

talentica



Sentiment Analysis

Movie Reviews

Sentiment Analysis is the method of extracting subjective information from any written content. It is being widely used in product benchmarking, market intelligence and advertisement placement. In this paper, we have demonstrated the process by analyzing a movie review using various Natural Language Processing techniques.

A White Paper

Sentiment Analysis: Movie Reviews



Sentiment Analysis reveals the emotions, beliefs and feelings of the author on a particular topic. It uses natural language processing and machine learning techniques to effectively apply general patterns and determine the attitude expressed in the written text.

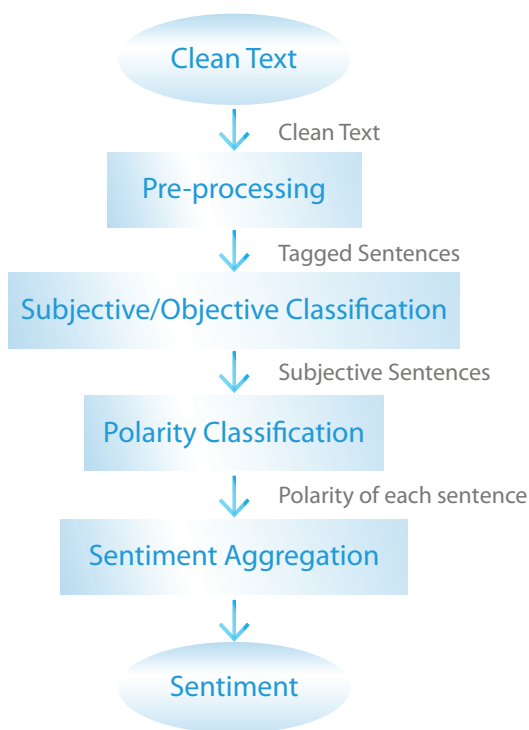
Sentiment Analysis has gained popularity in recent years due to its immediate applicability in business environment, such as summarizing feedback from the product reviews, discovering collaborative recommendations, or assisting in election campaigns.

In this paper, we describe a sentence level sentiment analysis system for movie reviews, which we built at Talentica Software.

Sentiment Analysis: Movie Reviews

Here's a review that we take as an example to explain how we went about analyzing the sentiment of movie reviews:

"Hands down, the best summer movie of 2011! This story is an origin story about how the Apes began to rise to power. The movie is intelligent, thought provoking, emotional, and damn well entertaining. The Weta team did a phenomenal job with their brilliant special effects. The only problem with this film is the acting. All the humans are cardboard clichés in this film."



The high level flow diagram of the NLP process.

Four Basic Steps

We followed a four step process to analyze the sentiment of each movie review:

- > Pre-processing and breaking the review into parts of speech
- > Identifying subjective and objective sentences
- > Classifying subjective sentences into those expressing positive and negative sentiments
- > Aggregating the overall sentiment

Sentiment Analysis: Movie Reviews

Steps of Sentiment Analysis:

- ❶ Pre-Processing
- ❷ Subjective/Objective Classification
- ❸ Polarity Classification
- ❹ Sentiment Aggregation

Pre-processing

As a first step we pre-processed the text to identify the various parts of speech, phrases and named entities.

POS Tagging

POS tagging is the process of assigning parts of speech such as noun, verb, adjective, adverb, etc to each word in a sentence. For example, we processed the following sentence - "And once again, the Weta team did a phenomenal job with their brilliant special effects", and tagged

- > "Weta", "team" & "job" as nouns
- > "did" as a verb
- > "phenomenal" & "brilliant" as adjectives
- > "again" as an adverb and
- > "with" as a preposition

Sentiment Analysis: Movie Reviews

Steps of Sentiment Analysis:

- 1 Pre-Processing
- 2 Subjective/Objective Classification
- 3 Polarity Classification
- 4 Sentiment Aggregation

Chunking

Chunking, also known as shallow parsing is the process of identifying phrases. So we split *"All the humans are cardboard clichés in this film"* into two noun chunks- *"all the humans"* and *"cardboard cliché in this film"*.

Named Entity Recognition

The process of NER, Named Entity Recognition, helps find named entities such as names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc. So 2011 is recognized as a date type in the following sentence, *"Hands down, the best summer movie of 2011"*.

We also applied Tokenizing, Sentence Splitting and Morphological Analysis in order to further clean the text.



