Domain Sentiment Knowledge Discovery

Research Student
Andrius Mudinas

Supervisors
Dell Zhang
Mark Levene

Department of Computer Science and Information Systems

Research Aims

Our research is motivated by the problems facing cross domain sentiment detection, its temporal evolution and dimensionality. Everyday a large number of opinion related documents are put on the Internet - people post product reviews, express their political views, and share their feelings. The ability to extract sentiments from such sources can provide invaluable information about people's views on various topics and even help predicting stock market movements.

To achieve higher level of accuracy sentiment analysis is typically performed on a specific domain. However once domain boundaries crossed, performance suffers - as example text from financial news substantially differ from Amazon reviews or Twitter messages - same words can have completely different sentiment values. Creating domain specific system can be expensive and time consuming. To address these challenges, we are developing an **automatic domain knowledge discovery** and **sentiment detection** system.

Research Methodology

We designed a **concept-level sentiment analysis system** that seamlessly integrates lexicon-based and learning-based approaches. It can work in both fully supervised or unsupervised modes (see figure 1). In unsupervised mode using shallow word embedding model Word2Vec it can discover target domain's linguistic context and automatically learn sentiment words. By integrating temporal and multidimensional sentiment analysis it can further enhance sentiment detection performance and get better insight into expressed opinion.

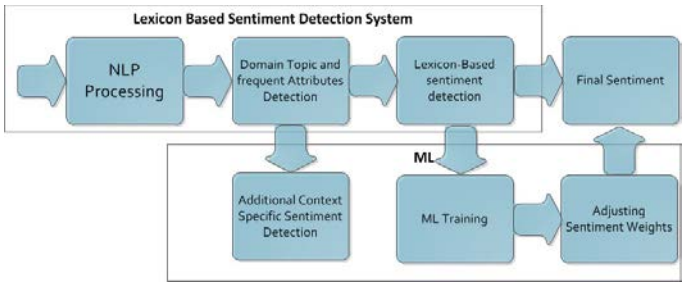

Figure 1. Software Architecture

Research Approach

We have developed a sentiment analysis tool *pSenti* that supports all steps of this process, and have conducted a preliminary evaluation of our methodology through a case study in various domains – **Amazon** Reviews, **Twitter** messages and **Medical** datasets.

Results so far demonstrated the potential benefits in terms of increased sentiment detection accuracy and the ability to extract domain specific information. We found that, words form distinct clusters in word embeddings vector space (fig. 2) and that information can be exploited to discover domain sentiment knowledge.

We are currently carrying out a more extensive evaluation using various datasets, aiming to demonstrate that this method could be universally applied to any target domain.

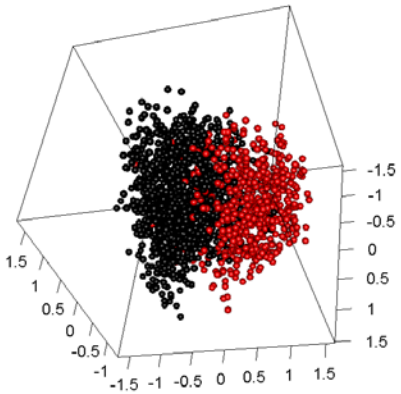

Figure 2. Sentiment Word Clusters

Publications

Andrius Mudinas, Dell Zhang, and Mark Levene. 2012. Combining lexicon and learning based approaches for concept-level sentiment analysis. Proceedings of WISDOM Workshop 2012.