An Approach for Thematic Relevance Analysis Applied to Textual Contributions in Discussion Forums

Crystiano José Richard Machado, University of Pernambuco, Recife, Brazil
Alexandre Magno Andrade Maciel, University of Pernambuco, Recife, Brazil
Rodrigo Lins Rodrigues, University Federal Rural of Pernambuco, Recife, Brazil
Ronaldo Menezes, University of Exeter, Exeter, UK

ABSTRACT

Discussion forums in learning management systems (LMS) have been shown to promote student interaction and contribute to the collaborative practice in the teaching-learning process. By evaluating the postings, teachers can identify students with learning difficulties. However, due to the large volume of posts that are generated on a daily basis in these environments, manual analysis becomes impractical. This article proposes a mechanism to support teaching through the thematic relevance analysis of the posts made by students in discussion forums. For this, text mining and metrics from network science were used to process and extract characteristics of the texts. Then, the processed texts were classified through supervised learning algorithms. The results show that the use of these techniques may generate potentially useful indicators for teachers to help them improve their pedagogical practices.

KEYWORDS

Discussion Forums, Learning Management Systems, Network Science, Supervised Learning, Support Systems, Teaching, Text Mining, Thematic Relevance Analysis

1. INTRODUCTION

Learning Management Systems (LMS) offer many features and tools that may be used as instruments of interaction between students and instructors. These instruments contribute to the learning experience because they support synchronous and asynchronous communication between all the participants in Distance Learning (DL) courses (Maciel, Rodrigues, & Carvalho, 2014). One of the most used features in these environments is the discussion forum. According to the Census EaD.br (2016), the discussion forum is a tool widely adopted by educational institutions as a way to communicate with students, being used approximately 72% of Brazilian distance learning courses.

In discussion forums, the number of posts at the end of a debate varies according to the theme, the interest of participation provoked by the topic, and even the profile of the students participating. Parts of the postings may contain interesting arguments or positions on the topic of discussion, such as citation to important concepts or arguments linking different concepts. On the other hand, there are often posts that do not contribute significantly to the discussion, including information that is not relevant to the topic.

DOI: 10.4018/IJDET.2019070103

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
The involvement of students in the forums is an important activity for the construction of knowledge and generally a sign of student interest. Through the analysis of students’ interactions in the tool, instructors can identify learning difficulties. Yet, due to the large number of messages that are created, a manual analysis of student posts can become a very time-consuming task. Furthermore, the analysis of the written content in forums may free instructors to spend more time focusing on causes that may be leading to possible lack of participation by students (Azevedo, Behar, & Reategui, 2011).

In this context, approaches that use Text Mining (TM) are becoming very promising in the process of knowledge discovery, by using techniques of analysis and extraction of data from unstructured texts (Morais & Ambrośio, 2007). More specifically, an option that may be considered together with natural language processing (NLP) is network science (NS) (Albert & Barabási, 2002; Newman, 2003), a concept that comes from statistical mechanics and makes extensive use of graph theory algorithms. The appropriate treatment of texts can provide relevant information about students, as well as their needs, knowledge, and interests (Panceri & Menezes, 2007).

Therefore, the objective of this work is to introduce an approach for modeling and development of a solution that allows the thematic relevance analysis of posts written by students in discussion forums. This approach prioritizes the identification of posts of students who may be facing difficulty in learning, while making sure the relevant contributions on the topic of the discussion are also identified. We developed a computational model that seeks to evaluate the thematic relevance of the posts made by students in relation to the opening text of the forum, which we assume to define the problematization of the discussion.

Besides this introduction, this article is organized into five other sections which present the related works, the proposed workflow, the experiments, the results and discussions, and the final considerations.

2. RELATED WORKS

Several scientific studies have been carried out in DL environments with a focus on the investigation of aspects associated with both the communication and the interaction of students in LMS (Chen & Looi, 2017; Machado, Lima, Maciel, & Rodrigues, 2016). The activities related to the communication and interaction of students in these virtual environments promote the Computer-Supported Collaborative Learning (CSCL) which, according to Bogarin, Cerezo, and Romero (2018), “is characterized by the sharing and construction of knowledge between participants using technology as their primary means of communication or as a common resource” (p. 10). From the computational point of view, several techniques have achieved promising results, among them Text Mining and Network Science deserve our special attention as they are related to the scope of this work.

2.1. Text Mining

According to Tan (1999), text mining (TM) extracts non-trivial patterns from texts. Inspired by data mining, which seeks to discover emerging patterns in structured data, TM aims to extract useful insights from unstructured or semi-structured data (Aranha & Passos, 2006). This process, however, can be accomplished through a variety of techniques, as in Li and Huang (2008), where a multidimensional analysis approach was proposed to investigate the contributions of students through a model that represents the texts in a vector space. In this technique, each document term becomes a dimensional characteristic. It can be seen, however, that the vector representation discards important information, such as the order in which the terms appear, where they appear, and the proximity between them (Schenker, 2003).

Another approach considers representing texts through graphs to overcome the aforementioned limitation. In this approach, the nodes of the graph generally represent the most frequent words whilst the associations between the nodes of the graph indicate the proximity between the words. In the context of graphical representations, we can highlight three interesting works. Chen, Kinshuk
Ethical Practice and Online Learning—A Contradiction? A Case Study
Donna Harper and Petra Luck (2009). Ethical Practices and Implications in Distance Learning (pp. 305-319).
www.igi-global.com/chapter/ethical-practice-online-learning-contradiction/18605?camid=4v1a