



IT Systems Engineering | Universität Potsdam

Natural Language Processing

Sentiment Analysis

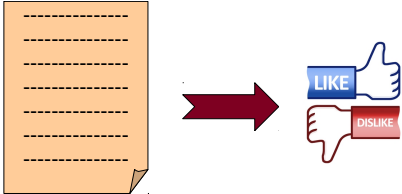
Potsdam, 7 June 2012

Saeedeh Momtazi

Information Systems Group

Sentiment Analysis

2



Outline

3

- 1 Applications
- 2 Task
- 3 Machine Learning Approach
- 4 Rule-based Approach

Outline

4

- 1 Applications
- 2 Task
- 3 Machine Learning Approach
- 4 Rule-based Approach

Hotel Reviews

HPI

Hasso Plattner Institut



The Ritz-Carlton, Berlin

Ranked #7 of 102 hotels in Berlin

Travelers' Choice® 2012 Winner

Show the lowest price for this hotel*

Show Prices

231 traveler photos

647 reviews from our community

Traveler rating

Travelers who booked this hotel also booked:

See which reviews travelers prefer · 23 traveler tips

"A really great place to stay in Berlin"

The price for the Ritz in Berlin was quite reasonable considering I was in a major European capital. That takes a lot of getting used to. It is like the prices you get staying in Chicago versus New York or San Francisco. The hotel itself is very new and nice. The service is great. Employees. They're a hooty...

"Great location and very comfortable hotel"

I person found this review helpful

My family of 4 stayed for 8 nights in mid-December/January. We had 2

Travelers also viewed...

- Grand Hotel Berlin
- Radisson Blu Hotel, Berlin
- Pullman Berlin Schwarzenberg

Browse nearby



Travelers who booked this accommodation also booked:

Sleep Rating



"A Complete Letdown"

I originally selected this hotel over others in Berlin... including Intercontinental... due to its top 3 TripAdvisor ranking at the time. I ended with dismay that ever since then its position had been continually slipping ever downwards!

Butler I searched into the full reviews. I will note that the best thing about this hotel is the breakfast. It is superb, and any criticism would only really come down to very personal taste and would be not picking anything in it catered for - a wonderful selection of fresh fruit, including berries, and cereals. Lots of different breads, rolls, and even bagels, and a toaster so you can toast sandwiches sandwiching one another, there is also an equally large array of pastries/cakes to enjoy with the bread. There is a good selection of cold cuts, fish, high quality cheeses, as well as the usual American style hot buffet complete with sandwiches made to order. All of this isn't enough, though. American pancakes. French toast, and waffles can all be made to order (at no extra charge), in short, this is an exemplary hotel breakfast in both terms of quality and variety. The service was usually attentive with glasses regularly topped up, but once it started to get busy (around 9.30 onwards, especially at the weekend) one did notice a drop in attention to detail, as well as a general winking to the bar area. If you are going to stay here, make sure you have breakfast included as it is the major reasoning factor of this property!

Upon arrival, a hair caped doorman opened my taxi door, and neatly removed the luggage and welcomed me to the hotel. If only things had continued like this... The girl at the front desk was polite, but far from welcoming as the doorman had been. Things were worse when we asked for my voucher. I explained that whilst I had booked through a third party I had e-mailed the hotel directly to confirm everything was as it should be and they had never mentioned the need for any voucher. She was very pleasant and went away to ask someone else, and still returned on the voucher. I advised that I could probably dig out an email once I had access to a computer and give it off to them. She still wasn't really satisfied, and a little huffy said that they would contact the travel agent for another copy. To add to this, I had been travelling since 4am that morning, whilst I did not expect the hotel to specifically know that, the front desk should always be very that most travellers will always be somewhat weary from their journey if it has not expired that they have checked into it and not contacted me if I had indeed needed to provide the voucher (which as it turned out I didn't).

A bell boy showed me up to my room and was extremely friendly and professional and showed me the various features of the room, how to operate touch-panel light switches etc. This was a thoughtful touch, and much appreciated. The room itself was a huge relief as I will discuss the rooms in more detail a little later and the double glazing was far from top quality, and a hissing noise could be heard where the window did not seal fully on the right side. I complained about this, and they advised someone would look at the window whilst I was out and the hissing noise was so loud and there was no noise explaining what had happened. I went to shower and get ready to go out, the hissing containing the vanity kit was empty which I was not best pleased with. I spoke to the duty manager in my stay out and he assured me he would locate the issue with the window. I returned from shower, still no noise and the hissing sound ever more apparent. I called the guest service department yet again (as soon as they were bored of my call) and they advised that someone had contacted the hotel to be normal. I can not remember who requires complete silence, but it was a very particular noise and the room attendant was pretty good so it would be very noticeable.

