Opinion Mining Results in the Education Domain

Ayoub Mohamed H. Elyasir

Faculty of Information Science and Technology
Multimedia University
Melaka, Malaysia

Abstract—**Opinions are essential for developing the services and facilities within the education domain. Students tend to express their opinions positively and negatively in the social network platform, such as Twitter and Facebook, toward various education universities. Collecting and analyzing those opinions will improve the overall education quality within the university's campus. In this paper, results are presented for the collected and analyzed tweets using Bag of Words (BoW) and Word Sense Disambiguation (WSD) approaches. The results include, tweets distribution for the chosen Malaysian universities, opinion mining timeline to illustrate the student responsiveness against the timeline, and polarity classification for the university entities and objects. The figure results are produced using nodebox and matplot libraries for python.**

Keywords—**opinion mining, data mining, polarity classification, bag of words, word sense disambiguation**

I.   Introduction

The internet became the most utilized communication method in our daily life which makes the education domain reliable and stronger. The universities provided different methods to evaluate their facilities, staff and services to enhance and make better environment for students. Online feedback is the most common method to obtain opinions and comments from the majority especially the opinions available in social network platform like Twitter and Facebook. We proposed opinion mining framework in the education domain (Elyasir & Anbananthen, 2012) that gathers, processes and analyzes the opinions' polarity to reflect the university's progress.

We developed opinion mining framework and built a corpus for the education domain to evaluate the Malaysian universities based on the opinions gathered and collected from the online resources, such as Twitter, Facebook and online blogs. Opinions on the Multimedia University were analyzed and classified based on the opinions' polarity. Two critical approaches in opinion mining are utilized, Bag of Words (BoW) and Word Sense Disambiguation (WSD), for building the framework to sufficiently collect, process the unstructured text and classify the data accordingly. BoW maps the opinion words in the feature vector to the formatting of the SentiWordNet (Esuli and Sebastiani, 2006) database. WSD provides the POS tagging with first time used technique (IOB) to tag the words based on their contextual comprehension.

This paper presents the results obtained from the opinion mining framework in the education domain including summary for the collected opinions and the corpus. The presented figure results are produced using nodebox and matplot libraries for python.

II.   The Collected Opinions

Total of 6156 opinions in three months for all universities and 1026 opinions for each university including, Multimedia University, University of Malaya, University Putra Malaysia, University Technology Malaysia, University Science Malaysia and University Kebangsaan Malaysia.

Name	Date modified
opinions_17-12-2012.txt	17/12/2012 1:19 PM
opinions_22-11-2012.txt	22/11/2012 2:56 AM
opinions_24-6-2012.txt	24/6/2012 9:46 PM
opinions_8-6-2012.txt	8/6/2012 5:02 PM
opinions_27-5-2012.txt	28/5/2012 3:26 AM
opinions_26-5-2012.txt	26/5/2012 5:09 AM
opinions_24-5-2012.txt	24/5/2012 11:31 AM
opinions_22-5-2012.txt	22/5/2012 10:55 PM
opinions_15-5-2012.txt	15/5/2012 2:13 AM

Figure 1: The collected opinions

Every corpus file consists of approximately 680 opinions. The corpus is formatted into 7 columns each one of them carries important information:

- First column is the hash tag that ensures resumption in case of service distortion and avoids opinion redundancy

- Second column is the university name

- Third column is the opinion language tag that detects the opinion language located in the fourth column

- Fourth column is the opinion

- Fifth column is the English translated version of the opinion if originally was not in English

- Sixth column is the polarity numerical value as either positive, negative or neutral based on the SentiWordNet database

- Seventh column is the timestamp which consists of the weekday, date and time

A. *Manual Tabulated Results*

The collected opinions contain noisy data which are not relevant to the topic of education domain and some others are not opinions but plain objective sentences. Table 1 below summarizes the information for the collected opinions in the corpus in four columns; column one is the university name and column two is the example for the collected opinion. Column three has three types of relevancy; (No) indicates that the sentence is not an opinion and not related to the education domain, (Yes but objective) indicates that the sentence is not an opinion but related to the education domain, (Yes) the sentence is an opinion and related to the education domain

