COM3110/4115/6150: Text Processing

Sentiment Analysis

Rob Gaizauskas

Department of Computer Science

University of She eld

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page31) 1 / 31

Learning Outcomes

By the end of the SA sessions, you will be able to:

Explain the relevance of the topic

Di erentiate between objective and subjective texts List the main elements in a sentiment analysis system

Provide a critical summary of the main approaches to the problem Explain how sentiment analysis systems are evaluated.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 2 / 31

Overview

De nition of the problem of sentiment analysis Approaches to sentiment analysis

Evaluation of sentiment analysis approaches

Based on survey and slides by Bing Liu (University of Illinois at Chicago), 2012.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 3 / 31

General goal

Certain texts, particularly on the Web, have emotions or sentiments or opinions, e.g.:

Blogs and microblogs (Twitter, etc.)

Social networks (Facebook, myspace, etc.)

User comments, like on Youtube, or on products, like on Amazon Review websites, like Rotten Tomatoes, yelp

Community websites, like Symantec Forums

Size of blogosphere: over 112 million blogs, 75,000 created each day, 1.2 million posts/day 1

Social networks like Twitter...

* <http://technorati.com/state-of-the-blogosphere/>

|  |  |  |
| --- | --- | --- |
| Rob Gaizauskas (U. She eld) | [COM3110/4115/6150: Text Processing](#page1) | 4 / 31 |

General goal

Extract opinions, sentiments and emotions expressed by humans in texts and use this information for business, intelligence, etc. purposes. Can't be done manually: huge volumes of opinionated text (esp. Big Data on the Web). Examples of applications:

Product review mining: Which features of the iPhone 6 customers like and which do they dislike?

Review classi cation: Is a movie review positive or negative?

Tracking sentiments toward topics over time: Is anger about the government policies growing or cooling down?

Prediction (election outcomes, market trends): Will the Tories win the next election?

Here: opinion = sentiment = emotion

Here: sentiment analysis = opinion mining

Although sentiment doesn't always express opinion: \I am sad today".

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 5 / 31

Importance of opinions

Whenever we need to make a decision, we may want to hear others' opinions

In the past: surveys, focus groups, consultants, opinions from friends and family

Nowadays: Word-of-mouth on the Web

User-generated media: one can express opinions on anything in reviews, forums, discussion groups, blogs ...

Opinions of global scale: no longer limited to one's circle of friends (individuals), small scale surveys, focus groups, etc. (businesses)

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 6 / 31

Importance of opinions

Individuals: interested in other's opinions when

purchasing a product or using a service,

nding opinions on political or other topics.

Businesses and organizations:

product and service benchmarking. market intelligence.

