ISSN:0975-9646

Filtering and Classification Of User Based On Social Media Data Using Memetic and Naive Bayes Methods

Gauri Joshi[1], Mr. Samadhan Sonawane [2]

Computer Engineering Department, SavitriBai Phule University of Pune
SITRC, Nasik, India.

***Abstract*— Social media allows the creation and interactions of user created content. Social medium places include Face book, Twitter etc. Student's casual discussion on social media focused into their educational experience, mind-set, and worry about the learning process. Examining data from such a social media can be challenging task. It pays attention on engineering students Twitter posts to know problem and troubles in their educational practices. Based on this outcome, a multi-label classification algorithm that is Naive Bayes Multi-label Classifier algorithm and memetic algorithm is applied to categorize tweets presenting students problems. Memetic algorithm is a population based approach with separate individuals learning for problem search. This study presents a tactic and outcome that demonstrate how casual social media data can present insight into student's incident**

***Keywords*— Social networking, Web-text analysis, Social network analysis, Data mining.**

I. INTRODUCTION

Data mining research has effectively produced several technique, tools and algorithms for managing huge amounts of data to answer real-world troubles. As social media is widely used for various purposes, vast amounts of user created data be present and can be made available for data mining. Data mining of social media can enlarge researcher's ability of understanding innovative experience, to the use of social medium and develop business intelligence to present good services and extend innovative opportunities. Main objectives of the data mining procedure are to collectively handle large-scale data, extract actionable patterns, and gain insightful knowledge. Social media sites such as Twitter, Face book present grand place to students to share happiness and struggle, sentiment and tension, and gain social support. On various social media sites, students talk about their everyday encounters in a comfortable and informal manner. This Students digital information gives huge amount of implicit information and a whole new viewpoint for educational researchers to know students experiences outside the prohibited classroom environment. This understanding can enhance education quality and thus improve student [3],[4].

In Figure 1.1 the width of hollow arrow indicates data volumes. Wider indicates more data volume. Black arrow represents data analysis, computation and result flow. The dashed arrow represents the parts that do not concern the central work. These workflows can be an iterative cycle.

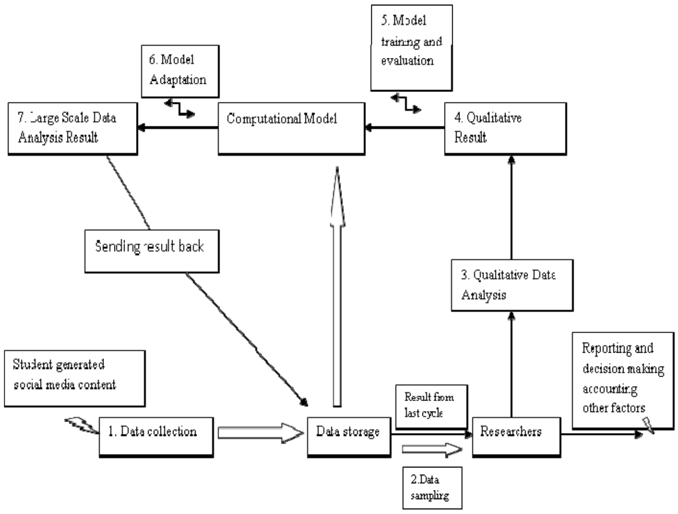

Figure 1.1: Workflow of social media data mining

II. LITRATURE REVIEW

Study overview of Mining social media data using Naïve Bayes algorithm, this article gives the Workflow to collect both qualitative analysis and large-scale data mining analysis and Methodology used Multiple label classification algorithms.

Study overview of Identification of Student's Behaviour in Higher Education from Social Media by using Opinion based Memetic Classifier this article gives to extracting this data from social media sites for identifying the student's behaviour and their opinions Methodology used Opinion based memetic classifier technique

Study overview of Mining Social Media Data for Understanding Students' Learning Experiences Multilable classifier Methodology used Naive Bayes Multi-label Classifier algorithm

study overview of A Web –Based Tool for Collaborative Social Media Data Analysis and findings includes Students informal conversation on social media methodology used for this is High light their educational experiences.

