

CourseObservatory: Sentiment Analysis of Comments in Course Surveys

Haydée G. Jiménez, Marco A. Casanova
Departamento de Informática
PUC-Rio
Rio de Janeiro, Brazil
{hjimenez, casanova}@inf.puc-rio.br

Bernardo Pereira Nunes
Depto de Informática, PUC-Rio
Depto de Ciência da Computação, UNIRIO
Rio de Janeiro, Brazil
bnunes@inf.puc-rio.br

Anna Carolina Finamore
Instituto Superior Técnico
Universidade de Lisboa
Lisboa, Portugal
anna.couto@ist.utl.pt

Abstract—This article describes experiments with a tool, called **CourseObservatory**, that applies sentiment analysis to comments made by students during course surveys. The main objective is to provide course coordinators and teachers with relevant information about the courses students take based on their qualitative feedbacks. The experiments use a dataset that contains comments in Portuguese found in questionnaires filled out by students from mid-2005 to the first semester of 2018, with a total of nearly 170,000 comments, after a cleaning process that removes blank comments. The experiments show that comments made by students are influenced by the final status achieved (approved or failed), among other facts.

Keywords—Sentiment Analysis; Educational Data Mining; Data Visualisation

I. INTRODUCTION

This article describes experiments with a tool, called **CourseObservatory**, that applies sentiment analysis to comments made by students during course surveys. The experiments with the tool use a dataset that contains comments made by the students of a Brazilian University about their teachers and courses. The dataset contains records from mid-2005 to the first semester of 2018, with a total of nearly 170,000 comments.

The comments are in Portuguese and are part of systematic surveys carried out during the enrollment for a new semester, when students are invited to evaluate the courses they took in the previous semester. The questionnaire contains multiple-choice questions about the usefulness of the course, the bibliography adopted and the pedagogical abilities of the teacher, and a free text area where students can voice their opinion about the course and the teacher. It is these comments that the article focuses on.

The main contribution of the article is to investigate how to provide the degree coordinator and the teachers with relevant information based on students' textual comments about the courses. Indeed, over the years, it became clear that a degree coordinator cannot fully benefit from such comments since their sheer number makes it nearly impossible to manually process them.

Specifically, this article addresses three questions: (1) Is there a dependency between the average teacher evaluation and the tendency of the comments of his/her students? (2) Can the tendency of a student's comment be influenced by the final status achieved in the course? (3) Can the tendency of students' comments help unveil how they reacted to changes in the way that the course is taught?

This article is organized as follows. Section 2 summarizes related work. Section 3 briefly describes the

CourseObservatory tool, the dataset used in the experiments, and assesses the performance of the tool. Section 4 shows how the tool can help address the questions posed above. Section 5 presents the conclusions and suggests future work.

II. RELATED WORK

Due to the explosion of social media, today the average citizen is overwhelmed by a large volume of text, available in digital format, expressing opinions, such as blogs, forum discussions, Twitter, among others. Sentiment analysis or opinion mining is an active research area devoted to the transformation of such data into useful knowledge by automatically analyzing natural language texts to extract the sentiments, opinions, and emotions they drive [1,2]. Based on this knowledge, decisions can be made regarding trends, products, etc. Ravi et al. [3] conducted extensive surveys about sentiment analysis, and summarized the different tasks and approaches.

Wilson et al. [4] analyzed the sentiment at a phrase-level, determining if an expression is neutral or polar; in the case it is polar, they then analyze the context. They presented four categories to classify sentiment at a phrase-level: positive, negative, both and neutral. This study proposed several ideas that will we will adopt in this paper.

Balahadia et al. [5] presented a tool for the analysis of comments made by students to help improve the performance of teachers. The tool evaluates both quantitative and qualitative information. However, the tool is not directly applicable to the investigation in this article since it is limited to English.

