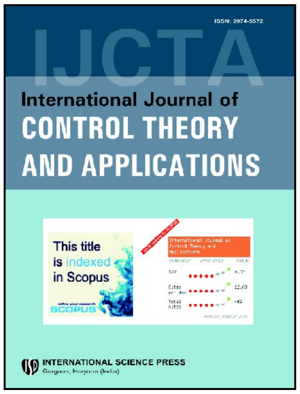

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 29 • 2017

Mining Social Networks for Analyzing Students Learning Experience and their Problems

Blessy Geo. V. M.[1] and S.Prasanna[2]

[1] *Research Schoar,Department of Computer Science, Vels University, Pallavaram, Chennai-600043*
[2] *Associate Professor, Department of MCA, Vels University, Pallavaram, Chennai-600043*

Abstract: On different social media sites, students discuss and contribute to their daily encounters in a casual and informal manner. Analyzing such data, though, can be difficult. The complication of students' experiences reflected from social media content requires human understanding problem of a student's experience exposes from social media sited require human investigation or communication. Examining data from such a social media can be demanding task. In this article, we develop a workflow to combine both qualitative analysis and large-scale data mining techniques. It pays a concentration on engineering student's Twitter posts to recognize the problem and the difficulty in their educational practices. Based on this result, a multi-label classification algorithm that is Naive Bayes Multi-label Classifier algorithm and Memetic classifier is applied to categorize tweets presenting students' problems. Memetic classifier is a population based approach to split the individuals' education for problem search, which has had their own advantages in solving optimization problems.

Keywords: Naïve Bayes Multi-label Classifier, Memetic Classifier, Social media, Statistical Analysis.

1. INTRODUCTION

A data mining study has effectively created several methods, tools, and algorithms for supervision huge amount of data in response to real-world difficulty. As social media are broadly used for a variety of purposes, enormous amounts of user shaped data are near and can be made accessible for data mining. Data mining of social media is able to enlarge ability of accepting innovative experience, the apple of social standard and increase business intelligence to hear good services and enlarges innovative opportunities. The main purpose of the data mining method is to collectively hold large-scale data, take out actionable patterns, and get insightful knowledge. Social media sites such as Twitter, a Face book present grand place for students to share pleasure and struggle, sentiment and strain, and gain communal support. The rising field of learning analytics and learning, data mining is listening carefully on analyzing structured data obtained beginning Course Management Systems (CMS), classroom knowledge usage, or prohibited online learning environments to notify educational resolution making. However, to the greatest of our knowledge, there is no research, establish to directly mine and examine student- posted content from unrestrained spaces on the social web with the obvious goal of understanding students' educational experiences [1].

The research aim of this study is to reveal a workflow of social media data sense-making for learning purposes, integrating equally qualitative analysis and large-scale data mining techniques to discover engineering students' casual conversations on Twitter, in order to understand the issues and troubles students meet in their learning experience [2].

2. LITERATURE REVIEW

Web (data) mining is one of the intelligent computing techniques in the context of Web data management. In general, Web mining is the means of utilizing data mining methods to induce and extract useful information from Web data information [3]. Web mining research has attracted a variety of academics and engineers for database management, information retrieval, artificial intelligence research areas, especially from data mining, knowledge discovery, and machine learning, etc. Basically, Web mining could be classified into three categories based on the mining goals, which determine the part of the Web to be mined: Web content mining, Web structure mining, and Web usage mining [4]. 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. Educational Data Mining (EDM) is the application of Data Mining (DM) techniques to; its objective is to examine this type of information in order to determine educational research problems [5].

3. PROPOSED METHODOLOGY

The research aim of this study is to reveal a workflow of social media data mining for learning purposes and to discover engineering student's informative conversations on Twitter, through which we appreciate the issues and troubles students come upon in their knowledge experiences. First a sample is taken from student and then it conducts qualitative investigation of that sample which is related to the engineering student's educational life. It found engineering students come upon problems such as heavy learning load, lack of social meeting, and sleep absence. Stand on these outcomes, the authors apply a multi-label classification algorithm to classify tweets presenting students' problems. Then Mimetic Algorithm is practical to make, more perfect result, it will filter after that used the algorithm to organize a detector of student problems.