Study overview of Analysing Social Media Data for Understanding Student's Problem for identifying Multi-label classification algorithm

Methodologies are Naive Bayes Multi-label Classifier algorithm.

III. Proposed System

The research goal of this study is to demonstrate a workflow of social media data mining for educational purposes and to explore engineering student's information conversations on Twitter, through which we understand issues and problems students encounter in their learning experiences. Classroom technology usage, or controlled online learning environments to inform educational decision- making [1] [5] [6] [7]. With the help of qualitative investigation and large scale data mining scheme, first a sample is taken from engineering student's educational life. We built a multi-label classifier to classify tweets based on the categories developed in analysis stage. For the task multilable classification, we use Memetic Algorithm. Memetic algorithm required large amount of data for sampling data so this one use for population based classification.

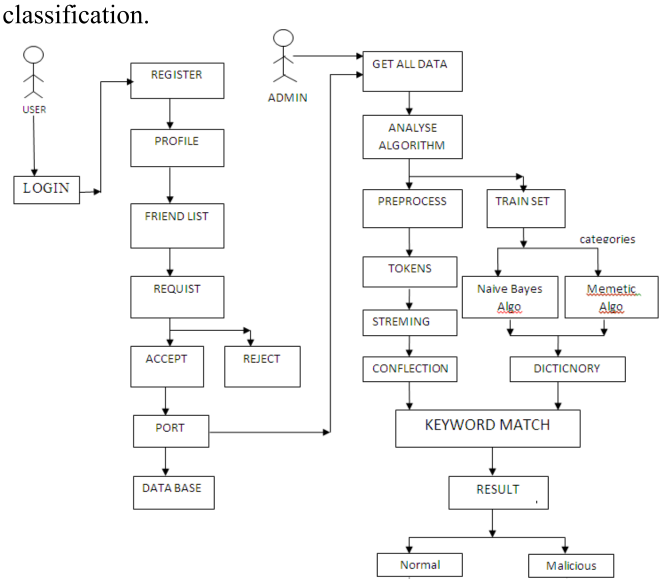

Figure 1.2: Block Diagram of Optimal Investigation of Social Media Data Mining using Memetic Algorithm

As shown in above figure, User will login to the system with the help of user name and password of its own. If user will authenticate then he/ she entered into the system else user will considered as unauthorized user. After that user will registered with system and make profile of self. Users also see the friend list of available data and send request to them. If request accept by any friend then connection make with port and maintain all data as a database. Simultaneously Admin get all data which is stored in the database. As per the data entered by user admin analyze methodology or algorithm for sorting out data into different categories for preprocessing like- Heavy study load, Sleeping problem, diversity issues. With the help of algorithm we will sort out keywords entered by user and with the help of qualitative investigation and large-scale data mining scheme, first sample is taken from engineering student's educational life. We built a multi-label classifier to classify tweets based on the categories developed in analysis stage. For the task multi label classification, we use Memetic Algorithm. Memetic algorithm required large amount of data for sampling data so this one use for population based classification.

Architecture

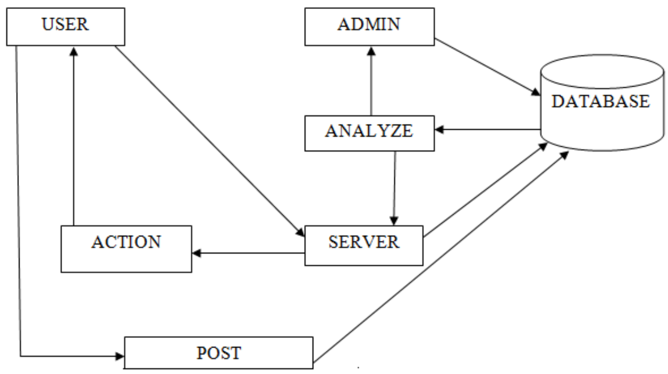

Figure 1.3: Architecture of Social Media Data Mining using Memetic Algorithm

IV. Methodology

A. *Naive Bayes Multi-lable Classifier*

We built a multi-lable classifier to classify tweets based on the categories developed in previous content analysis stage. For the task multi lable classification, Nave Bayes Classifier should be adapted to multi lable data. For multi-label classification, the situation is slightly more complicated, than the other classifier because each document gets assigned multiple labels.

