

A Bayesian Approach for Analyzing Academic Article Structure

Authors : Jia-Lien Hsu, Chiung-Wen Chang

Abstract : Research articles may follow a simple and succinct structure of organizational patterns, called move. For example, considering extended abstracts, we observe that an extended abstract usually consists of five moves, including Background, Aim, Method, Results, and Conclusion. As another example, when publishing articles in PubMed, authors are encouraged to provide a structured abstract, which is an abstract with distinct and labeled sections (e.g., Introduction, Methods, Results, Discussions) for rapid comprehension. This paper introduces a method for computational analysis of move structures (i.e., Background-Purpose-Method-Result-Conclusion) in abstracts and introductions of research documents, instead of manually time-consuming and labor-intensive analysis process. In our approach, sentences in a given abstract and introduction are automatically analyzed and labeled with a specific move (i.e., B-P-M-R-C in this paper) to reveal various rhetorical status. As a result, it is expected that the automatic analytical tool for move structures will facilitate non-native speakers or novice writers to be aware of appropriate move structures and internalize relevant knowledge to improve their writing. In this paper, we propose a Bayesian approach to determine move tags for research articles. The approach consists of two phases, training phase and testing phase. In the training phase, we build a Bayesian model based on a couple of given initial patterns and the corpus, a subset of CiteSeerX. In the beginning, the priori probability of Bayesian model solely relies on initial patterns. Subsequently, with respect to the corpus, we process each document one by one: extract features, determine tags, and update the Bayesian model iteratively. In the testing phase, we compare our results with tags which are manually assigned by the experts. In our experiments, the promising accuracy of the proposed approach reaches 56%.

Keywords : academic English writing, assisted writing, move tag analysis, Bayesian approach

Conference Title : ICECSE 2016 : International Conference on Electrical and Computer Systems Engineering

Conference Location : Montreal, Canada

Conference Dates : May 16-17, 2016