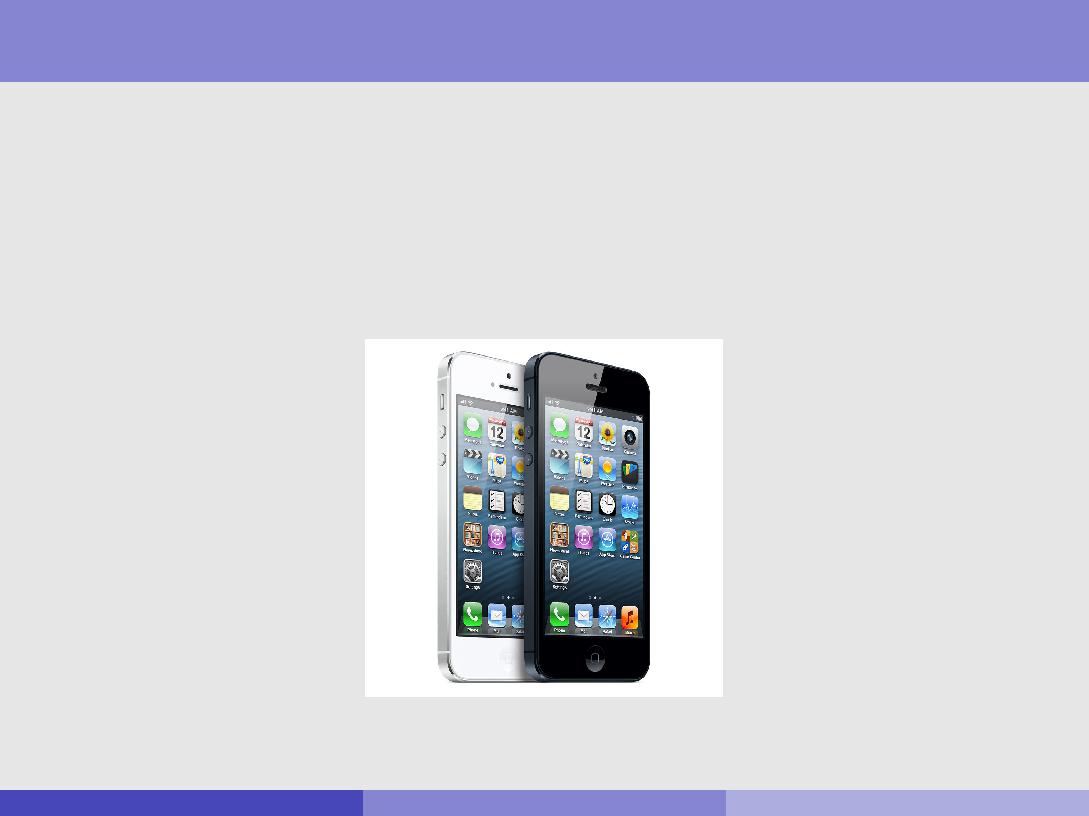
cost reduction: business spends a huge amount of money to nd consumer sentiments and opinions with consultants, surveys and focused groups, etc.

Ad placement: Placing ads in the user-generated content

Place an ad when one praises a product.

Place an ad from a competitor if one criticises a product.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 7 / 31



Motivations

\I bought the new iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life is not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, and wanted me to return it to the shop."

What do we see in this text? Positive or negative opinions?

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 8 / 31

Motivations

\I bought the new iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life is not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, and wanted me to return it to the shop."

Objective sentence

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 9 / 31

Motivations

\I bought the new iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life is not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, and wanted me to return it to the shop."

Positive and negative opinions about what?

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 10 / 31

Motivations

\I bought the new iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life is not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, and wanted me to return it to the shop."

Targets of opinions

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 11 / 31

Motivations

\I bought the new iPhone a few days ago. It was such a nice phone (I). The touch screen was really cool (I). The voice quality was clear too (I). Although the battery life is not long, that is ok for me (I). However, my mother was mad with me as I did not tell her before I bought the phone (mother). She also thought the phone was too expensive, and wanted me to return it to the shop.(mother)"

Holders of opinions

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 12 / 31

De nitions

Facts versus Opinions

Current text processing methods (e.g., web search, text mining) work with factual information.

Current search ranking strategy not appropriate for opinion retrieval. Sentiment analysis focuses on subjective statements - opinions,

sentiments, emotions: hard to express with a few keywords. E.g. What do people think of Motorola Cell phones?

Excellent phone, excellent service

Just double check with customer service to ensure the number provided

by amazon is for the city you wanted

I'd always eyed the nokia phones and had heard decent things about

t-mobile, so i gave it a whirl

It costed 500 dollars, not worth the price really

It costed 500 dollars!!!

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 13 / 31

Subjectivity analysis

Subjectivity classi cation is often the rst step for sentiment analysis:

subjective versus objective texts, e.g.:

Objective: I bought an iPhone a few days ago.