Tagging sentences in Pre-Processing

Steps of Sentiment Analysis:

- 1 Pre-Processing
- 2 Subjective/Objective Classification
- 3 Polarity Classification
- 4 Sentiment Aggregation

Subjective/Objective classification

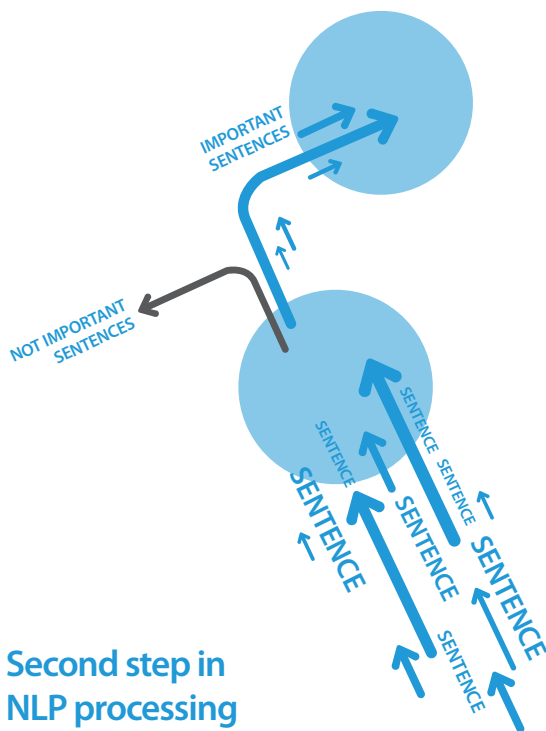
We then took the pre-processed clean text and classified the review into subjective (opinionated) and objective sentences.

Objective sentences don't play any role in calculating the sentiment of the text since they do not consist of sentiment-bearing words or phrases. For example the following sentence in the review is not opinionated thus will not play a role in determining the sentiment:

"This story is an origin story about how the Apes began to rise to power."

On the other hand, *"It's intelligent, thought provoking, emotional, and damn well entertaining"* definitely is opinionated/subjective and will be taken to the next level to determine whether the sentiment involved is positive or negative. So with this classification module we pruned all the not important sentences for future processing.

We used the Supervised Machine Learning technique for this classification task. In supervised classification, the classifier is trained on labeled examples that are similar to the test examples. We used different features of SVM training such as,



Second step in
NLP processing

Sentiment Analysis: Movie Reviews

Steps of Sentiment Analysis:

- ① Pre-Processing
- ② Subjective/Objective Classification
- ③ Polarity Classification
- ④ Sentiment Aggregation

Bag-of-Words

Bag-of-words is a model that takes individual words in a sentence as features. The text is represented as an unordered collection of words without taking grammar or order of words into consideration. Each word is conditionally independent from the others.

Presence of Number

We added a binary feature based on the presence or absence of the number in the sentence. Sentences that contain numbers generally are objective sentences.

Presence of Modal Verb

Presence of modal verbs like have to, must, can, should, wish, want, need play a major role in the classification of the subjective-objective sentences.

Presence of Polar Words

Polar words are the words which represent the sentiment like “good” and “bad”. A binary feature is added on the presence or absence of the polar word. Sentences which contain polar words generally are subjective sentences. Example: “Hands down, the best summer movie of 2011!”

Sentiment Analysis: Movie Reviews

Steps of Sentiment Analysis:

- ① Pre-Processing
- ② Subjective/Objective Classification
- ③ **Polarity Classification**
- ④ Sentiment Aggregation

Polarity Classification

In this step, we classified the subjective sentences identified into positive and negative sentences. Calculating the polarity of each sentence is very important to determine the overall sentiment. The input for this process comprised all the identified subjective sentences. For this task also we used different features of SVM training like,

Bag-of-words with Polarity

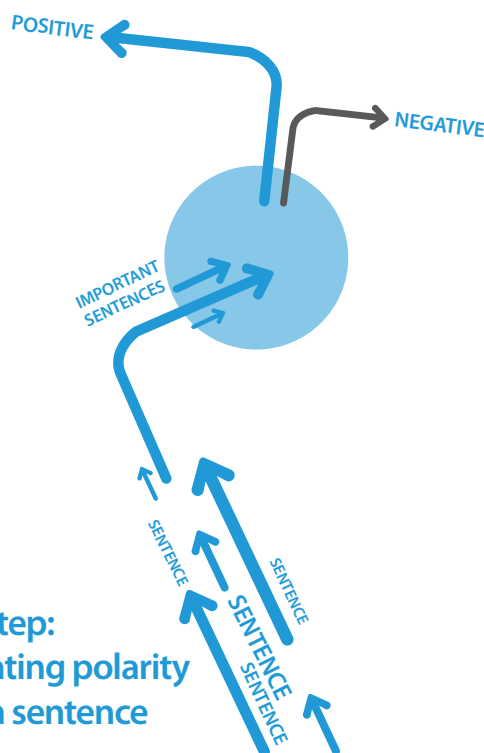
This is the same feature used in the last section. But we also added the polarity of the word. Polarity of word is calculated in the pre processing module. This is very important for handling negation. Details about this feature are given in the next section.