Menaha et al. [6] proposed a system based on the analysis of the repetition of keywords in a comment, which extracts the main topic to which the comment refers. Once the topic is identified, the system carries out a clustering process to classify the comments into positive or negative. By contrast, our course surveys are based on fixed questionnaires where the topics are pre-defined and the possible answers students may select for each topic are categorized from 1 (“completely disagree”) to 5 (“completely agree”).

Microsoft offers the possibility of adding a library to Excel to perform sentiment analysis. The tool accepts up to 10,000 records at a time, and returns the result of the analysis in a new tab. It classifies comments into 6 categories: positive, negative, very positive, very negative, neutral and none. The tool performs sentiment analysis in several languages, including Portuguese. For this reason, we chose this tool as a baseline.

Specifically, for Portuguese, one may quote, for instance, Souza and Vieira [7] and França and de Souza [8], which

applied sentiment analysis to tweets. Although tweets and comments have similarities, at least in terms of size, our study focuses on students' comments in the context of course surveys, unlike these references.

III. THE COURSEOBSERVATORY TOOL

A. Outline of the Tool

This section outlines the CourseObservatory Tool developed to apply sentiment analysis to students' comments. The major steps of the tool go as follows.

The first step separates the students' comments using punctuation marks, such as "." and ",", and keywords, such as "porém" ("although") and "mesmo assim" ("even though"). The idea is to be able to analyze each student's idea separately, and not as a whole, because relevant information could be lost when analyzing a comment as a single idea. The next step eliminates stop words and transforms phrases into a canonical form. For example, "não foi bom" ("it was not good") and "não é bom" ("it is not good") were both transformed to "não bom" ("not good"). In the second step, the user inputs a dictionary with keywords, based on an analysis of the most common terms that occur in the comments. For the construction of this dictionary, the user can adopt any tool he wishes to obtain a set of words that better represent the feelings to be analyzed in the comments. Then, he must manually classify each term as positive or negative. The core process uses the dictionary to classify each comment as: "positive", "negative", "mixed" (when the comment had both positive and negative phrases) and "unknown" (when it was not possible to classify the comment). For example, "nada a dizer" ("not comment")

B. Description of the Test Dataset

The dataset used to test the tool comes from systematic surveys carried out during the enrollment for a new semester, when students are invited – answering the survey is voluntary – to evaluate the courses they took in the previous semester. The questionnaire contains objective questions about the usefulness of the course, the bibliography adopted, and the pedagogical abilities of the teacher, among others. There is also a free text area where students can voice their opinion about the course (in Portuguese).

The dataset contains records from mid-2005 to the first semester of 2018, with a total of 918,439 records. After eliminating all blank comments, the dataset is left with over 168,760 records.

C. An Evaluation of the Tool

For the construction of the dictionary, an N-Gram Extraction Tool was applied, based on the study of Lü et al. [9]. Words from 1 to 5-gram, whose absolute frequency was larger than 100, were extracted and then analyzed to uncover which words reflect feelings to be placed in the dictionary. The dictionary consists of 211 terms, where 123 were manually classified as negative, 86 as positive and 2 as mixed. Several terms were very similar because it was necessary to consider the teacher's gender ("bom professor", "boa professora"), spelling mistakes ("exelente" instead of

"excelente"), and the young people's jargon ("mto" instead of "muito"). Although the dataset is composed of thousands of comments, the terms that make up the dictionary barely exceed 200 terms. Indeed, the way that students express positive and negative comments proved to be quite limited.

Table I shows the distribution of the classification of the comments that the tool produced.

TABLE I. DISTRIBUTIONS OF RECORDS

	<i>Positive</i>	<i>Negative</i>	<i>Mixed</i>	<i>Unknown</i>	<i>Total</i>
Comments	74,947	34,831	23,822	35,160	168,760

To check the performance of the tool, we randomly chose 150 records, manually classified the comments, and then measure the performance of the tool.

For the analysis, we adopted precision and recall, two metrics commonly used to measure the performance of search and pattern recognition systems. A high precision means that more relevant than irrelevant results are returned, while a high recall means that most of the relevant results are returned. Table II shows the results, which were above 0.75, indicating that the tool had a good performance.