The significant position in proposed study is, First, it intends a workflow to association and incorporate a qualitative investigate the methodology and large scale data mining techniques. It bases our data-mining algorithm on qualitative approaching consequential from human perception, so that it can gain deeper accepting of the data. Then apply the algorithm to another large-scale and unfamiliar dataset, so that the physical method is improved. Second, the paper provides bottomless insights into engineering students' educational experiences as reacting in casual, uncontrolled environments. Many issues and troubles such as study-life balance, lack of sleep, lack of social engagement, and lack of variety clearly materialize. These could bring attentiveness to educational examiner, policy-maker.

4. IMPLEMENTATION

This chapter describes the theoretical theories, and has been conducted for 500 Engineering students across India, who has commented on Face book, Twitter, What sup, etc. Discriminate analysis is used for the purpose of this survey. The survey was conducted to collect primary data through a well structured questionnaire. Secondary data were collected from the bank's records, websites etc. The data have been analyzed with the statistical tools like percentage analysis.

5. VARIABLE CONSIDERATION

The following variables are considered for the analysis.

a) Demographics

1. Gender
2. Age Group
3. Social Media
4. Family Income group
5. Region of India
6. Location
7. Engineering Stream

b) Analysis Results

Naive Bayes Classifier for Discrete Predictors

Table 1
Student segment name and number

Student Segment name	*Segment number*
Restless/Overloaded/Stressed/Fearful	1
Worries about Facility lacking/poor water-food-classroom-labs-library	2
Worries about Syllabus outdated/Curriculum old/Faculty not as per expectation	3
Worries about poor practical/training exposure- No softskills/personality development	4
Worries about internship/industry orientation/job placement	5
Absence of social meeting/alumni connection	6

This table 1 explains the analysis results of the student's segment name and their segment number the various student problems are discussed and the segment number is given.

5.1. A-priori probabilities

Student Segment: The student segment number is given with the A-priori probabilities.

Table 2
Student segment number and probabilities

1	*2*	*3*	*4*	*5*	*6*
0.164	0.172	0.168	0.162	0.174	0.160

Table 3
Conditional probabilities (SOCIAL MEDIA)

Student Segment	*Face book*	*Whatsapp*	*Twitter*	*YouTube*	*Integral*
1	0.232	0.305	0.232	0.110	0.122
2	0.360	0.221	0.163	0.163	0.093
3	0.321	0.202	0.262	0.143	0.071
4	0.284	0.346	0.136	0.086	0.148
5	0.310	0.310	0.195	0.115	0.069
6	0.238	0.313	0.175	0.188	0.088

The above table 3 depicts the student's segment number and the probabilities of students who use social media as face book, Whatsapp, twitter, YouTube and integral.

5.2. Memetic Classifier

Test 1: the Male and female are independent in terms of student classification in the Six Segments

Hypothesis definition:

H0: There is no significance difference between the two genders (Male and Female) in terms of the student classification in the Six Segments.

H1: There is a significance difference between the two genders (Male and Female) in terms of the student classification in the Six Segments.

Variables considered:

- Gender
- Student Segment

Statistical Technique

Chi-Square test of independence of attributes.

Calculation

The observed matrix (O) is as below.

Table 4
Custom Table for Age Group

	Segment of Engineering Students						
Gender	*Restless/ Overloaded/ Stressed/ Fearful*	*Worries about Facility lacking/ poor water-food-classroom-labs-library*	*Worries about Syllabus outdated/ Curriculum old/ Faculty not as per expectation*	*Worries about poor practical/ training exposure-No soft skills/ personality development*	*Worries about internship/ industry orientation/ job placement*	*Absence of social meeting/ alumni connection*	*Total*
Male	55	59	52	54	57	53	330
Female	27	27	32	27	30	27	170
Total	82	86	84	81	87	80	500

The above table 4 depicts the segment of engineering students according to their age gender. The matrix satisfies the minimum count of 5 within each cell.

Degrees of Freedom: The observed values table has 2 rows &6 columns. Hence, the degrees of freedom is (rows-1) x (columns -1) i.e. (2-1) x (6-1) = 5 d.f.

Testing level: 95% significance level (i.e. alpha = 0.05)

Table 5
SPSS Results-Chi-Square Value Calculation for Gender

Pearson Chi-Square Tests		
		Segment of Engineering Students
Gender	Chi-square	0.957
	d.f.	5
	Sig.	0.966

This table 5 interprets the social media usage students according to their gender.

