**Aspect and Entity Extraction for Opinion Mining**

**1 Introduction**

With the growth of *social media* (i.e., reviews, forum discussions, and blogs) on the Web, individuals and organizations are increasingly using the opinions in these media for decision making. Automated opinion mining is thus needed, as subjective biases and mental limitations can be overcome with an objective opinion mining system.

Basically, researchers have studied opinion mining at three levels of granularity, namely, document level, sentence level, and aspect level.

For example, in a product review, the reviewer usually writes both positive and negative aspects of the product, although the general sentiment on the product could be positive or negative. To obtain more fine-grained opinion analysis, we need to delve into the aspect level.

**2 Aspect-Based Opinion Mining Model**

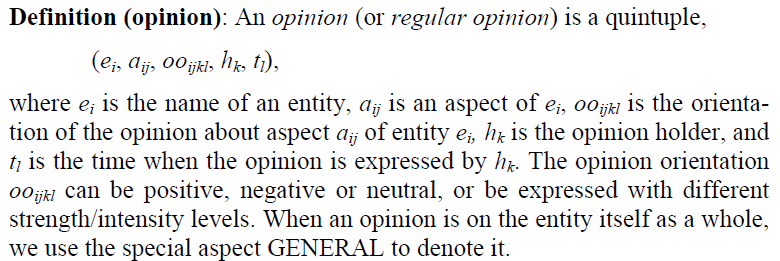
***2.1 Model Concepts***

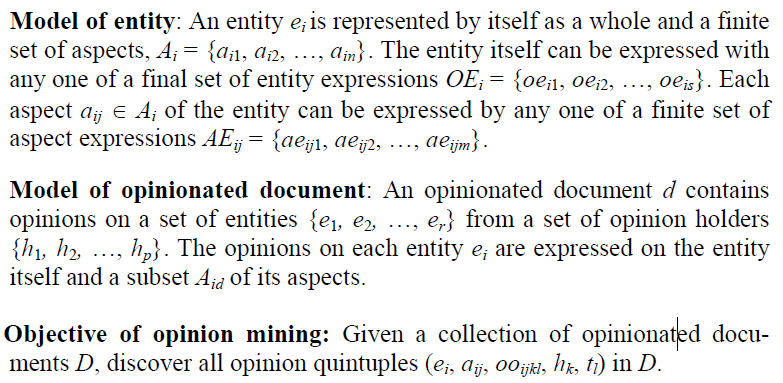
**Definition** (**entity**)**:** An *entity e* is a product, service, person, event, organization, or topic.

**Definition** (**aspect and aspect expression**)**:** The *aspects* of an entity *e* are the components and attributes of *e*. An *aspect expression* is an actual word or phrase that has appeared in text indicating an aspect.

**Definition** (**entity expression**)**:** An *entity expression* is an actual word or phrase that has appeared in text indicating a particular entity.

**Definition** (**opinion holder**)**:** The *holder* of an opinion is the person or organization that expresses the opinion.





***2.2 Aspect-Based Opinion Summary***

**3 Aspect Extraction**

**Format 1** − **Pros, Cons and the detailed review:** The reviewer is asked to describe some brief Pros and Cons separately and also write a de-tailed/full review.

**Format 2** − **Free format:** The reviewer can write freely, i.e., no separation of pros and cons.

***3.1 Extraction Approaches***

There are two types of aspect expressions in opinion documents: *explicit aspect expression* and *implicit aspect expression*.

**3.1.1 Exploiting Language Rules**

Hu and Liu (2004a) first proposed a method to extract product aspects based on association rules. The idea can be summarized briefly by two points: (1) finding frequent nouns and noun phrases as frequent aspects. (2) using relations between aspects and opinion words to identify infrequent aspects. The basic steps of the approach are as follows.

