Mining sentiments in SMS texts for teaching evaluation

Chee Kian Leonga,*, Yew Haur Leeb, Wai Keong Makc

aCentre for Applied Research, SIM University, 461 Clementi Road, Singapore 599491, Singapore
bSchool of Business, SIM University, 535A Clementi Road, Singapore 599490, Singapore
cEdtrix Solutions, VBOX 882284, Singapore 919191, Singapore

ARTICLE INFO

Keywords:
Sentiment mining
SMS texts
Education

ABSTRACT

This paper explores the potential application of sentiment mining for analyzing short message service (SMS) texts in teaching evaluation. Data preparation involves the reading, parsing and categorization of the SMS texts. Three models were developed: the base model, the "corrected" model which adjusts for spelling errors and the "sentiment" model which extends the "corrected" model by performing sentiment mining. An "interestingness" criterion selects the "sentiment" model from which the sentiments of the students towards the lecture are discerned. Two types of incomplete SMS texts are also identified and the implications of their removal for the analysis ascertained.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

In education, it is typical to have an end-of-course evaluation in which the students can express their opinions about the teachers or instructors as well as the content and delivery of the course. Such evaluation generally involves a Likert scale (Likert, 1932) in which the students will indicate their opinions on the course by checking how strongly they agree or disagree with carefully constructed statements about the teachers, the course content and the delivery of the course. Typically, there are five levels: strongly agree, agree, neutral, disagree and strongly disagree.

The proliferation of handphones and usage of short message service (SMS) among students offer an opportunity for such evaluation to be conducted immediately after a lesson. In this way, prompt feedback can be provided to the instructors and the necessary actions can be taken during the course.

Short message service (SMS) is a component of the Global System for Mobile Communications (GSM) series of standards in 1985. SMS is a means of sending short text messages not longer than 160 characters (including spaces) between mobile phone devices. SMS text messaging is the most ubiquitous data application in the world, with annual worldwide SMS volumes estimated to have increased from 2.6 trillion in 2007 to nearly 5.5 trillion SMS in end-2009 (Portio Research, 2010). The term SMS is often used as a synonym for all types of short text messaging, as well as the activity of creating SMS texts. In this paper, we use the term "SMS texts" to refer to the individual SMS text messages sent.

This paper examines the opportunities and challenges of obtaining teaching feedback from students via SMS texts. Specifically, it explores the potential applications of text mining on SMS texts used for teaching feedback. To the best of our knowledge, the present paper is the first to explore such applications in educational data mining.

In addition, the paper also contributes to the literature on text mining (Feldman & Sanger, 2007) and sentiment mining (Liu, 2010). Recent research on sentiment analysis or opinion mining (Liu, 2010) has focussed on the mining of massive volume of texts with opinions or sentiments. Unlike most texts, however, SMS messages are comparatively short and offer a number of challenges. Firstly, because of its nature, the responses given will be different for different users. So if there are 100 students in a class, no two students would give the same SMS feedback even when linguistic rules are observed and no spelling errors are committed. Secondly, SMS messages are pervaded by abbreviated shortcuts (e.g. oic), emoticons (e.g. :) ). Thirdly, it is common for SMS texts to contain "noises" such as spelling and grammatical errors, incomplete sentences, false starts, repetitions and pause filling words such as um and uh. The challenge, in this case, is to extract insights from an analysis of such noisy and unstructured data in SMS texts. In exploring these challenges, the paper paves the way for further research by proposing a systematic approach to analyze such SMS texts for education purposes.

This paper is organized as follows. Section 2 provides some background on the project for using SMS texts for teaching evaluation. Section 3 discusses some of the data pre-processing steps required before modeling can be performed on the SMS texts. Modeling and model evaluation are also explored. In Section 4, the results of the sentiment analysis are presented and the implications...
of incomplete SMS texts are considered. Finally, Section 5 concludes and proposes possible directions for future research.

2. Background

The study makes use of SMS text messages generated from a SMS Response System (SMSRS), which is a platform independent web-application system\(^1\) developed by one of the co-author (Wai Keong Mak). The system allows audience in a lecture, tutorial or seminar to send in their comments, responses and feedback via SMS. The response can be in the format of a response to a multiple-choice question, a poll or a simple text message using a mobile phone's SMS facility. For responses of the first two types which involve response of a single alphabet, the answers can be immediately tabulated, graphed and displayed to the class via a website. The focus of this study is on the third type involving a SMS message in the form of free text.

