Abstract

Although, the fair amount of works in sentiment analysis (SA) and opinion mining (OM) systems in the last decade and with respect to the performance of these systems, but it still not desired performance, especially for morphologically-Rich Language (MRL) such as Arabic, due to the complexities and challenges exist in the nature of the languages itself. One of these challenges is the detection of idioms or proverbs phrases within the writer text or comment. An idiom or proverb is a form of speech or an expression that is peculiar to itself. Grammatically, it cannot be understood from the individual meanings of its elements and can yield different sentiment when treats as separate words. Consequently, in order to facilitate the task of detection and classification of lexical phrases for automated SA systems, this paper presents AIPSeLEX a novel idioms/ proverbs sentiment lexicon for modern standard Arabic (MSA) and colloquial. AIPSeLEX is manually collected and annotated at sentence level with semantic orientation (positive or negative). The efforts of manually building and annotating the lexicon are reported. Moreover, we build a classifier that extracts idioms and proverbs, phrases from text using n-gram and similarity measure methods. Finally, several experiments were carried out on various data, including Arabic tweets and Arabic microblogs (hotel reservation, product reviews,
and TV program comments) from publicly available Arabic online reviews websites (social media, blogs, forums, e-commerce web sites) to evaluate the coverage and accuracy of AIPSeLEX.

References

Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA), Atlanta, Georgia, 2013, pp. 55–64.

Index Terms

Computer Science  Information Sciences

Keywords

Sentiment lexicon  Idioms lexicon  AIPSeLEX.