and family members.

REFERENCES

[1] Xin Chen, Mihaela Vorvoreanu, and Krishna Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences", IEEE Transaction on Learning Technologies, 2014.

[2] Xindong Wu, Xingquan Zhu, Gong-Qing Wu and Wei Ding, "Data Mining with Big Data", Knowledge and Data Engineering, IEEE Transactions, 2014.

[3] Joris Klerkx, Katrien Verbert and Erik Duval, "Enhancing Learning with Visualization techniques", In Handbook of Research on Educational Communications and Technology, Springer, 2014.

[4] Julie S Hui, Elizabeth M Gerber and Steven P Dow, "Crowd-based design activities helping students connect with users online", ACM, 2014.

[5] Thin Nguyen, Dinh Phung, Brett Adams and Svetha Venkatesh, "Mood sensing from social media texts and its applications", Knowledge and Information systems, 2014.

[6] Andrei Yakushev and Sergey Mityagin, "Social networks mining for Analysis and Modeling Drugs Usage", Elsevier, 2014.

[7] Elvira Popescu, "Providing collaborative learning support with Social Media in an integrated environment", World Wide Web,Springer, 2014.

[8] Fei Liu, Maria Vasardani and Timothy Baldwin, "Automatic identification of locative expressions from social media text: A comparative analysis", In Proceedings of the 4th International workshop on Location and the Web, ACM, 2014.

[9] Abdullah Gok, Alec Waterworth and Philip Shapira, "Use of web mining In studying innovation", Scientometrics, 2014.

[10] Yue Gao, Fanglin Wang, Huanbo Luan and Tat-Sang Chau, "Brand data gathering from live social media streams", ACM, 2014.

[11] Carlos Castillo, Mohammed El-Haddad, Jurgen Pfeffer and Matt Stempeck, "Characterizing the life cycle of online news stories using social media reactions", ACM conference on Computer supported cooperative work and social computing, 2014.

[12] William Ribarsky, Derek Xiaoyu Wang, and Wenwen Dou, "Social Media Analytics for competitive Advantage", Elsvier, 2014.

[13] Reihanch Rabbany, Samira Elatia, Mansoureh Takaffoli and Osmar R Zaiane, "Collaborative learning of students in online discussion forums: A social network analysis perspective", In Educational Data Mining, Springer2014.

[14] M.Vorvoreanu, Xin Chen and K.Madhavan, " A Web-based tool for Collaborative social media data analysis", International Conference on Social computing and its Application, 2013.

[15] Marek Opuszko and Johannes Ruhland, "Classification analysis in Complex online social networks using semantic web technologies", In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining, IEEE Computer Society, 2012.

[16] Cristobal Romero and Sebastian Ventura, "Educational Data Mining : A Review of the state of the art", IEEE Transactions, 2010.

[17] Sheela Gole, Bharat Tidke, "A survey of Big Data in social media using data mining techniques"(ICACCS -2015), Jan. 05 – 07, 2015

[18] Sitaram Asur,Bernardo A. Huberman, "Predicting the Future With Social Media" IEEE 2010

[19] Krushikanth R. Apala, Merin Jose, Supreme Motnam, C.-C. Chan, Kathy J. Liszka, "Prediction of Movies Box Office Performance Using Social Media" University of Akron.

[20] Bosnjak Zita ; Bosnjak Sasa, "Students' behavior on social media sites — A data mining approach" 2010