In this particular study, the setting is a physics lecture given by a lecturer in a junior college, the equivalent of high school. The physics lecturer delivered his lecture, after which the students were asked to submit their feedback on the lecture via SMS texts. The lecturer was interested in finding out how the students felt about the lecture. The feedback provided by the students would allow him to moderate his delivery of subsequent lectures.

3. SMS texts: text mining and sentiment mining

3.1. Data preparation

Analysis of SMS texts can be difficult because of the incompleteness of the data. Firstly, the default limitation on the maximum number of characters which can be stored for each SMS text results in only a limited amount of text which can be analyzed. Secondly, many of these SMS texts are incomplete. In this section, we detail the data preparation process and also deal with the issue of incomplete SMS texts.

The data preparation process of SMS texts is made up of 3 phases, as shown in Fig. 1. They are:

1. Reading the SMS corpus or collection of SMS texts.
2. Parsing of the SMS texts;
   - **Part of speech (POS) tagging**: Each term is tagged with a POS tag. In this analysis, POS tagging is an intermediate step that is carried out to identify concepts. Table 1 gives a description of the different POS tags used in this analysis.
   - **Stemming**: Inflection refers to the change in form in words to mark gender, number or tense. For example, the root word prevent has the following inflections – prevents, prevented and preventing. Stemming is the process of mapping all the variants to its root word.
   - **Synonyms**: Certain terms or concepts may mean the same thing. For example, cancer of the thyroid and thyroid cancer refer to the same thing. These will need to be mapped to each other.
   - **Exclude list**: In this analysis, pronouns, particles and prepositions (except of) are not extracted during parsing since these POS are usually to be redundant in text mining. For additional concepts that are redundant, they are populated in the exclude list.
   - **Type/entity extraction**: For each concept extracted, a type or entity is also identified. A type is defined as a semantic grouping of concepts. Types include higher-level concepts, positive and negative words and qualifiers, contextual qualifiers, first names, places, organizations, amongst others.
   - **Exclude list**: In this analysis, pronouns, particles and prepositions (except of) are not extracted during parsing since these POS are usually to be redundant in text mining. For additional concepts that are redundant, they are populated in the exclude list.

3.2. Exploratory data analysis and visualization

Exploratory data analysis of the prepared SMS could be performed at either the concept or category level.

Exploratory data analysis at the concept level involves viewing a list of concepts extracted, the relevant statistics in terms of frequency and percentage of occurrence of the respective concepts as well as the number and percentage of documents in the corpus that contain the concepts. Fig. 2 displays the exploratory analysis for the most frequently occurring concepts. There are a total of 118 concepts available for scoring. The corresponding percentage and frequency of the occurrence of the concept are shown in the third and fourth column while that for the occurrence in the number of documents are given in the fifth and sixth column. As shown in the diagram, the most frequently occurring concept is lecture: it occurs in 18.4% of the SMS corpus and 32.8% of the SMS texts in the corpus. This is followed by the concept pace and to a lesser extent, fun, lesson and feedback.

1 URL: www.smsrs.com.
In exploratory data analysis at the category level, a list of categories are extracted. Further, the relevant statistics in terms of number of concepts contained in the category and the corresponding number of documents in that category are also obtained. In Table 3, the categories extracted are in the first column with the corresponding number of concepts listed in the adjacent column. The category with the highest number of concepts is lecture (7), followed by pace (6), jokes (5) and teacher (4).

To explore the lecture category further, we explore the “lecture” category compared to other categories. From Fig. 3, the category lecture occurs in all the SMS texts in 44 of 119 SMS texts in the SMS corpus, whereas pace and feedback occur in only 11 and 8 SMS texts, respectively.

Since there are a number of concepts extracted, visualization of the relationship between concepts can provide a handle on interesting patterns amongst concepts. Fig. 4 displays the links between concepts. The concept lecture has the most links with links to 15 concepts. The thickness of the line shows the number of documents with both concepts and the thickest appear to the link between lecture, ppl and pace, which appear together in 12 documents. The next thickest is the link between lecture, explanation and feedback appearing together in 10 documents.

3.3. Modeling and evaluation

Once the SMS corpus has been parsed and categorized, the next step is to perform modeling of the SMS texts. In this paper, we consider three models: the base model, the “corrected” model and the “sentiment” model.

- **Base model**: The whole SMS text corpus is being analyzed without adjusting for spelling error.
“Corrected” model: Any spelling error is classified under the same concept. For instance, the word slp will be classified under the concepts sleep and sleepy, expltn and explanation under the concepts explanation.

“Sentiment” model: This extends the “corrected model” by performing sentiment mining.

Table 4 presents a comparison of the three models. The question is: given the three models, which one would be selected? To the best of our knowledge, there is currently no guidance in the literature regarding the selection of text mining models. We proceeded as follows.

