Short Text Classification Using Part of Speech Feature to Analyze Students' Feedback of Assessment Components

Authors : Zainab Mutlaq Ibrahim, Mohamed Bader-El-Den, Mihaela Cocea
Abstract : Students' textual feedback can hold unique patterns and useful information about learning process, it can hold information about advantages and disadvantages of teaching methods, assessment components, facilities, and other aspects of teaching. The results of analysing such a feedback can form a key point for institutions' decision makers to advance and update their systems accordingly. This paper proposes a data mining framework for analysing end of unit general textual feedback using part of speech feature (PoS) with four machine learning algorithms: support vector machines, decision tree, random forest, and naive bays. The proposed framework has two tasks: first, to use the above algorithms to build an optimal model that automatically classifies the whole data set into two subsets, one subset is tailored to assessment practices (assessment related), and the other one is the non-assessment related data. Second task to use the same algorithms to build an optimal model for whole data set, and the new data subsets to automatically detect their sentiment. The significance of this paper is to compare the performance of the above four algorithms using part of speech feature to the performance of the same algorithms using n-grams feature. The paper follows Knowledge Discovery and Data Mining (KDDM) framework to construct the classification and sentiment analysis models, which is understanding the assessment domain, cleaning and pre-processing the data set, selecting and running the data mining algorithm, interpreting mined patterns, and consolidating the discovered knowledge. The results of this paper experiments show that both models which used both features performed very well regarding first task. But regarding the second task, models that used part of speech feature has underperformed in comparison with models that used unigrams and bigrams.