Impact of Topic Modeling on Rule-Based Persian Metaphor Classification and its Frequency Estimation

Hadi Abdi Ghavidel
Language and Linguistics Center
Sharif University of Technology
Tehran, Iran
hadi_stlt@yahoo.com

Parvaneh Khosravizadeh
Language and Linguistics Center
Sharif University of Technology
Tehran, Iran
khosravizadeh@sharif.ir

Afshin Rahimi
University of Melbourne
Melbourne, Australia
arahimi@student.unimelb.edu.au

Received: January 12, 2014-Accepted: December 29, 2014

Abstract—**The impact of several topic modeling techniques have been well established in many various aspects of Persian language processing. In this paper, we choose to investigate the influence of Latent Dirichlet Allocation technique in the metaphor processing aspect and show this technique helps measure metaphor frequency effectively. In the first step, we apply LDA on Persian or so-called Bijankhan corpus to extract classes containing the words which share the most natural semantic proximity. Then, we develop a rule-based classifier for identifying natural and metaphorical sentences. The underlying assumption is that the classifier allocates a topic for each word in a sentence. If the overall topic of the sentence diverges from the topic of one of the words in the sentence, metaphoricity is detected. We run the classifier on whole the corpus and observed that roughly at least two and at most four sentence in the corpus carries metaphoricity. This classifier with an f-measure of 68.17% in a randomly 100 selected sentences promises that a LDA-based metaphoricty analysis seems efficient for Persian language processing.**

Keywords-Impact, LDA, Persian language, Metaphoricity