TABLE I.   TWEETS DISTRIBUTION FOR EACH UNIVERSITY

University	Collected Opinion	Relevancy	Relevance Percentage
Multimedia University	Principles of Interactive Multimedia: "Principles of Interactive Multimedia" introduces all the contributory fie...	No	15 % = 154 out of 1026
Multimedia University	I'm at Multimedia University (MMU) (Bukit Beruang, Melaka) w/ 7 others	Yes but objective	23 % = 236 out of 1026
Multimedia University	To my opinion mmu lab is one of the best lab i see.....because i been to other university here in malaysia and all i think is that my university is the best even though we work hard to cope with there rules	Yes	62 % = 636 out of 1026
University of Malaya	I thought dad is sending me to work but he drag me to a wedding! (@ 9th Residential College, University Of Malaya)	No	13 % = 134 out of 1026
University of Malaya	1st time masok (@ Universiti Malaya (University of Malaya)	Yes but objective	26 % = 267 out of 1026
University of Malaya	"if you can survive the University of Malaya; you can survive anywhere, everywhere." - Datuk Johan Jaaffar; Chairman, Media Prima Bhd	Yes	61 % = 625 out of 1026
University Putra Malaysia	Wonderful conversation with School Leadership students from University of Malaya in Kuala Lampur this evening (@ Universiti Putra Malaysia (UPM))	No	27 % = 278 out of 1026
University Putra Malaysia	doing my research bout Universiti Putra Malaysia :) even rumah sebelah-sebelah je -_____-"	Yes but objective	13 % = 134 out of 1026
University Putra Malaysia	Universiti Putra Malaysia hooray hooray ~(..)~ Hello # twtupcampus	Yes	60 % = 615 out of 1026
University Technology Malaysia	This weblog was creatively designed for all bachelor candidates\' in Universiti Teknologi Malaysia Melaka (UTeM)	No	30 % = 307 out of 1026
University Technology Malaysia	See you guys tomorrow ☺ (@ Universiti	Yes but objective	18 % = 185 out of 1026

80

	Teknologi Malaysia (UTM))		
University Technology Malaysia	Missed UTM :((@ Universiti Teknologi Malaysia (UTM) w/ 3 others)	Yes	52 % = 533 out of 1026
University Science Malaysia	Sending my mom to be warded for her operation... (@ Hospital Universiti Sains Malaysia (HUSM))	No	18 % = 185 out of 1026
University Science Malaysia	Adobe TTT certificate of completion for Universiti Sains Malaysia (USM)	Yes but objective	28 % = 288 out of 1026
University Science Malaysia	IPS is fantastic and responsive - Universiti Sains Malaysia	Yes	54 % = 554 out of 1026
University Kebangsaan Malaysia	I'm at Hospital Universiti Kebangsaan Malaysia (HUKM) (Kuala Lumpur, WP Kuala Lumpur) w/ 4 others	No	10 % = 103 out of 1026
University Kebangsaan Malaysia	It's been a decade not visiting this place! (at Universiti Kebangsaan Malaysia)	Yes but objective	57 % = 585 out of 1026
University Kebangsaan Malaysia	Lectures in (UKM) only speak Bahasa ☹	Yes	33 % = 338 out of 1026

III. Findings Through The Opinions

Our framework evaluates the universities based on the collected opinions' polarity from the online resources. Figure 2 displays the polarity for the collected opinions by our framework. The left side of each bar represents the amount of positive opinions while the right side is the negative amount of opinions. From figure 2, we noticed that the top three universities had greater amount of subjective opinions due to their rapid policy changes and active respond from the student community. For example, Multimedia University kept changing its attendance, visa and immigration, grading and campus policies and rules that forced greater number of students to tweet or post their comments.
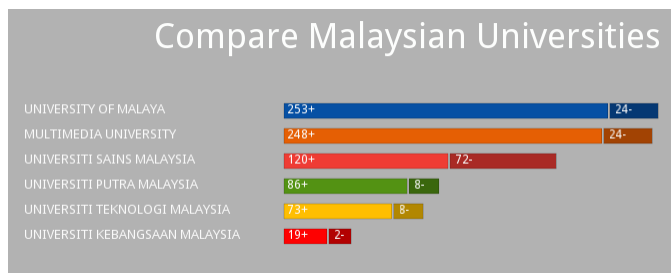

Figure 2: Polarity Classification

Figure 2 is the overall evaluation for the Malaysian universities as an outcome for the framework. We further analyze the opinions collected on Multimedia University. Figure 3 presents the timeline for the collected opinions based on the idea of "cause and effect" where the student opinions are depicted through the bar spikes against the time plot. The peaks above the dates are positive and the below ones are negative, in which the higher the peak the bigger is the student reaction polarity.