**Step 1**: Find frequent nouns and noun phrases.

**Step 2**: Find infrequent aspects by exploiting the relationships between aspects and opinion words.

**3.1.2 Sequence Models**

**Hidden Markov Model**

**Conditional Random Fields**

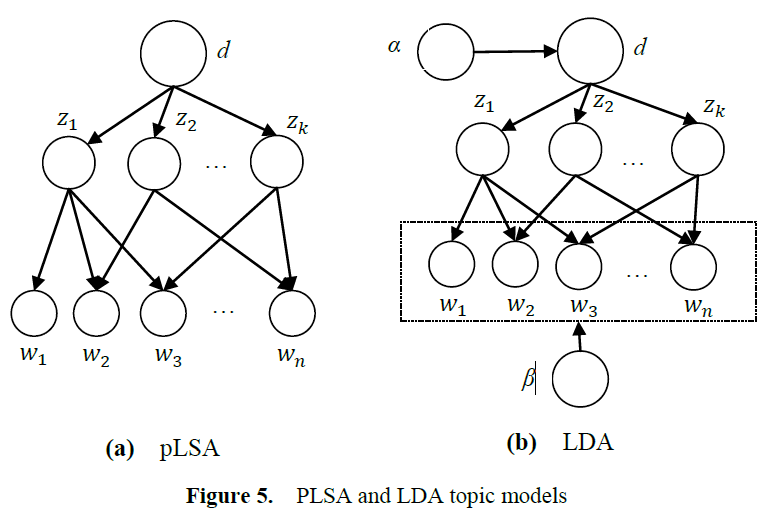
**Token**: This feature represents the string of the current token.

**Part of Speech**: This feature represents the POS tag of the current token. It can provide some means of lexical disambiguation.

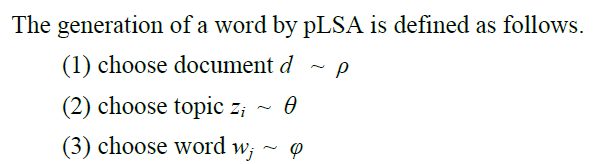
**Short Dependency Path**: Direct dependency realtions show accurate connections between a target and an opinion expression. Thus, all tokens which have a direct dependency relation to an opinion expression in a sentence are labelled.

**Word Distance**: Noun phrases are good candidates for opinion targets in product reviews. Thus token(s) in the closest noun phrase regarding word distance to each opinion expression in a sentence are labelled.

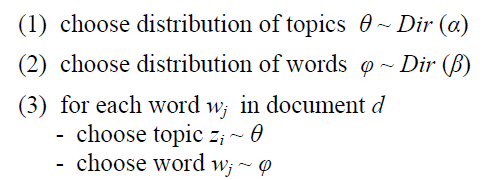
**3.1.3 Topic Models**



**Probabilistic Latent Semantic Analysis**



**Latent Dirichlet Allocation (LDA)**



**3.1.4 Miscellaneous Methods**

***3.2 Aspect Grouping and Hierarchy***

***3.3 Aspect Ranking***

A product may have hundreds of aspects. Sometimes, we need to identify important one from reviews, which are more influential for people’s decision making.

***3.4 Mapping Implicit Aspect Expressions***

***3.5 Identifying Aspects that Imply Opinions***

**Step 1:** *Candidate Identification*:

**Step 2:** *Pruning*:

***3.6 Identifying Resource Noun***

**4 Entity Extraction**

***4.1 Extraction Methods***

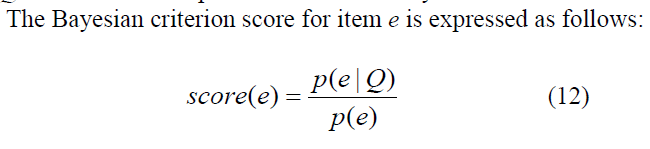
**4.1.1 PU Learning**

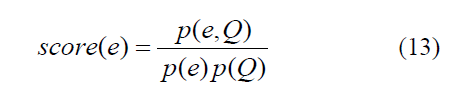
**Generating candidate entities**

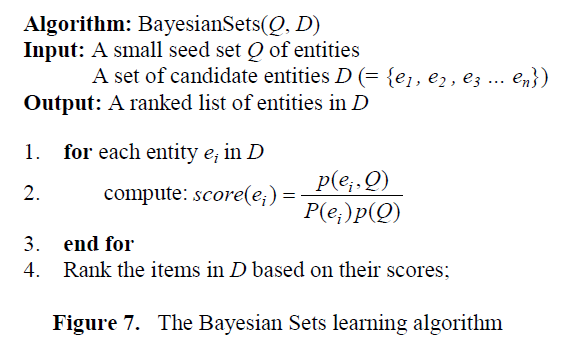
**Generating positive and unlabeled sets**

**Ranking entity candidates**

**4.1.2 Bayesian Sets**







**5 Summary**

With the explosive growth of social media on the Web, organizations are increasingly relying on opinion mining methods to analyze the content of these media for their decision making. Aspect-based opinion mining, which aims to obtain detailed information about opinions, has attracted a great of deal of attention from both the research community and industry.

For aspect extraction, existing solutions can be grouped into three main categories:

1. using language dependency rules,
2. using sequence learning algorithms such as HMM and CRF
3. using topic models

The current F-1 score results range from 0.60 to 0.85 depending on domains and datasets.