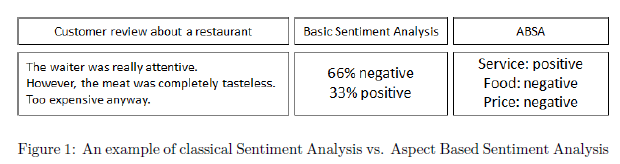
**Almost Unsupervised System for Aspect Based Sentiment Analysis**

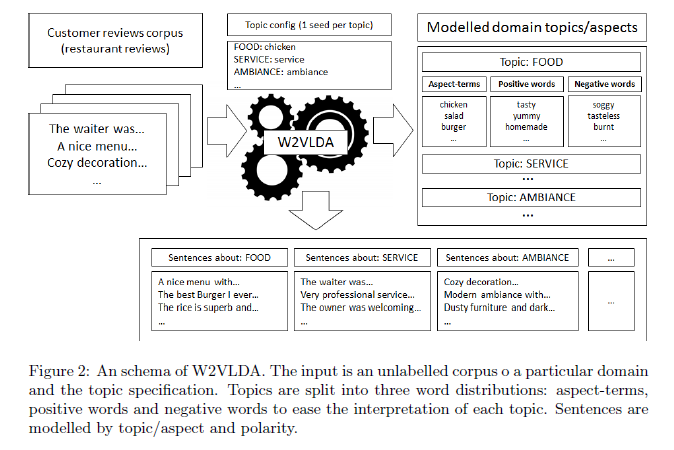
1. Introduction

Opinion Mining is a subfield of Natural Language Processing (NLP) that deals with the automatic analysis of opinions shared by humans in different contexts, like in customer reviews.

An ABSA system should be capable of classifying each opinion according to the aspects relevant for each domain in addition to classifying its sentiment polarity (usually positive, negative or neutral).



Our system performs at a word level three subtasks simultaneously: aspect classification, aspect-term/opinion-word separation, and sentiment polarity classification.



The system is based on a topic modelling approach combined with continuous

word embeddings and a Maximum Entropy classifier.

A well-known unsupervised method for text modelling documents is Latent Dirichlet Allocation (LDA).

2. Related work

In summary, combining topic modelling, continuous word embeddings and a minimal topic definition, our proposed approach can model customer reviews in different languages and domains performing three subtasks at the same time: aspect classification, sentiment classification and aspect terms/opinion-words separation. To our knowledge, no other almost unsupervised system tries to perform these three tasks at the same time and without requiring any pre-existing language or domain dependent resource.

1. System description

The main objective of the W2VLDA system is to perform the three tasks (detecting aspects, opinions and their polarity) of Aspect Based Sentiment Analysis at the same time. That is, to classify pieces of text into a predefined set of domain aspects and classify their sentiment polarity as positive or negative.

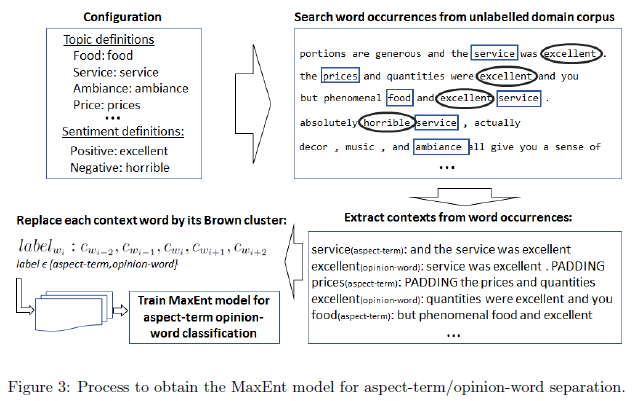
3.1. Topics and sentiment configuration

3.2. Aspect-term and opinion-word separation

Part of the outcome of the system consists of the aspect-term/opinion word

separation into differentiated word classes. In order to achieve this separation

without adding any language dependent tool or resource, the system uses Brown clusters to model examples of aspect-terms and opinion-words and train a MaxEnt-based classification model.



* 1. Combining everything in a topic model

The core of the system consists of an LDA-based topic model, extended to include the aspect-term/opinion-word separation and the positive/negative separation for each topic.

Word embeddings are a very popular way of representing words as the input for a

variety of machine learning techniques and are known for encoding interesting

syntactic and semantic properties

1. Evaluation

4.1. Resources and experimental setting

For topic classification we use the dataset from (Ganu et al., 2009) which contains

restaurant reviews labelled with domain-related categories (e.g. food, staff, ambience) for English. For sentiment classification, we use the Laptops and

DIGITAL-SLR dataset (Jo and Oh, 2011), consisting of English reviews of

electronic products with their corresponding 5-star rating.

4.2. Comparison with other LDA based approaches

4.3. Multilingual evaluation on SemEval2016 dataset

4.4. Assessing the seed words impact

Since the proposed approach heavily relies on the seed words (i.e. seeds

words are the only source of supervision to guide the algorithm to the desired

goal), it is interesting to evaluate the impact of different seed words and their

combination.

4.5. Aspect-term/Opinion-word separation evaluation

5. Conclusions and future work

In this document, we have presented W2VLDA, a system that performs

aspect and sentiment classification with almost no supervision and without the need of language or domain specific resources. In order to do that, the

system combines different unsupervised approaches, like word embeddings or

Latent Dirichlet Allocation (LDA), to bootstrap information from a domain

Corpus.

The only supervision required by the user is a single seed word per desired aspect and polarity. Because of that, the system can be applied to datasets of different languages and domains with almost no adaptation. The resulting topics and polarities are directly paired with the aspect names selected by the user at the beginning, so the output can be used to perform Aspect Based Sentiment Analysis. In addition, the system tries to separate automatically aspect terms and opinion words, providing more clear information and insight to the resulting domain aspects vocabulary. We evaluate W2VLDA for aspect classification using customer reviews of several domains and compare it against other LDA-based approaches. We also evaluate its performance using a subset of the multilingual SemEval 2016 task 5 ABSA dataset. As future work, it would we interesting to include an automated way to deal with stop-words and other words that do not carry information for the ABSA task. A better-integrated handling of multi-word and negation expressions could also improve the results. On the other hand, the are more specialised word embeddings related to sentiment analysis (Rothe et al., 2016), and it would be interesting to study if different word embeddings bring improvements to the method keeping a minimal supervision.