Through manual identification of the corpus, we noticed that the users express their opinions in the internet as a reaction to their university changing policy which makes the opinions decision or policy oriented. For example, Multimedia University began to force an attendance policy where the students are required to fulfill minimum of 80% of the overall attendance, this policy had a negative impact on the students' opinions in the beginning of June 2012 as shown in Figure 3 and 5.
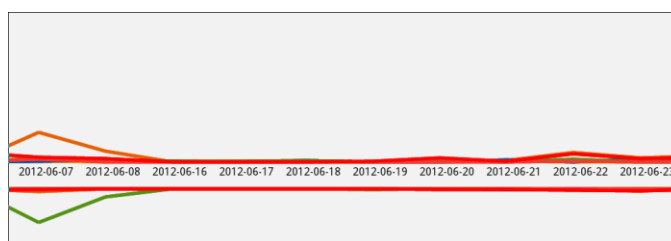

Figure 3: Opinion Timeline from 07/06/2012 – 23/06/2012

Provided those opinion timelines, we enable the top managers in the Multimedia University to explore the reaction of students after the policy changes or a new service is provided. The timelines are very effective for the decision making and direct the future plans such that, Figure 4 depicts the beginning of the "cause and effect" opinion timeline where the negative responses are considerably low while positive is at the peak in 22/05/2012 as a result of the students' satisfaction with the hostel maintenance and its new look in Multimedia University, Melaka campus.
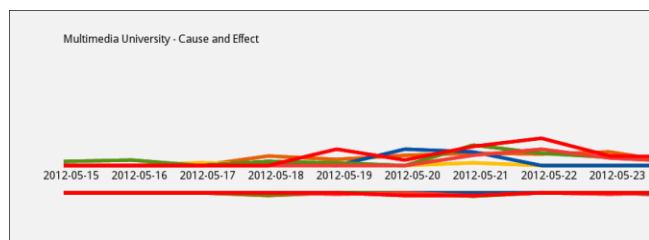
Figure 4: Opinion Timeline from 15/05/2012 - 23/05/2012

In Figure 5 the reaction of the new hostel changes and construction continue to have positive impact on the students up to 27/05/2012 together with the Deepvali holiday and a few job openings in the Multimedia University that left most of the student in a positive mood. An opinion contradiction happened at the end of Figure 3 and 5 as a result of the attendance policy changes that force the students to attend 80% of the overall registered credit hours for each subject as well as the PTPTN applications and approvals that mostly amplify the negative side of the graph.

The opinion timeline figures, 3-5, show the impact of the top management decisions on the students across certain duration of time. We created equalizer figures using matplot library in python to graph and plot all the features in our corpus vector. Equalizer figures help in detailed understanding for the satisfaction status on each feature and object inside the corpus. Figure 6 depicts the overall opinion polarity on Multimedia University's certain objects with polarity count on the Y axis and the objects distributed in the X axis. Figure 7 and 8 are zoom in for the figure 6, whereby they clearly display the respective objects and their corresponding polarity counted in the scale of 0 to 10.
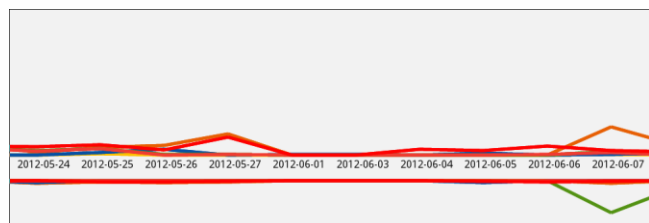

Figure 5: Opinion Timeline from 24/05/2012 - 07/06/2012
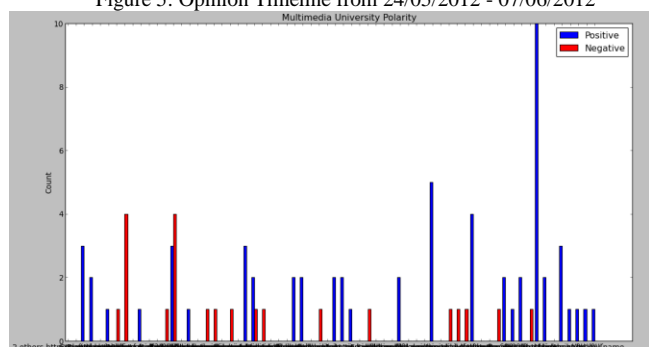

