ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Using Social Data for understanding Student Experience and Decision Crisis Correlation

Thahin Taj A[1], Dr.R.Balu[2]

Research Scholar, Department of Computer Science, Sudharsan College of Arts and Science, Perumanadu-622104, Pudukkottai District, Tamil Nadu, India.

Associate Professor, Department of Computer Science, Sudharsan College of Arts and Science, Perumanadu-622104, Pudukkottai District, Tamil Nadu, India.

ABSTRACT: Students' 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. Analyzing 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, we developed a workflow to integrate both qualitative analysis and large-scale data mining techniques. We focused on engineering students' Twitter posts to understand issues and problems in their educational experiences. We first conducted a qualitative analysis on samples taken from about 25,000 tweets related to engineering students' college life. We found engineering students encounter problems such as heavy study load, lack of social engagement, and sleep deprivation. Based on these results, we implemented a multi-label classification algorithm to classify tweets reflecting students' problems. We then used the algorithm to train a detector of student problems from about 35,000 tweets streamed at the geo-location of Purdue University. This work, for the first time, presents a methodology and results that show how informal social media data can provide insights into students' experiences.

KEYWORDS: Education, computers and education, social networking, web text analysis

I. INTRODUCTION

To learn student experience from social media like twitters using workflow. To integrate 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. Twitter about problems in their educational experiences they are:- Heavy study load, Lack of social engagement, Negative emotion, Sleeping problems. Learning analytics and educational data mining has focused on analyzing structured data obtained from course management systems (CMS). Classroom technology usage or controlled online learning environments to inform educational decision-making. Twitter is a popular social media site. Its content is mostly public and very concise (no more than 140 characters per tweet). Twitter provides free APIs that can be used to stream data. Therefore, I chose to start from analyzing students' posts on twitter. Naive Bayesian Classification algorithm to be used in these concepts. I found Naive Bayes classifier to be very effective on dataset compared with other state-of-the-art multi-label classifiers.

Social media sites such as Twitter, Facebook, and YouTube provide great venues for students to share joy and struggle, vent emotion and stress, and seek social support. On various social media sites, students discuss and share their everyday encounters in an informal and casual manner. Students' digital footprints provide vast amount of implicit knowledge and a whole new perspective for educational researchers and practitioners to understand students' 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 students' 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 slangs, the unpredictability of locations, and timing of students posting on the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

web, as well as the complexity of students' 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 methods such as surveys, interviews, focus groups, classroom activities to collect data related to students' learning experiences. These methods are usually very time consuming, thus cannot be duplicated or repeated with high frequency. 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. The emerging field of learning analytics and educational data mining has 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 students' learning experiences.

Fig.1. The workflow we developed for making sense of social media data integrates qualitative analysis and data mining algorithms. The width of gray arrows represents data volumes – wider indicates more data volume. Black arrows represent data analysis, computation, and results flow. The dashed arrows represent the parts that do not concern the central work of this paper. This workflow can be an iterative cycle.

The research goals of this study are 1) to demonstrate a workflow of social media data sense-making for educational purposes, integrating both qualitative analysis and large-scale data mining techniques as illustrated in Fig. 1; and 2) to explore engineering students' informal conversations on Twitter, in order to understand issues and problems students encounter in their learning experiences. We chose to focus on engineering students' posts on Twitter about problems in their educational experiences mainly because:

1. Engineering schools and departments have long been struggling with student recruitment and retention issues [8]. Engineering graduates constitute a significant part of the nation's future workforce and have a direct impact on the nation's economic growth and global competency.
2. Based on understanding of issues and problems in students' life, policymakers and educators can make more informed decisions on proper interventions and services that can help students overcome barriers in learning.
3. Twitter is a popular social media site. Its content is mostly public and very concise (no more than 140 characters per tweet). Twitter provides free APIs that can be used to stream data. Therefore, we chose to start from analyzing students' posts on Twitter.

PUBLIC DISCOURSE ON THE WEB

The theoretical foundation for the value of informal data on the web can be drawn from Goffman's theory of social performance. Although developed to explain face-toface interactions, Goffman's theory of social performance is widely used to explain mediated interactions on the web today. One of the most fundamental aspects of this theory is the notion of front-stage and back-stage of people's social performances. Compared with the frontstage, the relaxing atmosphere of back-stage usually encourages more spontaneous actions. Whether a social setting is front-stage or back-stage is a relative matter. For students, compared with formal classroom settings, social media is a relative informal and relaxing back-stage. When students post content on social media sites, they usually post what they think and feel at that moment. In this sense, the data collected from online conversation may be more authentic and unfiltered than responses to formal research prompts. These conversations act as a zeitgeist for students' experiences.

