

# iMovie: A Prototype Sentiment Search Engine for Movie Reviews

Dongling Chen

School of Information Engineering  
Shenyang University  
Shenyang, 110044, P.R.China  
e-mail: 13504183184@163.com

Feng Yang

School of Information Engineering  
Northeastern University  
Shenyang 110004, P.R.China  
e-mail: syucdl@msn.com

**Abstract**—In this paper, we present our system iMovie to provide an efficient sentiment search tool for online movie reviews. iMovie implements a ontology tree model to represent the complex relationship between features and opinion words in movie reviews, which guarantees system's accuracy. iMovie integrates flexible fine-grained sentiment search interface as well as relation search functions. The internal mechanism and the usage of iMovie are illustrated in this paper.

**Keywords**- Sentiment Search; Opinion Mining; Search Engine; Sentiment Analyzer; Mining Sentiment

## I. INTRODUCTION

With the rapid development of Web 2.0 technology, the Web has become the largest public data source in the world. More and more people are willing to express their personal opinions and attitudes on review websites, such as Amazon (<http://www.amazon.com>) and IMDB (<http://www.imdb.com>). The online opinions have become the most important credit evaluation for e-commerce and personal guidelines. A lot of research work has been published for mining and summarizing opinions in product reviews [1,2,4,5]. However, there are still several challenges to be tackled for searching the sentiments in movie review domain.

- (1) People prefer to talk about more scattered topics and use more flexible styles when writing movie reviews. Therefore, the commented features in movie review are much richer than those in product review [3,5].
- (2) Most existing opinion mining prototype systems focus on summarizing people's opinions in the reviews[6,7]. However, there are huge amount of movie reviews and thousands of actors which are only accessible through search engines when specified keyword is given.
- (3) The sentiments in reviews contain both feature targets and orientations. The traditional topic-based search engines usually provide single-dimension keyword based interface, which could not meet the need of multi-dimensions sentiment search.

To tackle these challenges, in this demonstration we present iMovie, a prototype sentiment search engine for online movie reviews. iMovie has three novel features.

First of all, iMovie utilizes a semi-supervised ontology tree model to represent the complex relationship between the features and opinion words in movie reviews. The proposed ontology tree model can ensure the accuracy for mining the sentiments in the reviews. Secondly, both the full review texts and the summarization results are indexed in iMovie for sentiment retrieval. Thirdly, iMovie not only provides a fine-grained sentiment search interface based on keywords and polarities, but also integrates other components such as actor relation search and similar movie search, which together constitute a new multi-dimensions sentiment search engines. In the next sections, we will further illustrate the system architecture and interface demonstration of iMovie.

## II. SYSTEM OVERVIEW

The architecture of iMovie is shown in Fig .1 Generally, iMovie is composed of data store layer, search interface, and several system core functional components.

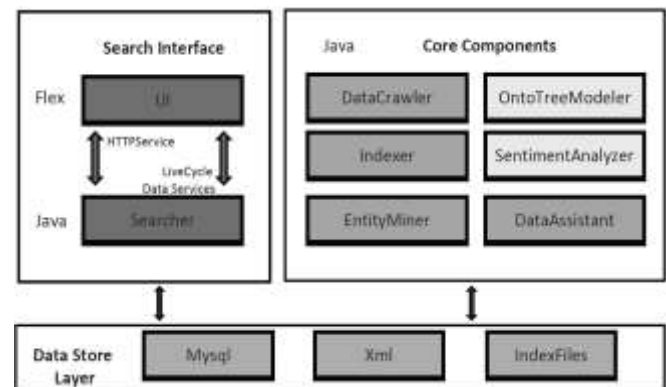


Figure 1. System Architecture of iMovie

iMovie utilizes Flex technology to provide a convenient and friendly user interface for searching the sentiments in movie reviews. The DataCrawler component can traverse the entire website and extract raw review texts as well as the structured data such as movie's title, directors and actors when crawling. Lucene is used to index the crawled data in Indexer. The EntityMiner component focuses on mining the relationship between the movies and actors. We design a DataAssistant component to provide a unified data access tool for different kinds of processed data. The extracted and indexed data are stored using MySQL, XML and index files. We give a brief

introduction to the implementation of Onto Tree Modeler and Sentiment Analyzer component here.

**The topic model description.** The topic is usually the central idea of an article, a word, a sentence. In general the user comments are based on a particular topic. In the car shopping guide network, user comments refer to the theme of "cars". in the movie review website, user comments refer to the theme of "film"; in the electronic commerce website, user comments is specific to the purchase of goods.