Algorithm

1. Start
2. Select data
3. Create user profile
4. Stored all data into dataset
5. If data of type structured then .....if no goto step 10
6. Analyze algorithm
7. Select trainset
8. Select Naive Bayes Multi-lable Classifier
9. Apply formula

$$p(c'|\, d_i) = \frac{p(\, d_i|c').p(c')\ \alpha\ \Pi^{k}{}_{k\text{-}1}\ p(\, w_{ik}\ |\ c'\ ).p(c')}{p(\, d_i\,)}$$

10. Preprocess data
11. Make tokens
12. Perform steaming
13. Generate confection
14. Go to step no 08
15. Find out keyword match
16. Generate result.

B. *Memetic Algorithm*

Memetic algorithms (MA) represent one of the recent growing areas of research in evolutionary computation. The term MA is now widely used as a synergy of evolutionary or any population-based approach with separate individual learning or local improvement procedures for problem search. Quite often, MA are also referred to in the literature as Baldwinian evolutionary algorithms (EA), Lamarckian EAs, cultural algorithms, or genetic local search.

ALGORITHM

1. Start
2. Select data
3. Create user profile
4. Stored all data into dataset
5. If data of type structured then …..if no goto step 10
6. Analyze algorithm
7. Select trainset
8. Select Memetic Algorithm
9. Apply formula

Probability (accept) $\{1 \Leftrightarrow \Delta f > 0$

$$\frac{e^{k*\Delta f}}{f_{max}-f_{avg}}$$

Otherwise

Where k is a normalization factor, $\Delta f = f(i') - f(i)$, and we are assuming a maximisation problem. This mechanism seamlessly induces the Memetic Algorithm to oscillate between periods of intense exploitation whenever the spread of fitness in the population, i.e $f_{max}-f_{avg}$, is large, and periods of vigorous exploration when that spread is confined to a narrow interval.

10. Pre-process data
11. Make tokens
12. Perform steaming
13. Generate confection
14. Go to step no 08
15. Find out keyword match
16. Generate result.

V. MATHEMATICAL MODEL

A. Problem Description

Let S be a Social Media Data Mining tool which will prevent malicious query;

such that S={U,R,P,F,A,B,G,C,T,S,E,D,K|Φs} where

U represents normal user;U={ |Φu} and

R represents registration of user; R= $\{r_0,r_1,\ldots r_n|\Phi r\}$and

P represents profile created by user; P=$\{p_0,p_1,\ldots p_n|\Phi p\}$and

F represents friend list; F= $\{f_0, f_1,\ldots f_n|\Phi f\}$and

A represents request; A= $\{a_0, a1|\Phi a\}$ and

B represents ports; B= $\{b_0, b_1, \ldots b_n|\Phi b\}$and

G represents get all data; G= $\{g_0, g_1,\ldots g_n|\Phi g\}$and

C represents categories; C= $\{c_0,c_1,\ldots c_n|\Phi c\}$and

S represents streaming; S= $\{s_0, s_1, \ldots s_n|\Phi s\}$and

E represents confection; E= $\{e_0, e_1 ,\ldots e_n|\Phi e\}$and

D represents dictionary; D= $\{d_0,d_1, \ldots d_n|\Phi d\}$and

K represents keyword match; K= $\{k_0, k_1,\ldots k_n|\Phi k\}$.