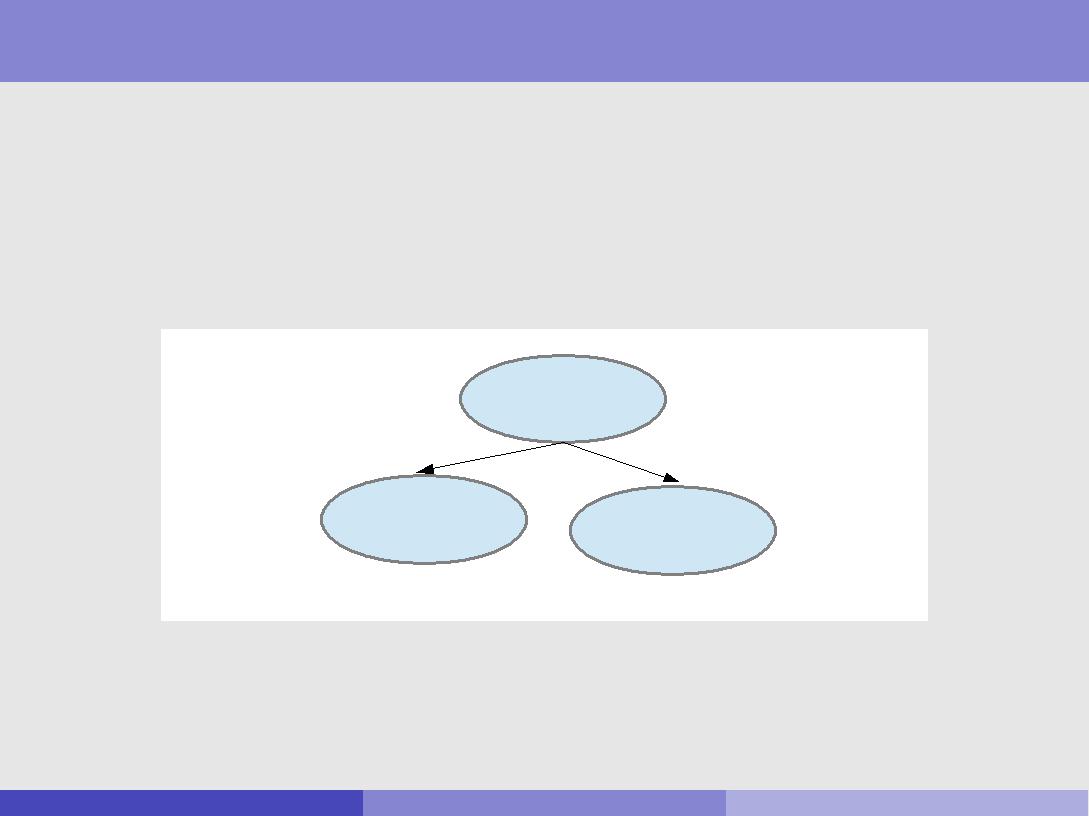
Subjective: It is such a nice phone.

However:

Subjective sentences do not always express positive or negative opinions, e.g.: I think he came yesterday.

Objective sentences can express opinion indirectly, e.g.: My phone broke in the second day.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 14 / 31



Sentiment Analysis

Target objects

Product, person, event, organization, or topic: o. It is represented as

A hierarchy of components, sub-components, etc.

Each node represents a component and has a set of attributes.

**iPhone 5**

{size, appearance, price,...}

**Battery** **Screen**

|  |  |  |
| --- | --- | --- |
| {battery life, size, weight, ...} | {sensitivity, brightness, ...} |  |
|  |  |

An opinion can be expressed on any component or attribute of the component { call them both \features" of the object.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 15 / 31

Bing Liu's model for Sentiment Analysis

An opinion is a quintuple (oj ; fjk ; soijkl ; hi ; tl ), where:

oj is a target object.

fjk is a feature of the object oj .

soijkl is the sentiment value of the opinion of the opinion holder hi (usually the author of the post) on feature fjk of object oj at time tl .

soijkl is positive, negative, neutral, or a more granular rating, such as 1-5 stars as in movie reviews.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 16 / 31

Sentiment Analysis

For example:

\I bought the new iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life is not long, that is ok for me. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, and wanted me to return it to the shop."

oj : iPhone

fjk : phone, screen, voice quality, battery life, price

soijkl : positive, positive, positive, negative, negative

opinion holder hi : I, I, I, I, mother

time tl : post's date

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 17 / 31

Sentiment Analysis

The task of opinion mining is: given an opinionated document:

Discover all quintuples (oj ; fjk ; soijkl ; hi ; tl ), or Discover some of these components

With that, one can structure the unstructured:

Traditional data and visualisation tools can be used to slice, dice and visualise the results.