Usually a theme has a number of attributes, where each attribute description consists of a plurality of feature words. At the same time, the user comment is the description of the subject specific to one or more attributes. For example, the user can be the "car" as the theme of "price" attribute to comment; the user can "film" on the topic of "music" property comment; users can "commodity" the theme of "quality" attribute to comment.

At the same time, emotion tendentiousness of each attribute was described by some emotional words. According to the Chinese word collocations, general different properties will correspond to different emotional words. For example, users can use the "great music" to express appreciation for music; "superlative" to express the actor's approval; with "reliable quality" to express the satisfaction of commodity quality.

Through the study of the above rules, emotion tendentiousness topic model is selected in this paper to analyze user comments. Next, this paper use the definition of 3.1 to define each concept 3.7 to detail the topics in the model.

**Definition 3.1 (Theme) theme:** by name, attribute set top concept consisting of a theme T. T by the topic name, attribute set composition, said method for  $T = \langle tn, A \rangle$ . Where TN is the topic name, A for the attribute collection.

**Definition 3.2 (set of attributes):** set consists of the theme of the attributes for the property sets A, A by several attributes, the representation method for  $A = \{a_1, a_2, \dots, a_k\}$ , where a is the attribute (Attribute).

**Definition 3.3 (property):** describe the concept of the theme of a feature for attribute a. A by the attribute name, set of feature words, the emotional word set and the characteristics of emotional words set composition, said method for  $\langle aname, FW, OW, FOWP \rangle$ . Where aname is the attribute name, FW is the set of feature words, OW is the emotional word sets, FOWP is characteristic of emotional words collection.

**Definition 3.4 (set of feature words):** describe the characteristics of theme words set of properties for a set of feature words FW, FW composed of a plurality of nouns and noun phrases composition, said method for  $FW = \{fw_1, fw_2, \dots, fw_i\}$ , where FW is the character, word (Feature Word).

**Definition 3.5 (emotional word set):** describes the theme of some attribute emotional tendency to emotion word set for emotional words set OW, OW is composed of a plurality of emotional words, said method for  $OW = \{ow_1, ow_2, \dots, ow_j\}$ , where ow is the emotional words (Opinion Word).

**Definition 3.6 (characteristics of emotional words to set):** consists of a set of feature sentiment words on the set of characteristics of emotion for the collection of FOWP, FOWP by a number of characteristics of emotional words

on the composition, said method for  $FOWP = \{fop_1, fop_2, \dots, fop_k\}$ , fowp, which is characterized of emotional words (Feature Opinion Word Pair).

**Definition 3.7 (characteristics of emotional words):** appear in the same sentence feature words and emotional words as features of emotional words on fowp, said method for  $\langle aname, FW, ow, count \rangle$ , where aname is the attribute name, FW as the feature words, ow is the emotional words, count is the number of occurrences.

**Storage model of theme.** The topic model uses XML files to store constructed. The contents of the DTD file of the XML file as shown in Fig .2.

```
<?xml version="1.0" encoding="UTF-8" ?>
<topicDictionary>
  <topicModel name="电影">
    <attributes>
      + <attribute name="导演">
        - <attribute name="剧情">
          + <featureWords>
            + <opinionWords>
          - <wordPairs>
            <wordPair featureWord="内容" opinionWord="深" count="1" />
            <wordPair featureWord="剧情" opinionWord="淋漓尽致" count="1" />
            <wordPair featureWord="剧情" opinionWord="简明" count="2" />
            <wordPair featureWord="剧情" opinionWord="扣人心弦" count="1" />
            <wordPair featureWord="剧情" opinionWord="紧张刺激" count="1" />
            <wordPair featureWord="结尾" opinionWord="精彩" count="3" />
            <wordPair featureWord="内容" opinionWord="逻辑性强" count="1" />
            <wordPair featureWord="人物" opinionWord="丰满" count="1" />
            <wordPair featureWord="故事" opinionWord="生动" count="1" />
            <wordPair featureWord="故事" opinionWord="感人" count="2" />
            <wordPair featureWord="剧本" opinionWord="弱" count="1" />
            <wordPair featureWord="剧情" opinionWord="张力十足" count="1" />
            <wordPair featureWord="故事" opinionWord="紧凑" count="1" />
            <wordPair featureWord="段落" opinionWord="出彩" count="1" />
            <wordPair featureWord="故事" opinionWord="简单" count="4" />
          
```

Figure 2. an example of topic model

**Sentiment analysis framework based on topic model.**

Based on the analysis of sentiment topic model, first of all need to be divided into different categories of comments attribute this theme, such as several major attribute of the film by the director, actor, plot, music etc.. Therefore, attribute feature words, emotional words in this task is the focus of analysis. The characteristics of the word contained in comments can be divided to comment on the specific property, and includes comments in the emotional words can be divided into 3 different comments of the sentiment polarity category.