B. Activity

Let fu be a rule of U into S such that each user sends request to the Server. fu(u0; u1:un) S□
{s0}_S
ff (f0; f1:fn) S□ {f0}_S
fp(p0; p1:upn) S□ {p0}_S
ft(t0; t1:tn) S□ {t0}_S
fg(g0; g1:gn) S□ {g0}_S
fc(c0; c1:cn) S□ {c0}_S
fb(b0; b1) S□ {b0}_S
fd(d0; d1:dn) S□ {d}_S
fe(e0; e1:en) S□ {e0}_S
33
Optimal Investigation of Social Media Data Mining using Memetic Algorithms
fa(a0; a1:an) S□ {a0}_S
fn(n0;n1:nn) S□ {n0}_S
fm(m0;m1:mn) S□ {m0}_S

VI. CONCLUSION

Mining social media data is helpful to researchers in learning analytics, educational data removal, and learning skill. It gives a way to examining social medium statistics that conquer the main restrictions of both physical qualitative analysis and huge scale computational study of user produced textual content. Two algorithms are useful for this classification first the Naive Bayes Multi-label Classifier and Second the Memetic algorithm. It notifies educational manager, and other applicable assessment makers to expand further accepting of engineering students institution understanding.

ACKNOWLEDGMENT

The authors would like to thank Prof. Samadhan Sonavane and Prof. A. D. Potgantwar Head of Computer Engineering Department, SITRC, Nashik for his help. This work was supported in part by the Sandip Foundation.

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 con- ference 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 or Engineering Education Annual Conference and Exposition,2008.

[4] C.J. Atman, S.D.Sheppard, J. Turns, R.S. Admans, L. Fleming, R. Stevens, R.A. Strev- eler, K. smith, R. Miller, L.Leifer, K. Yasuhara, and D.Lund, \Enabling Engineering student success: The Final report for the Center for the Advancement of Engineering Education", Morgan and 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] E.Go_man, The presentation of self in everyday life. Lighting source Inc., 1959.

[8] A. Go, R. Bhayani and L. Hunge,\Twiiter sentiment classi_cation using distant super- vision, "CS224N, project report Sanford, PP. 1-12,2009.

[9] X.Chen, M. Vorvoreanu, and K. Madhavan, \ A Web -Based Tool for Collaborative Social Media Data Analysis", presented at the Thired International Conferrance on Social Computng and Its application, Karlsruhe, Germany, 2013.

[10] J.L.Devore, Probability and statistics for engineering and the sciences. Duxbury Press, 2012.

[11] G. soumakas, I. Katakis, and I. Vlahavas , \Mining multi- lable data", Data Mining and knowledge discovery handbook, pp.667-685, 2010.

[12] G. Siemens and R. S. d Baker, \Learning analytics and educational data mining: To- wards communication and collaboration", in

Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 2012, pp. 252-254.
[13] Rajeshri Thorat, Ekta Pawar, Pranali Shendekar, Kajal Lokhande, ApekshaMengade \studyoverview of Mining social media data using Nave Bayes algorithm" March 2015
[14] Pooja R. Takle, Prof. Narendra Gawai \study overview of Identi_cation of Student's Behavior in Higher Education from Social Mediaby using Opinion based Memetic Clas- si_er" March 2015
[15] Xinchen, MihaelaVorvoreanu and Krishna Madhavan\Mining Social Media Data for Understanding Students' Learning Experiences" 2014
[16] X.Chen, M. Vorvoreanu, and K. Madhavan\studyoverview of A Web -Based Tool for Collaborative Social Media Data Analysis" January 2014
[17] Pallavi K. Pagare \study overview of Analyzing Social Media Data for Understanding Student's Problem "Dec. 2014
[18] M.Rost, L. Barkhuus, H. Cramer, and B.Brown\study overview of Representation and communication; challenges in interpreting large social media datasets" 2013
[19] Mattias Rost Louise Barkhuus Henriette Cramer Barry Brown\studyoverview ofRepre- sentation and Communication:Challenges in Interpreting Large Social Media Datasets" February 2013