Fine-grained Sentiment Analysis of Product Comments Based on Product Features

Zheng Weihong

BUPT No.: 2013213217

QM No.: 130801991

Project Introduction

- Product comments reflect the overall attitude of buyers towards to its different features (quality, price, etc.)
- Customers and vendors like to know better about different aspects of a product.
- The huge amount of comments brings difficulty to the assessment on a product.

So, the objective is :

Provide a technical way to analyze buyers' sentiment to each feature of a specific product.

Project Workflow

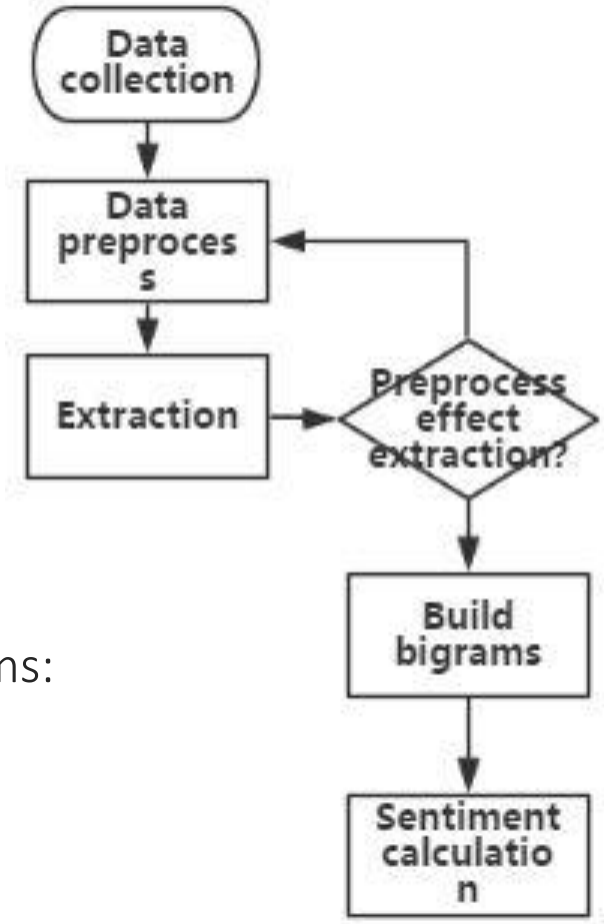
Collect data from e-commerce website

Preprocess the data:
- word segmentation
- eliminating stop words
- Part-of-speech tagging

Features and opinion words extraction
- Morpheme-based method
- PMI method

Building (“feature”, “opinion word”) bigrams:
- Eg. (“quality”, “good”), (“style”, “good”)

Sentiment calculation
- Sentiment classification
- Calculation method



Finished Work

Data collection:

- Crawler program based on scrapy and xpath
- 5000 comments from Taobao

Data preprocess:

- Jieba, Sogou thesaurus
- dynamic programming, HMM-based model with Viterbi algorithm
- HIT stop words dictionary

Preliminary feature and opinion words extraction:

- Feature: TF-IDF algorithm with manually-set dictionary, special tag “f”
- Opinion words: Most of them are adjectives with tag “a”

Attempt of building bigrams:

- NLTK, not very ideal

Problems and Solutions

Different expression of features

"the style is good" & "it looks good"

Solution: Group features into explicit and implicit group. For explicit features, the morpheme-based method can be used to extract similar feature words. And for implicit features, a method named PMI is used to find those features.

Reference: "Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis"

Building bigrams

NLTK, ("quality", "handsome")

Solution: Build up an opinion words dictionary for each feature, establish bigrams by the easy permutation and combination and then apply a filter to select diagrams.

Reference: "Study on Chinese Text Sentiment Classification"

Next step

Evaluate the feature and opinion words extraction and complete it

Build up bigrams

Sentiment calculation

Finish the final report