Before a commentary on the emotional tendency analysis, have a look first to feature words - Review sentiment word representation. For a review, it may contain multiple character words and a plurality of emotional words. The part between feature words and emotional words with matching relationship of language model, this paper defines it as feature - affective words on; then there is nothing between partial feature words and emotional words matching relationship. At the same time in the analysis, the need to extract these free feature words and emotional words. The features of the word free to assist the comment attribute classification, free of the emotional words to assist the review sentiment orientation calculation. Fig .3 shows a comment may contain features of emotional words, free feature words, free emotional words.

**OntoTreeModeler.** As we have discussed above, there are more commented features in movie reviews than in product reviews. Therefore, it is difficult to manually define the total hierarchical relationships of the features in movie reviews. In this component, we utilize a ontology

tree model to represent the underlying relationships between the features and opinions in movie reviews, as shown in Fig . 3.

The ontology tree model for movie reviews has three layers[2], i.e. movie root, aspects and feature-opinion layer, which is learned in a semi-supervised way. For learning the model, firstly we define several concrete aspects such as Music, Story, Acting and so on. Secondly, we manually label the features, opinion words and their pairs in review training set. Then, the language patterns are learned from these labels, which could increase the coverage for

extracting the feature-opinion word pairs. With the help of a semantic knowledge base, finally we can form the ontology tree model for the sentiments in movie reviews.

**Sentiment Analyzer.** Based on the ontology tree model, the SentimentAnalyzer component extracts the features, opinion words and their pairs in reviews to calculate the fine-grained aspect level sentiment orientations[8,9,10]. When calculating the overall sentiments of a review, the component also considers the isolated features and opinion words, i.e. the items that are not defined in the ontology tree model.

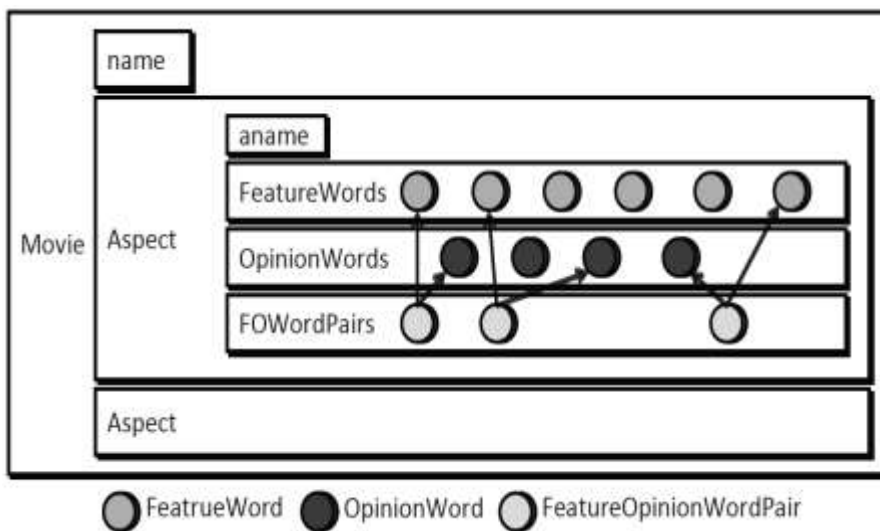


Figure 3. The ontology tree model for the sentiments in movie reviews



Figure 4. The sentiment search interface of iMovie

### III. SYSTEM DEMONSTRATION

The demonstration of iMovie includes the following two parts:

**System internals.** We will show how the review data is crawled, parsed and indexed from the biggest Chinese

movie review website Douban ([http:// www.douban.com](http://www.douban.com)). We will illustrate the working mechanism of the review parser and the building process of the ontology tree model. The labeling system for extracting the feature-opinion word pairs will be shown on site. We will also demonstrate



IEEE/WIC/ACM International Conference on Web Intelligence (WI-2013), 2013.

- [10] Z. Chen, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Exploiting Domain Knowledge in Aspect

Extraction. Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2013), October 18-21, 2013, Seattle, USA.