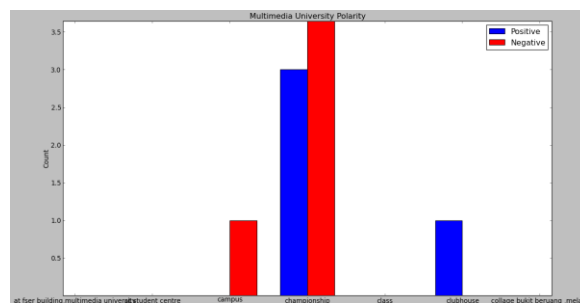Figure 6: Multimedia University Polarity

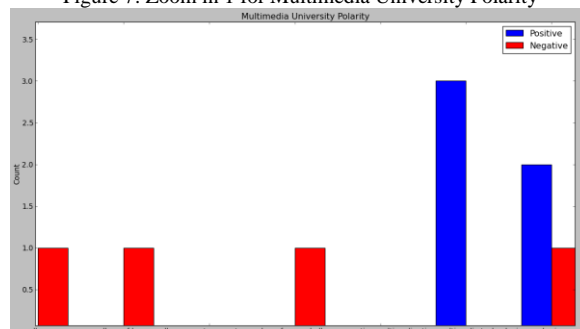Figure 7: Zoom in 1 for Multimedia University Polarity

Figure 8: Zoom in 2 for Multimedia University Polarity

A. Results Obtained Through "cause and effect" and Equalizer Figures

From the "cause and effect" figures on Multimedia University:

- Students are not satisfied with the attendance policy

- The hostel maintenance in Melaka campus had good satisfaction

- Some incidents in Melaka campus resulted in aggressive responses from the students about the campus security

- Job openings through the seminars and the announcements in Multimedia University bulletin board had positive impact on the students

- PTPTN applications and local English tests had negative impression

From the "equalizer" figure:

- Faculty of Engineering and Technology in Melaka had negative impact on students because of certain proprietary policies. The faculty dean was negatively mentioned several times, about 28 times in June corpus

- Football championship motivated the students to post and tweet positive opinions

- Various events and activities in cyberjaya left positive impression on students and outside tweeters

- Negative popularity for the security guards

- Cyberjaya campus in Multimedia University had greater satisfaction than Melaka

IV. **Conclusion**

The proposed framework (Elyasir & Anbananthen, 2012) is to analyze the collected opinions from Twitter, Facebook and some online blogs, which are related to the Malaysian universities in general and Multimedia University in particular. Then, we evaluate some Malaysian universities and provide thorough analysis on Multimedia University with various output representations to help in the decision making process. Evaluating the universities has no specific criteria or characteristics rather all public opinions are gathered and classified according to their respective polarity, positive, negative or neutral

Figure 2 illustrates the positivity/negativity of the collected opinions from the internet, with 1st place for University of Malaya, 2nd for Multimedia University, 3rd for University Sains Malaysia, 4th for University Putra Malaysia, 5th for University Technology Malaysia, 6th and last for University Kebangsaan Malaysia. Figure 3-5 depict how the opinions are significantly affected by the rules and regulations updates in the institution corresponding to certain time duration, in which helps the top managements to respond appropriately according to the classified polarity. Figure 6-8 illustrate the opinion polarity on the respective objects using matplot library and their equalizer models. They show the most commented objects in Multimedia University campus including, regulations and rules, campus facilities and events. These representations are extremely useful to understand the interesting objects among the students, and whether classified as positive or negative.

A. *Future Work*

Extensive filtration process must be embedded to filter out false opinions, rude and swearing words, advertisements and spamming just like the case in the "flame framework" proposed by Spertus (1997), whereby it contains natural language processing and vector utilization to detect the "flames". This step is necessary as we noticed that majority of the opinions contain inappropriate words used by angry students commenting on their universities.

Construct standardized corpus for education domain to collect the publics' opinions on various universities and institutions worldwide, just like the case in other domains such as movies popular database and corpus (IMDB). This helps researchers all over the world to train their classifiers on one and for all corpus databases to obtain results with standard benchmark in the education domain.

References

[1] Elyasir, A., & Anbananthen, K. (2012). Opinion Mining Framework in the Education Domain. International Journal of Machine Learning and Data Mining (waset), vol 6. (pp. 34-40).

[2] Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. Proceedings of LREC, 6, pp. 417-422.

[3] Spertus, E., & others. (1997). Smokey: Automatic recognition of hostile messages. Proceedings of the National Conference on Artificial Intelligence, (pp. 1058-1065).