# Sentiment-Aspect Extraction based on Restricted Boltzmann Machines

F. Linlin Wang[1] F. Kang Liu[2] F. Zhu Cao[1] F. Jun Zhao[2]
F. Gerard de Melo[1]

[1]Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China

[2]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

September 2, 2015

# Outline


1. Motivation
2. Previous works
3. Pre-processing
4. Model
   - Basic Setup
5. Dataset
6. Model
   - Results on aspect words
   - Results on sentiment words
7. Conclusion

# Motivation

## Demand of automated sentiment and aspect words selection

There amount of reviews and comments on various social platforms is huge. For efficient sentiment analysis we need to identify the sentiment and aspect words efficiently.

## Review analysis

The e-commerce sites need to extract informations from large amount of reviews available on their products.

# Previous works

- Rule based
- Supervised learning methods like SVMs and CRFs
- Unsupervised methods like LDA

# Pre-processing

- Part of speech tagging
- Tf-Idf
- Removing stop words
- Assigning Prior probabilities using Tf-Idf,LDA and SentiWordNet

# Outline


1. Motivation
2. Previous works
3. Pre-processing
4. Model
   - Basic Setup
5. Dataset
6. Model
   - Results on aspect words
   - Results on sentiment words
7. Conclusion

# Basic Setup

- Created a variance of RBM called novel Sentiment-Aspect Extraction RBM
- The hidden units of the above RBM contains heterogeneous nodes instead of homogeneous like in standard RBMs
- Separate kind of hidden nodes for aspects, sentiments and background
- Finally, stochastic gradient descent to find the optimum parametres

a restaurant review dataset, which contains 1,644,923 tokens and 52,574 documents in total.

Documents in this dataset are annotated with one or more labels from a gold standard label set S = Food, Staff, Ambi-ence, Price, Anecdote, Miscellaneous.

# Outline


1. Motivation
2. Previous works
3. Pre-processing
4. Model
   - Basic Setup
5. Dataset
6. Model
   - Results on aspect words
   - Results on sentiment words
7. Conclusion

# Results on aspect words

- Three major aspects (food, staff and ambience) were used to test the accuracy of the model
- For food precision, recall and F1 score were, respectively, 0.891, 0.854 and 0.872
- For staff and ambience the above stated scores were 0.819, 0.582, 0.680, 0.805, 0.592 and 0.682 respectively

# Outline


1. Motivation
2. Previous works
3. Pre-processing
4. Model
   - Basic Setup
5. Dataset
6. Model
   - Results on aspect words
   - Results on sentiment words
7. Conclusion

# Results on sentiment words

- The model was also used to check sentiment classification accuracy.
- The model scored 0.788 versus SentiWordnet's 0.703.

Extraction of RBM model to jointly extract review aspects and sentiment polarities in an unsupervised dataset.

Outperforms LDA.