The next day I returned and was offered a change of room, and a complimentary



Viewed hotels

The Ritz-Carlton, Berlin

The Ritz-Carlton, Berlin, Best Western and special

Explore Berlin

Free Berlin Guide

Get it now

Überbuchungen zum Preis von 0

in über 1.200 Hotels

Jetzt buchen

AccorHotels.com

Value
Location
Sleep Quality

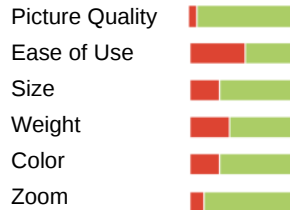
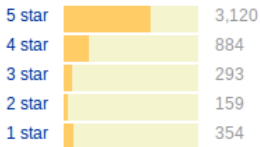
Cleanliness
Service

Product Reviews

6

★★★★☆ (4,810)

4.3 out of 5 stars



Social Media

7



The screenshot shows the Tweetfeeling interface with a search bar containing 'julia roberts' and a search button. Below the search bar, it displays 'Try some Twitter trends:' followed by a list of trending terms. A visual representation shows a green smiley face with '48' below it, a red frowny face with '21' below it, an equals sign, and '70%'.

Try some Twitter trends: [Rise & Grind](#) [Happy Birthday Jeremy](#) [Last Week Of School](#) [Say It Back](#) [MTV Movie Awards](#) [Richard Dawson](#) [Tiger Woods](#)

48 = 70% 21

Search

|| Search

#NottingHill's on yay! [julia roberts](#) get in me. I don't even know what Hugh Grant is.

@Vanna2409 lol I loved [julia roberts](#) in this movie!

RT @carnauval: Charlize Theron digs her Evil Queen persona way better than [julia roberts](#). Theron balances her OTT acts & fragility well. She saves #SWATH.

@sydd_thekydd lol I know I saw this tweet and decided to steal it. The one with [julia roberts](#) sucked.

Eat Pray Love with [julia roberts](#) is better than the book #mustsee

[julia roberts](#) was great in Pretty Space Woman.

@cheaty I've heard good things about it! The other Snow White movie w/ [julia roberts](#) was great as well.

I want someone to love me like Hugh grant loves [julia roberts](#) in notting hill

RT @saratopss: She loves [julia roberts](#)' movies <http://t.co/rrMaXsrT>

RT @iHeartLABlush: I love pretty woman [julia roberts](#) is amazing!

Event Analysis and Prediction

- Analyzing the side effects of events in different communities
- Predicting the election results
- Predicting the Stock exchange
- ...

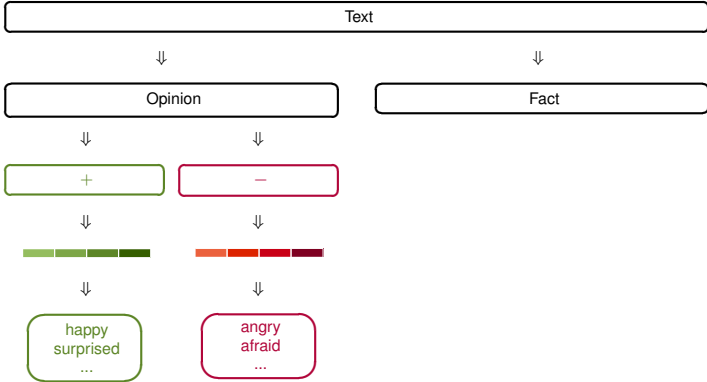
Outline

9

- ① Applications
- ② Task
- ③ Machine Learning Approach
- ④ Rule-based Approach

Sentiment Analysis Levels

10



Advanced Sentiment Analysis

11

- Opinion holder
- Opinion target / aspect

Students like Wikipedia because it is easy to use and it sounds authoritative.

op holder *target*

I had a nice stay in this hotel and the rooms were very clean.

aspect

- Mixed opinions

The restaurant has an amazing view but it is very dirty.