Many studies show that social media users may purposefully manage their online identity to "look better" than in real life. Other studies show that there is a lack of awareness about managing online identity among college students, and that young people usually regard social media as their personal space to hang out with peers outside the sight of parents and teachers. Students' online conversations reveal aspects of their experiences that are not easily seen in formal classroom settings, thus are usually not documented in educational literature. The abundance of social media data provides opportunities but also presents methodological difficulties for analyzing large-scale informal textual data. The next section reviews popular methods used for analyzing Twitter data.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

MINING TWITTER DATA

Researchers from diverse fields have analyzed Twitter content to generate specific knowledge for their respective subject domains. For example, Gaffney analyzes tweets with hash tag #iran Election using histograms, user networks, and frequencies of top keywords to quantify online activism. Similar studies have been conducted in other fields including healthcare, marketing, and athletics, just to name a few. Analysis methods used in these studies usually include qualitative content analysis, linguistic analysis, network analysis, and some simplistic methods such as word clouds and histograms. In our study, we built a classification model based on inductive content analysis. This model was then applied and validated on a brand new dataset. Therefore, we emphasize not only the insights gained from one dataset, but also the application of the classification algorithm to other datasets for detecting student problems. The human effort is thus augmented with large-scale data analysis. Below we briefly review studies on Twitter from the fields of data mining, machine learning, and natural language processing. These studies usually have more emphasis on statistical models and algorithms. They cover a wide range of topics including information propagation and diffusion, popularity prediction, event detection, topic discovery, and tweet classification, to name a few.

LEARNING ANALYTICS AND EDUCATIONAL DATA MINING

Learning analytics and educational data mining (EDM) are data-driven approaches emerging in education. These approaches analyze data generated in educational settings to understand students and their learning environments in order to inform institutional decision-making. The present paper extends the scope of these approaches in the following two aspects. First, data analyzed using these approaches typically are structured data including administrative data (e.g., high school GPA and SAT scores), and student activity and performance data from CMS (Course Management Systems) or VLE (Virtual Learning Environments) such as Blackboard (http://www.blackboard.com/). For example, researchers at Purdue University created a system named Signals that mines student performance data from Blackboard course system such as time spent reading course materials, time spent engaging in course discussion forums, and quiz grades. Signals give students red, yellow, or green alerts on their progress in the course taken in order to promote self-awareness in learning. Our study extends the data scope of these data-driven approaches to include informal social media data. Second, most studies in learning analytics and EDM focus on students' academic performance. We extend the understanding of students' experiences to the social and emotional aspects based on their informal online conversations. These are important components of the learning experiences that are much less emphasized and understood compared with academic performance.

INDUCTIVE CONTENT ANALYSIS

Because social media content like tweets contain a large amount of informal language, sarcasm, acronyms, and misspellings, meaning is often ambiguous and subject to human interpretation. Rost et. al argue that in large scale social media data analysis, faulty assumptions are likely to arise if automatic algorithms are used without taking a qualitative look at the data. We concur with this argument, as we found no appropriate unsupervised algorithms could reveal in-depth meanings in our data. For example, LDA (Latent Dirichlet Allocation) is a popular topic modeling algorithm that can detect general topics from very large scale data. LDA has only produced meaningless word groups from our data with a lot of overlapping words across different topics. There were no pre-defined categories of the data, so we needed to explore what students were saying in the tweets. Thus, we first conducted an inductive content analysis on the #engineering Problems dataset. Inductive content analysis is one popular qualitative research method for manually analyzing text content. Three researchers collaborated on the content analysis process.

II. RELATED WORK

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

PENETRATING THE FOG: ANALYTICS IN LEARNING AND EDUCATION - G. SIEMENS AND P. LONG - 2011

Attempts to imagine the future of education often emphasize new technologies ubiquitous computing devices, flexible classroom designs, and innovative visual displays. But the most dramatic factor shaping the future of higher education is something that we can't actually touch or see: big data and analytics. Basing decisions on data and evidence seems stunningly obvious, and indeed, research indicates that data-driven decision-making improves organizational output and productivity.1 for many leaders in higher education, however, experience and "gut instinct" have a stronger pull. Meanwhile, the move toward using data and evidence to make decisions is transforming other fields. Notable is the shift from clinical practice to evidence-based medicine in health care. The former relies on individual physicians basing their treatment decisions on their personal experience with earlier patient cases.