TABLE II. RESULTS OF THE PERFORMANCE EVALUATION.

	Precision	Recall
Positive	0.93	0.83
Negative	0.90	0.76
Mixed	0.77	0.83

IV. APPLICATIONS OF THE COURSEOBSERVATORY TOOL

This section shows how to use the CourseObservatory tool help answer the questions posed in the introduction.

Consider the first question: Is there a dependency between the average teacher evaluation and the tendency of the comments of his/her students?

To help answer this question, one may compare the average evaluation obtained by a given teacher per semester, computed from the class evaluation questionnaire, with the distribution of the classification of the students' comments, obtained with the CourseObservatory tool.

Consider Fig. 1, constructed as described in the previous paragraph, for a given teacher over several semesters. It was observed that he/she is a professor who usually receives more than ten comments per semester and his/her semester evaluation is usually above 80. In the three semesters his evaluation was below 75 points, the periods 2005.2, 2006.1 and 2017.1, in two of them, he/she received a considerable number of negative comments. But on the semesters 2010.1 and 2012.1, where he/she obtained an evaluation of 85 points, he/she received more negative comments than positive ones. It must be pointed out that the participation of students in the course surveys is not mandatory, but for a course or professor to be evaluated he must have been evaluated by a representative percentage of the total number of students enrolled.

Consider now the second question: Can the tendency of a student's comment be influenced by the final status she/he achieved in the course?

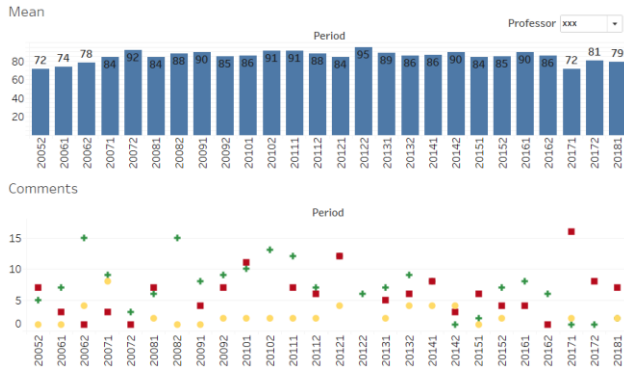


Figure 1. Comparison between average and comments.

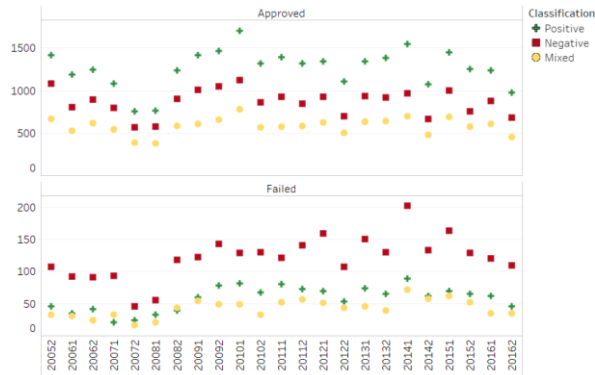


Figure 2. Distribution of comments by period and final situation.

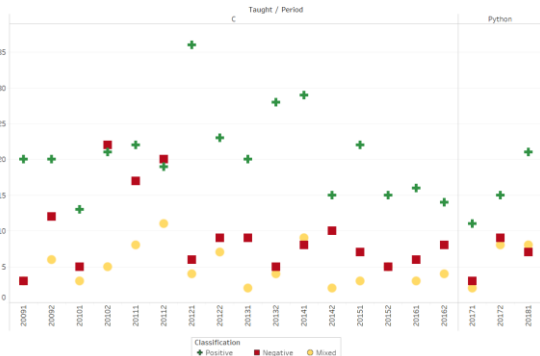


Figure 3. Distribution of comments by period.