In deciding between the base model and the “corrected” model, one could observe that with an increased volume of SMS texts, the increase in spelling mistakes may result in a multiplication of concepts which are unnecessary. In this case, the base model has 121 concepts while the “corrected” model has pruned the number concepts to 118, which is not a significant reduction. On the other hand, the “sentiment” model increased the number of concepts to 164 but also provided additional information on the type of concepts (positive and negative). One could perhaps make the choice of one model amongst the three on the basis of “interestingness”. What is perhaps “interesting” to the lecturer would be how the students find his lecture, particularly the positive and negative aspects. Since the “sentiment” model provides such information, it is the model of choice.

4. Discussion

4.1. Sentiments towards lecture

Fig. 5 presents the text link analysis of the positive and positive qualifier with other concepts. The link with the highest global count of 6 is that between “pace” and the positive qualifier “good”. This is followed by that between “lecture” and the positive quali-
fier “interesting”. In general, the text link in the “positive” map appears to be quite populated, suggesting that the lecture pace is generally well-received by most students.

Fig. 6 displays the text link analysis of the negative and negative qualifier with other concepts. There are two links with global count equal to 2: the link between “lecture” and the negative qualifier “too long” and that between “examples” and the negative qualifier “more”. In general, there are fewer negative comments on the lecture than positive comments.

4.2. Incomplete SMS texts

Earlier, we mentioned that there are two types of incomplete SMS messages: Type 1 incompleteness due to the limitations on the maximum number of characters which can be stored for each SMS in the online feedback system; and Type 2 incompleteness involving the respondent sending a single alphabetical letter instead of a complete message. In this section, we discuss whether removing such incomplete SMS texts will result in additional information.

To this end, we analyze the results of the ‘sentiment’ model under four cases:

1. Complete: both types of incompleteness are retained;
2. Only Type 1 incompleteness is removed.
3. Only Type 2 incompleteness is removed.
4. Both types of incompleteness are removed.

We determine the number of concepts and types generated in each of these cases. The results are summarized in Table 5.

From Table 5, it can be noted that the removal of either or both types of incompleteness does not affect the number of types generated by the ‘sentiment’ model. Similarly, removing Type 2 incompleteness alone does not affect the number of concepts generated. In contrast, the removal of Type 1 incompleteness reduces the number of concepts generated by about 8%. This is not surprising since the removal of a single alphabet SMS text probably does not contribute to any loss of concepts while the removal of a lengthy but incomplete SMS text can lead to potential loss of concepts. In short, it would appear that the analysis for this dataset is robust to the removal of both types of incompleteness.

5. Concluding remarks

In this paper, we explore the potential application of text mining and sentiment mining for analyzing short message service (SMS) texts in teaching evaluation. As far as we are aware, this paper is the first to explore such applications in educational data mining. Data pre-processing involves the reading, parsing and catar-

Table 4

<table>
<thead>
<tr>
<th>Types of concepts</th>
<th>Number of concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base model</td>
<td>“Corrected” model</td>
</tr>
<tr>
<td>Not available</td>
<td>Not available</td>
</tr>
<tr>
<td>Positive/negative</td>
<td>Positive/negative</td>
</tr>
</tbody>
</table>

Fig. 5. Positive aspects of the lecture.

Fig. 6. Negative aspects of the lecture.
eegorization of the SMS texts in the SMS corpus. We develop three models: the base model, the “corrected” model which adjusts for spelling errors and the “sentiment” model which extends the “corrected” model by performing sentiment mining. Next, we apply an “interestingness” criterion to select the “sentimental” model. Based on this model, we discuss the sentiments of the students towards the lecture. Finally, we ascertain that the results of the analysis is robust to the removal of either or both types of incomplete SMS texts. In exploring these challenges, the paper paves the way for further research by proposing a systematic approach to analyze such SMS texts for educational purposes.

In particular, the sentiment analysis of the SMS provides the lecturer with timely information to gauge the positive and negative aspects of the lecture delivery and reflect on possible improvement. This would be an improvement over the end of course evaluation which only captures the performance of the lecturer at the end of the course, by which time corrective actions can hardly be taken until the course is offered in subsequent semesters.

A potential area of future research is trend analysis across similar evaluations obtained over a period of time (Lent, Agrawal and Srikant, 1997). Such analysis requires the stamping of date and time on SMS texts for different SMS corpora so that comparisons can be made on the evaluation of the lecturer in several lessons and possibly pertaining to different topics in the syllabus. In this way, a holistic picture of the lecturer’s teaching performance can be achieved.

References


