Aspect Extraction with Automated Prior Knowledge Learning

1 Introduction

Learning quality knowledge:

Extraction guided by learned knowledge:

In summary, this paper makes the following contributions:

1. It proposes to exploit the big data to learn prior knowledge and leverage the knowledge in topic models to extract more coherent aspects.

2. It proposes an effective method to learn quality knowledge from raw topics produced using review corpora from many different domains.

3. It proposes a new inference mechanism for topic modeling, which can handle incorrect knowledge in aspect extraction.

2 Related Work

Aspect extraction has been studied by many researchers in sentiment analysis using supervised sequence labeling or classification and using word frequency and syntactic patterns.

3 Overall Algorithm

This section introduces the proposed overall algorithm. It consists of two main steps: learning quality knowledge and using the learned knowledge.

4 Learning Quality Knowledge

This section details Step 1 in the overall algorithm, which has three sub-steps: running LDA (or AKL) on each domain corpus, clustering the resulting topics, and mining frequent patterns from the topics in each cluster.

4.1 Topic Clustering

After running LDA (or AKL) on each domain corpus, a set of topics is obtained.

4.2 Frequent Pattern Mining

4.3 Knowledge Representation

5 AKL: Using the Learned Knowledge

5.1 Plate Notation

In short, our modeling contributions are (1) the capability of handling more expressive knowledge in the form of clusters, (2) a novel Gibbs sampler to deal with inappropriate knowledge.

5.2 The Gibbs Sampler

6 Experiments

6.1 Experimental Settings

6.2 Topic Coherence

6.3 User Evaluation

6.4 Sensitivity to Clustering Parameters

6.5 Test on New Domains

7 Conclusions

This paper proposed an advanced aspect extraction framework which can learn knowledge automatically from a large number of review corpora and exploit the learned knowledge in extracting more coherent aspects. It first proposed a technique to learn knowledge automatically by clustering and FPM. Then a new topic model with an advanced inference mechanism was proposed to exploit the learned knowledge in a fault-tolerant manner. Experimental results using review corpora from 36 domains showed that the proposed method outperforms state-of-the-art methods significantly.