Interpretation: The chi-square value as per Chi-square test is 0.957 (5 degree of freedom, 5% significance level). The significant value p=0.966 is greater than 0.05, the set 5% significance level.

Responses about the learning issues in their engineering study

1. Overloaded with too much of studies/work
2. Inadequate rest/ sleep
3. Stress related to exam fear
4. Poor campus environment [location, nature etc.]
5. No proper water/food/refreshment at college
6. Poor class rooms/labs/library
7. Outdated syllabus/curriculum
8. Faculty is not up to the mark/not high professional
9. Not much practical exposure/training
10. No soft-skills development/personality development aspects
11. Inadequate internship/industry orientation
12. No job guarantees / job placements
13. Absence of social meeting
14. Absence of Alumni connection

Test 2: The age-group distribution is independent of the student classification in the six segments.

Hypothesis definition

H0: There is no significance difference between the age-groups in terms of the student's classification in the Six Segments.

H1: There is a significance difference between the two genders (Male and Female) in terms of the student's classification in the Six Segments.

Variables considered

- Age-groups
- Student Segment

Statistical Technique

Chi Square test of independence of attributes.

Calculation

The observed matrix (O) is as below.

Table 6
Custom table for Age-group

	Segment of Engineering Students						
Age-Group	*Restless/ Overloaded/ Stressed/ Fearful*	*Worries about Facility lacking/ poor water-food-classroom-labs-library*	*Worries about Syllabus outdated/ Curriculum old/Faculty not as per expectation*	*Worries about poor practical/ training exposure-No soft skills/ personality development*	*Worries about internship/ industry orientation/ job placement*	*Absence of social meeting/ alumni connection*	*Total*
17 years	2	1	3	2	0	0	8
18 years	27	21	29	23	27	28	155
19 years	16	22	15	19	14	19	105
20 Years	25	29	27	26	31	28	166
21 years	7	10	7	8	9	3	44
22 years	5	3	3	3	6	2	22
Total	82	86	84	81	87	80	500

This table 6 depicts the age group of the students who are using social media from age 17 to 22 years. The matrix does not satisfy the minimum count of 5 within each cell. Hence, some age –groups are merged, and the revised results are taken.

Degrees of Freedom: The observed values table has 4 rows &6 columns. Hence, the degrees of freedom is (rows-1) x (columns -1) i.e. (4-1) x (6-1) = 3x5=15d.f.

Testing level: 95% significance level (i.e. alpha = 0.05)

Table 7
SPSS Results-Chi-Square Value Calculation for Age Group

Pearson Chi-Square Tests		
		Segment of Engineering Students
Age group	**Chi-square**	10.37
	d.f.	15
	Sig.	0.795

The above table 7 explains the Pearson Chi-Square test of the students according to their age group.

CONCLUSION

Mining social media data is cooperative to researchers in learning analytics, educational data elimination, and learning skill. It gives a way to investigate social intermediate statistics that defeat the main limitations of both corporal qualitative analysis and huge scale computational study of user formed textual contented. From this work done and survey of Interpreting Students Behavior is cooperative to find the drawbacks of the accessible classification algorithm. The newly implemented Memetic Classifier has analyzed the very large amount of data in a short period of time. This newly developed system provides the classification or clustering of student problems and issues and also provides the opinion analysis of the database. As per the survey and work done this system very useful in institutes, organizations, universities. We can enhance in image processing like images, emoticons, videos, etc. This system is also very useful for industry, manufacturing companies, banking sectors, government sectors etc. in future for identification of employees' actions, their behaviors, product feedback, for banking feedback and related to their services.

REFERENCE

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

[2] J.M. DiMicco and D.R. Millen, "Identity Management: Multiple Presentations of Self inFacebook," Proc. The Int'l ACM Conf. SupportingGroup Work, pp. 383-386, 2007.[3] M. Vorvoreanu and Q. Clark, "Managing Identity Across Social Networks," Proc. Poster Session at the ACM Conf. Computer SupportedCooperative Work, 2010.

[3] X. Chen, M. Vorvoreanu, and K. Madhavan, "A Web-Based Tool for Collaborative Social Media Data Analysis," Proc. Third Int'lConf. Social Computing and Its Applications, 2013.

[4] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining Multi-label data,"Data mining and knowledge discovery handbook, pp. 667-685, 2010.

[5] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining Multi-label data,"Data mining and knowledge discovery handbook, pp. 667-685, 2010.