

Dynamic Bag-of-Word Construction through Amazon Customer Reviews

Fuehres, H. Krauss, J. Nann, S. Gloor, P. Schoder, D.

Classic sentiment retrieval focuses on a particular domain like product or movie reviews and applies dictionary/bag-of-word methods. These approaches are static in regard to the structure of their bag-of-word and the domain being analyzed. This project's goal is to construct a dynamic bag-of-word (DBoW) for any given context through utilizing Amazon's product categories as a taxonomy and Amazon's customer reviews as a training dataset. By applying social network analysis metrics, relevance levels of individual customer reviews will be considered in order to achieve higher DBoW quality (e.g. helpful reviews will be considered as more relevant). This will lead to a holistic sentiment model suitable for classification of documents from any given domain.

The approach consists of two steps: first, a given word or phrase, representing the context/subject domain, is classified based on Amazon's product taxonomy. Fuzziness is applied to find all potential matches. In case of multiple matches the match with the highest degree of similarity will be chosen, or manual selection will be applied. Second step is the construction of a dictionary consisting of positive and negative words/phrases specifically adapted to the identified domain. Construction uses positive and negative customer reviews for products from the domain as training dataset. The social network of customers will provide clues about a review's individual impact for the DBoW construction. Result is a domain-specific DBoW suitable for classifying documents from the subject domain as positive or negative