Negation Handling

Negation plays an important role in polarity analysis. One of the example sentences from our corpus –

"This is not a good movie" had the opposite polarity from the sentence *"This is a good movie"*, although the features of the original model would show that they were of the same polarity.

Third step:
Calculating polarity
of each sentence



Sentiment Analysis: Movie Reviews

Steps of Sentiment Analysis:

- 1 Pre-Processing
- 2 Subjective/Objective Classification
- 3 Polarity Classification
- 4 Sentiment Aggregation

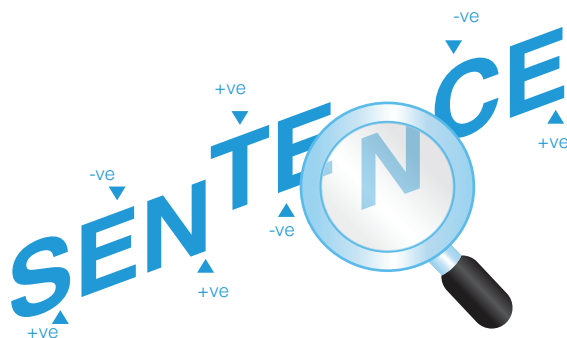
So in order to handle the word “good” in first and second sentences differently, we added the polarity of the word to it. The negation handling module changed the polarity of “good” to negative.

Positive Words Count

We calculated the number of positive words in the sentence and added it as a feature. This is a very important feature because if there are more positive words then the sentence tends to be a positive sentence. For example, “*It's intelligent, thought provoking, emotional, and damn well entertaining*” has four positive words so it is a positive sentence.

Negative Words Count

We also calculated the number of negative words present in the sentence and added it as a feature. For example, “*The only problem with this film is the acting*” has one negative word.



- ⦿ +ve polarity
- ⦿ -ve polarity

Identifying polarity of the word

Steps of Sentiment Analysis:

- ① Pre-Processing
- ② Subjective/Objective Classification
- ③ Polarity Classification
- ④ Sentiment Aggregation

Aggregating Sentiment from Sentences

This is the last step in the NLP processing. The main task of this step is to aggregate the overall sentiment of the movie review from the sentences which were tagged positive and negative in the previous step. We just used the simple aggregation scheme.

If the number of positive sentences was greater than the number of negative sentences, then we considered the overall sentiment of the review to be positive with probability number of positive sentences by total subjective sentences and vice versa.

We can conclude that our example review is positive since the number of positive sentences is more compared to the number of negative sentences.



Sentiment Analysis: Movie Reviews

Results

Having applied all the above steps, we manually cross-checked the results for 250 reviews. We found that our Sentiment Analysis system was accurate in 203 cases. So the precision of the system stands at 81%.

In the next step of evaluation, we selected 5 movies and randomly selected 10 reviews for each movie. Out of 50 reviews, 34 were correctly guessed by the system. So the precision of the system on random text stands at 68%.

Moving Forward

Though we have a fairly good accuracy rate today, to improve our process further we plan to

- > Create and annotate more training data in future to enhance the precision of our system
- > Take the probability of each sentence to calculate the overall sentiment instead of concluding on the basis of number of positive and negative sentences

About Talentica

At Talentica we help technology companies develop products. We have successfully built core intellectual property for several customers in their journey from pre-funded startups to profitable acquisitions. We have the deep technological expertise, proven track record and the best talent to build products successfully.

Whether you are an ISV with an Enterprise product or offer your product as a service over the Web or Mobile, Talentica's advanced approach to outsourced product development with the unique Build Operate Transfer model helps you build great products.

Contact Us @

India

Tel: +91 20 4075 1111 [Main Board]

Tel: +91 20 4075 1177 [Sales Enquiries]

USA

Tel: +1 408 332 5790

Fax: +1 408 332 5791



Email: info@talentica.com

Web: www.talentica.com
