

Cross-Lingual Opinion Mining using Semantic Features

Shima Esmaeili Taft

Azadeh Shakery

School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

ABSTRACT

Opinion mining or sentiment analysis is a subtask of text mining that analyzes the sentiment orientation of subjective documents. Both supervised and unsupervised methods have been proposed in the literature for this task. Supervised methods generally perform better than unsupervised methods, but they require a large set of labeled training dataset in the same domain and language of the test dataset. Creating a large training dataset is costly, and thus it is desired to make use of available datasets in one language to train the model in another one. Obviously using the available dataset directly won't have the desired result, so how to transfer the information from the source language to the target language is the challenge. In this paper we propose a cross-lingual opinion mining method which makes use of the available training data in one language to build a classifier and classify new documents in another language. To this end, a bilingual dictionary is used to overcome the language barrier, which is an available translation resource even in resource lean languages. The proposed method suggests dividing the features of both languages into two categories; pivot features and non-pivot features. Then using an unlabeled opinion dataset in both languages, a bipartite graph between these two categories of features is constructed. Bilingual semantic features are extracted by clustering this graph and documents in both languages are transferred into a unified semantic space. Experiment results on an English-German dataset show the significantly better performance of the proposed method compared to other cross-lingual methods.

Keywords: Opinion Mining, Sentiment Analysis, Pivot Feature, Semantic Feature, Cross-Lingual, Domain-Independent Feature, Domain-Specific Feature, Bipartite Graph, Classification.