Other Names

- Opinion mining
- Opinion extraction
- Sentiment mining
- Subjectivity detection
- Subjectivity analysis

- Machine learning methods
⇒ classification

- Rule-based methods
⇒ dictionary oriented

Outline

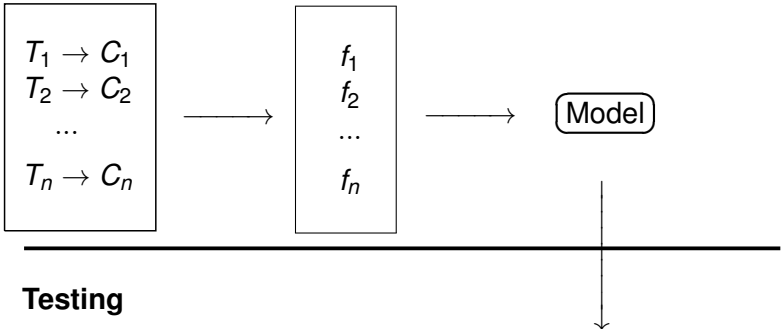
14

- ① Applications
- ② Task
- ③ Machine Learning Approach
- ④ Rule-based Approach

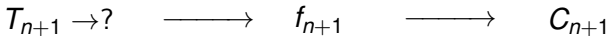
Machine Learning Approach

15

Training



Testing



Sentiment Classification

- Using any kinds of supervised classifiers
 - K Nearest Neighbor
 - Support Vector Machines
 - Naïve Bayes
 - Maximum Entropy
 - Logistic Regression
 - ...

Features

17

■ Word

- All words or adjectives?
All words works better than adjectives only

- Word occurrence or frequency?
Word occurrence is more useful than frequency
 - Using binary value for words
 - Replace all word counts higher than 0 in each text by 1

Features

■ Negation

- Negation words change the text polarity
 - Adding prefix *NOT_* to every word between negation and next punctuation

*“I did **not** like the restaurant location, but the food ...”*

I did not **NOT-like** **NOT-the** **NOT-restaurant** **NOT-location** but the food ...

Features

19

- Other emotions
 - Considering emoticons as additional features
 - :)
 - :(

Fine-grained Analysis

20

- Dealing with finer classes of sentiment
-3,-2,-1,+1,+2,+3

- Approaches
 - Using multiclass classifier (6 classes in this case)
 - Using two level classifier
 - First level: polarity classifier (positive or negative)
 - Second level: strength classifier (1 or 2 or 3)

Outline

21

- ① Applications
- ② Task
- ③ Machine Learning Approach
- ④ Rule-based Approach

Rule-based Approach

22

Training

~~$T_1 \rightarrow C_1$
 $T_2 \rightarrow C_2$
 ...
 $T_n \rightarrow C_n$~~



good
love
brave
intelligent
nice
...

bad
hate
lie
ugly
poor
...

Testing

$T_{n+1} \rightarrow ?$ \longrightarrow C_{n+1}

Rule-based Approach

23

- Looking for opinionated words in each text
- Classifying the text based on the number of positive and negative words

- Considering different rules for classification
 - Fine-grained dictionary
 - Negation words
 - Booster words
 - Idioms
 - Emoticons
 - Mixed opinions
 - Linguistic features of the language

Rule-based Approach

24

- Fine-grained Dictionary

"It was a *good* song."



"The song was *excellent*."



Rule-based Approach

25

- Negation Words

“The song was good.”



*“The song was **not** good.”*



Rule-based Approach

■ Booster Words

"The song was interesting."



*"The song was **very** interesting."*



*"The song was **somewhat** interesting."*



Rule-based Approach

27

- Idioms

“shock horror”



Rule-based Approach

28

- Mixed Opinions

*“The song was **good**, but I think its title was **strange**.”*



Rule-based Approach

29

- German Linguistic Features

*"I do **not** love the song."* ⇒ *"Ich liebe **nicht** das Lied."*
*"Ich liebe das Lied **nicht**."*

Opinion Dictionary

30

■ English

- Subjectivity Clues (2005)
- SentiSpin (2005)
- SentiWordNet (2006)
- Polarity Enhancement (2009)
- SentiStrength (2010)

■ German

- GermanPolarityClues (2010)
- SentiWortSchatz (2010)
- GermanSentiStrength (2012)

- Using opinion words as a feature in the algorithms
- Ignoring other words in the text

Adjectives alone do not work well,
but opinion words are the best features to be used