The latter is about carefully designed data collection that builds up evidence on which clinical decisions are based. Medicine is looking even further toward computational modeling by using analytics to answer the simple question "who will get sick?" and then acting on those predictions to assist individuals in making lifestyle or health changes.3Insurance companies also are turning to predictive modeling to determine high-risk customers. Effective data analysis can produce insight into how lifestyle choices and personal health habits affect long-term risks.4 Business and governments too are jumping on the analytics and data-driven decision-making trends, in the form of "business intelligence."

Higher education, a field that gathers an astonishing array of data about its "customers," has traditionally been inefficient in its data use, often operating with substantial delays in analyzing readily evident data and feedback. Evaluating student dropouts on an annual basis leaves gaping holes of delayed action and opportunities for intervention. Organizational processes—such as planning and resource allocation—often fail to utilize large amounts of data on effective learning practices, student profiles, and needed interventions. Something must change. For decades, calls have been made for reform in the efficiency and quality of higher education. Now, with the Internet, mobile technologies, and open education, these calls are gaining a new level of urgency. Compounding this technological and social change, prominent investors and businesspeople are questioning the time and monetary value of higher education.5 Unfortunately, the crescendo of calls for higher education reform lacks a foundation for making decisions on what to do or how to guide change.

REPRESENTATION AND COMMUNICATION: CHALLENGES IN INTERPRETING LARGE SOCIAL MEDIA DATASETS - M. ROST, L. BARKHUUS, H. CRAMER - 2013

Online services provide a range of opportunities for understanding human behavior through the large aggregate data sets that their operation collects. Yet the data sets they collect do not unproblematic ally model or mirror the world events. In this paper we use data from Foursquare, a popular location check-in service, to argue for the importance of analyzing social media as a communicative rather than representational system. Drawing on logs of all Foursquare check-ins over eight weeks we highlight four features of Four square's use: the relationship between attendance and check-ins, event check-ins, commercial incentives to check-in, and lastly humorous check-ins These points show how large data analysis is affected by the end user uses to which social networks are put.

ACADEMIC PATHWAYS STUDY: PROCESSES AND REALITIES - R. STREVELER, AND K. SMITH - 2008

This system describes the evolution and implementation of the Academic Pathways Study (APS), a five year, multi-institution study that addresses questions about the education and persistence of undergraduates in engineering. The APS is the largest element of the Center for the Advancement of Engineering Education (CAEE), funded by NSF for the advancement of engineering learning and teaching. Parts of this paper address questions that engineering education researchers may have about the organizational and technical infrastructure that supported this project, or about its overall implementation (e.g. subject recruitment, data collection methods, and participation rates). Other parts of the paper

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

address questions that researchers and engineering faculty and administrators might have regarding how to explore the findings and insights that are emerging from this extensive longitudinal and cross-sectional study of students' pathways through engineering. In addition, the paper serves as a good starting point for researchers who might be interested in doing additional analysis using APS data. Specific topics include: research team and leadership; research design and methodology; the four study cohorts and their respective contributions; some of the challenges and solutions; and implications for engineering education and future research.

ENABLING ENGINEERING STUDENT SUCCESS: THE FINAL REPORT FOR THE CENTER FOR THE ADVANCEMENT OF ENGINEERING EDUCATION - D. LUND - 2010

Today‘s engineering graduates will solve tomorrow‘s problems in a world that is advancing faster and facing more critical challenges than ever before. This situation creates significant demand for engineering education to evolve in order to effectively prepare a diverse community of engineers for these challenges. Such concerns have led to the publication of visionary reports that help orient the work of those committed to the success of engineering education. Research in engineering education is central to all of these visions. Research on the student experience is fundamental to informing the evolution of engineering education. A broad understanding of the engineering student experience involves thinking about diverse academic pathways, navigation of these pathways, and decision points-how students choose engineering programs, navigate through their programs, and then move on to jobs and careers. Further, looking at students ‘experiences broadly entails not just thinking about their learning (i.e., skill and knowledge development in both technical and professional areas) but also their motivation, their identification with engineering, their confidence, and their choices after graduation.

In actuality, there is not one singular student experience, but rather many experiences. Research on engineering student experiences can look into systematic differences across demographics, disciplines, and campuses; gain insight into the experiences of underrepresented students; and create a rich portrait of how students change from first year through graduation. Such a broad understanding of the engineering student experience can serve as inspiration for designing innovative curricular experiences that support the many and varied pathways that students take on their way to becoming an engineer. However, an understanding of the engineering student experience is clearly not enough to create innovation in engineering education. We need educators who are capable of using the research on the student experience. This involves not oly preparing tomorrow‘s educators with conceptions of teaching that enable innovation but also understanding how today‘s educators make teaching decisions.

