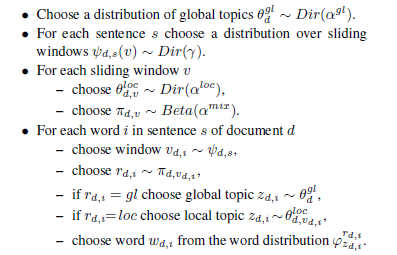
A Joint Model of Text and Aspect Ratings for Sentiment Summarization  
1 Introduction

Instead, we propose an unsupervised model that leverages aspect ratings that frequently accompany an online review. First, ratable aspects normally represent coherent topics which can be potentially discovered from co-occurrence information in the text. Second, we hypothesize that the most predictive features of an aspect rating are features derived from the text segments discussing the corresponding aspect

2 The Model

In this section we describe a new statistical model called the Multi-Aspect Sentiment model (MAS), which consists of two parts. The first part is based on Multi-Grain Latent Dirichlet Allocation (Titov and McDonald, 2008), which has been previously shown to build topics that are representative of ratable aspects. The second part is a set of sentiment predictors per aspect that are designed to force specific topics in the model to be directly correlated with a particular aspect.  
2.1 Multi-Grain LDA

MG-LDA models two distinct types of topics: global topics and local topics. As in LDA, the distribution of global topics is fixed for a document (a user review). However, the distribution of local topics is allowed to vary across the document.



2.2 Multi-Aspect Sentiment Model

Thus, if we model the prediction of aspect ratings jointly with the construction of explicitly associated topics, then such a model should benefit from both higher quality topics and a direct assignment from topics to aspects. This is the basic idea behind the Multi-Aspect Sentiment model (MAS).

The distribution of a sentiment rating ya for each rated aspect a is computed from two scores. The first score is computed on the basis of all the ngrams, but using a common set of weights independent of the aspect a. Another score is computed only using n-grams associated with the related topic, but an aspect-specific set of weights is used in this computation. More formally, we consider the log-linear distribution:



2.3 Inference in MAS

3 Experiments

3.1 Qualitative Evaluation

3.2 Sentence Labeling

A primary advantage of MAS over unsupervised models, such as MG-LDA or clustering, is that topics are linked to a rated aspect, i.e., we know exactly which topics model which aspects.

4 Related Work

5 Conclusions

In this paper we presented a joint model of text and aspect ratings for extracting text to be displayed in sentiment summaries. The model uses aspect ratings to discover the corresponding topics and can thus extract fragments of text discussing these aspects without the need of annotated data. We demonstrated that the model indeed discovers corresponding coherent topics and achieves accuracy in sentence labeling comparable to a standard supervised model. The primary area of future work is to incorporate the

model into an end-to-end sentiment summarization system in order to evaluate it at that level.