Unsupervised Text Feature Extraction for Academic Chatbot using Constrained FP-Growth

ABSTRACT

In the edge where conversation merely involves online chatting and texting one another, an automated conversational agent is needed to support certain repetitive tasks such as providing FAQs, customer service and product recommendations. One of the key challenges is to identify and discover user's intention in a social conversation where the focus of our work in the academic domain. Our unsupervised text feature extraction method for Intent Pattern Discovery is developed by applying text features constraints to the FP-Growth technique. The academic corpus was developed using a chat messages dataset where the conversation between students and academicians regarding undergraduate and postgraduate queries were extracted as text features for our model. We experimented with our new Constrained Frequent Intent Pattern (cFIP) model in contrast with the N-gram model in terms of feature-vector size reduction, descriptive intent discovery, and analysis of cFIP Rules. Our findings show significant and descriptive intent patterns was discovered with confidence rules value of 0.9 for cFIP of 3-sequence. We report an average feature-vector size reduction of 76% compared to the Bigram model using both undergraduate and postgraduate conversation datasets. The usability testing results depicted overall user satisfaction average mean score is 4.30 out of 5 in using the Academic chatbot which supported our intent discovery cFIP approach.