We also need to be concerned about creating the capacity to do such research in short, we need more researchers. One promising approach is to work with educators who are interested in engaging in research, supporting them as they negotiate the space between their current activities and their new work in engineering education research. To fully support this process, we must also investigate what is required for educators to engage in such a path.

THE STATE OF LEARNING ANALYTICS IN 2012: A REVIEW AND FUTURE CHALLENGES - R. FERGUSON - 2012

Learning analytics is a significant area of technology‐enhanced learning that has emerged during the last decade. This review of the field begins with an examination of the technological, educational and political factors that have driven the development of analytics in educational settings. It goes on to chart the emergence of learning analytics, including their origins in the 20th century, the development of data-driven analytics, the rise of learning-focused perspectives and the influence of national economic concerns. It next focuses on the relationships between learning analytics, educational data mining and academic analytics. Finally, it sets out the current state of learning analytics research, and identifies a series of future challenges.

MICROBLOGGING IN CLASSROOM: CLASSIFYING STUDENTS' RELEVANT AND IRRELEVANT QUESTIONS IN A MICROBLOGGING SUPPORTED CLASSROOM - S. CETINTAS, L. SI, H. AAGARD - 2011

Microblogging is a popular technology in social networking applications that lets users publish online short text messages (e.g., less than 200 characters) in real time via the web, SMS, instant messaging clients, etc. Microblogging can be an effective tool in the classroom and has lately gained notable interest from the education community. This paper proposes a novel application of text categorization for two types of microblogging questions asked in a classroom, namely relevant (i.e., questions that the teacher wants to address in the class) and irrelevant questions. Empirical results

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

and analysis show that using personalization together with question text leads to better categorization accuracy than using question text alone. It is also beneficial to utilize the correlation between questions and available lecture materials as well as the correlation between questions asked in a lecture. Furthermore, empirical results also show that the elimination of stopwords leads to better correlation estimation between questions and leads to better categorization accuracy. On the other hand, incorporating students' votes on the questions does not improve categorization accuracy, although a similar feature has been shown to be effective in community question answering environments for assessing question quality.

III. Conclusion And Future Work

Our study is beneficial to researchers in learning analytics, educational data mining, and learning technologies. It provides a workflow for analyzing social media data for educational purposes that overcomes the major limitations of both manual qualitative analysis and large scale computational analysis of user-generated textual content. Our study can inform educational administrators, practitioners and other relevant decision makers to gain further understanding of engineering students' college experiences. As an initial attempt to instrument the uncontrolled social media space, we propose many possible directions for future work for researchers who are interested in this area. We hope to see a proliferation of work in this area in the near future. We advocate that great attention needs to be paid to protect students' privacy when trying to provide good education and services to them.

References

[1] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," Educause Review, vol. 46, no. 5, pp. 30- 32, 2011.

[2] 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.

[3] 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.

[4] C. J. Atman, S. D. Sheppard, J. Turns, R. S. Adams, L. Fleming, R. Stevens, R. A. Streveler, K. Smith, R. Miller, L. Leifer, K. Ya suhara, and D. Lund, "Enabling engineering student success: The final report for the Center for the Advancement of Engineering Education," Morgan & Claypool Publishers, Center for the Advancement of Engineering Education, 2010.

[5] R. Ferguson, "The state of learning analytics in 2012: A review and future challenges," Knowledge Media Institute, Technical Report KMI-2012-01, 2012.

[6] R. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," Journal of Educational Data Mining, vol. 1, no. 1, pp. 3–17, 2009.

[7] S. Cetintas, L. Si, H. Aagard, K. Bowen, and M. Cordova- Sanchez, "Microblogging in Classroom: Classifying Students' Relevant and Irrelevant Questions in a Microblogging-Supported Classroom," Learning Technologies, IEEE Transactions on, vol. 4, no. 4, pp. 292–300, 2011.

[8] C. Moller-Wong and A. Eide, "An Engineering Student Retention Study," Journal of Engineering Education, vol. 86, no. 1, pp. 7–15, 1997.

[9] National Academy of Engineering, The engineer of 2020: Visions of engineering in the new century. Washington, D.C.: National Academies Press, 2004.

[10] E. Goffman, The Presentation of Self in Everyday Life. Lightning Source Inc, 1959.

[11] E. Pearson, "All the World Wide Web's a Stage: The performance of identity in online social networks," First Monday, vol. 14, no. 3, pp. 1–7, 2009.

[12] J. M. DiMicco and D. R. Millen, "Identity management: multiple presentations of self in facebook," in Proceedings of the 2007 international ACM conference on Supporting group work, 2007, pp. 383–386.