At the end of the semester, a student receives the final status in each of the course in which she/he is enrolled: "AP" for approved and "RM" for failed. The students receive this final status before making the teacher's evaluation.

Fig. 2 shows the classification of the comments of the students according to their final status in each course, including all comments by period. It can be seen that most of the comments in a period are made by students who were approved and that the majority the comments are positive. It is also possible to observe that, by contrast, most of the comments of students who failed are negative. Then, this brief analysis in fact suggests that the type of comment made by a student is associated with the final status achieved.

Finally, consider the third question: Can the tendency of students' comments help unveil how they reacted to changes in the way that the course is taught?

To illustrate how to answer this question, we selected an introductory programming course offered by the Informatics Department to all engineering students. From 2009 to the second semester of 2016 the course was taught in C, and since 2017 it is taught in Python.

Regardless of the number of students enrolled in the course, the percentage of students who made comments remains similar, between 7% and 16% per period. Fig. 3 shows that most of the comments made about these courses are positive. Negative comments were generally in the range of 5 to 10 per period. Thus, this tool could also provide indications about how students perceived modifications in a course curriculum.

V. CONCLUSIONS AND FUTURE WORK

We first outlined the CourseObservatory tool that applies sentiment analysis to comments made by students in course evaluation surveys. Then, we assessed the performance of the tool and compared it with a baseline tool. Finally, we showed how the tool can help answer three basic questions that involve a sentiment analysis of students' comments.

In addition to classifying the comments as positive or negative, it would be interesting to investigate the social, cultural and psychological characteristics of the student (who evaluates), the teacher (who is evaluated) and the course (what is evaluated). For example, the context of a Religion student/teacher/course is quite different from that of a Calculus student/teacher/course.

REFERENCES

- [1] B. Liu. "Sentiment analysis and opinion mining". Synthesis lectures on human language technologies, 5(1), 1-167. 2012.
- [2] B. Liu, and L. Zhang. "A survey of opinion mining and sentiment analysis." In Mining text data, pp. 415-463. Springer, Boston. 2012.
- [3] K. Ravi and V. Ravi. "A survey on opinion mining and sentiment analysis". Knowledge-Based Systems, v.89 n.C, pp 14-46, Nov 2015.
- [4] T. Wilson, J. Wiebe , P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis", Proceedings of the HLT-EMNLP 2005, pp.347-354, Oct. 06-08, Vancouver, British Columbia, Canada.
- [5] F.F. Balahadia, M.C.G. Fernando and I.C. Juanatas, "Teacher's Performance Evaluation Tool Using Opinion Mining with Sentiment Analysis", IEEE Region 10 Symp. (TENSymp), pp. 95-98. 2016.
- [6] R. Menaha, R. Dhanaranjani, T. Rajalakshmi and R. Yogarubini. "Student Feedback Mining System Using Sentiment Analysis". IICATR, v. 6, pp. 1-69, 2017. DOI: 10.7753/IICATR0601.1009.
- [7] M. Souza and R. Vieira. "Sentiment analysis on twitter data for portuguese language". PROPOR'12, Helena Caseli, Aline Villavicencio, Antônio Teixeira, and Fernando Perdigão (Eds.). Springer-Verlag, Berlin, Heidelberg, pp. 241-247. 2012.
- [8] K. França de Souza. "UniLex: Método Léxico para Análise de Sentimentos Textuais sobre Conteúdo de Tweets em Português Brasileiro". Ph.D. Dissertation. PUC de Minas Gerais, Minas Gerais, Brazil. Abakós, v.5, n.2, pp. 79-96. 2017.
- [9] X. Lü, L. Zhang and J. Hu (2005) Statistical Substring Reduction in Linear Time. In: Su KY., Tsujii J., Lee JH., Kwong O.Y. (eds) Natural Language Processing – IJCNLP 2004. IJCNLP 2004. Lecture Notes in Computer Science, vol 3248. Springer, Berlin, Heidelberg.