Qualitative and quantitative analysis can be done.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 18 / 31

Sentiment Analysis

Granularity level:

Document level: classify a document (e.g., a movie review) based on the overall sentiment expressed by opinion holder into, e.g.: positive, or negative (and neutral). Assumption: Each document focuses on a

single object and contains opinions from a single opinion holder:

(oj ; fjk ; soijkl ; hi ; tl ), where oj = fjk

Sentence level: idem, but for (subjective) sentences, so these need to be identi ed rst.

Feature level: sentence/doc-level sentiment analysis can have mixed opinions and they do not tell what people like/dislike. A positive/negative opinion on an object does not mean that the opinion holder likes/dislikes everything. Better to nd opinions on component and/or its attributes. This allows all sorts of analyses.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 19 / 31

Sentiment Analysis

Granularity level - feature level (ctd) - Steps:

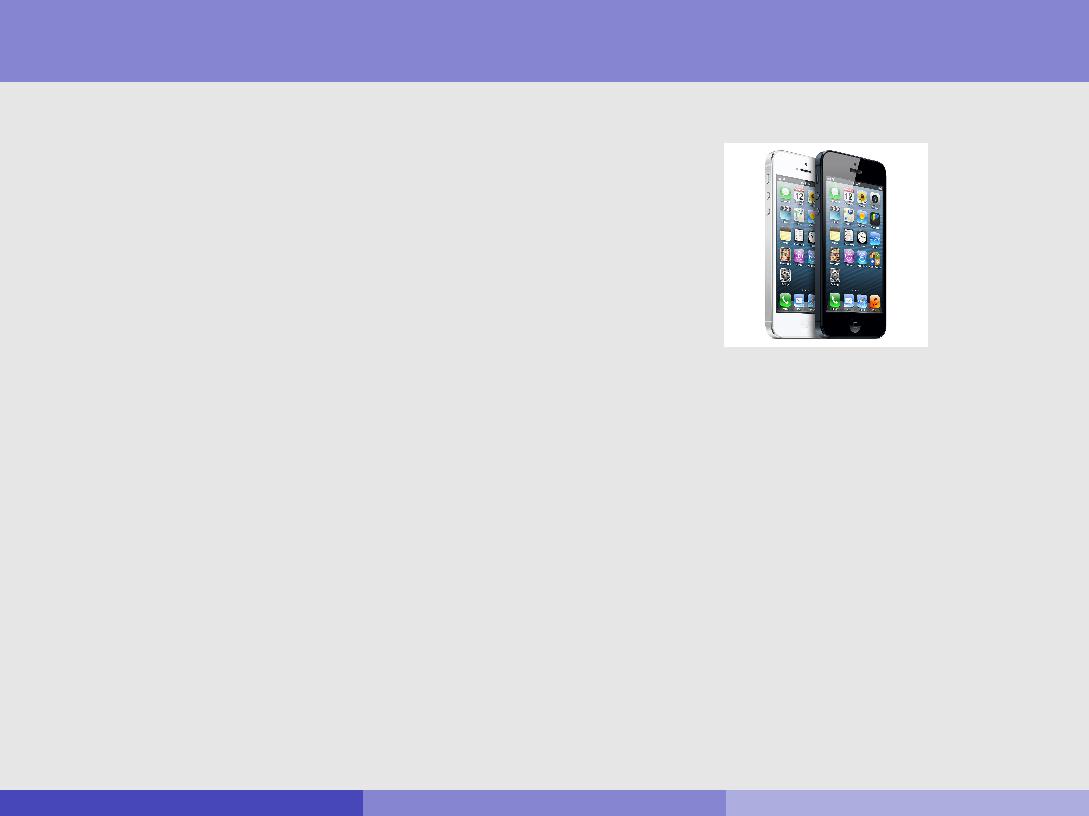
Identify and extract object features that have been commented on by an opinion holder (e.g., a reviewer).

Determine whether the opinions on the features are positive, negative or neutral.

Group synonym features, e.g. screen and touch screen.

Optional: produce a feature-based opinion summary of multiple reviews.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 20 / 31



Sentiment Analysis

Granularity level - feature level (ctd):

\I bought the new iPhone a few days ago. ..."

Feature Based Summary:

Feature1: touch screen

Positive:

The touch screen was really cool.

The touch screen was so easy to use and can do amazing things. ...

Negative:

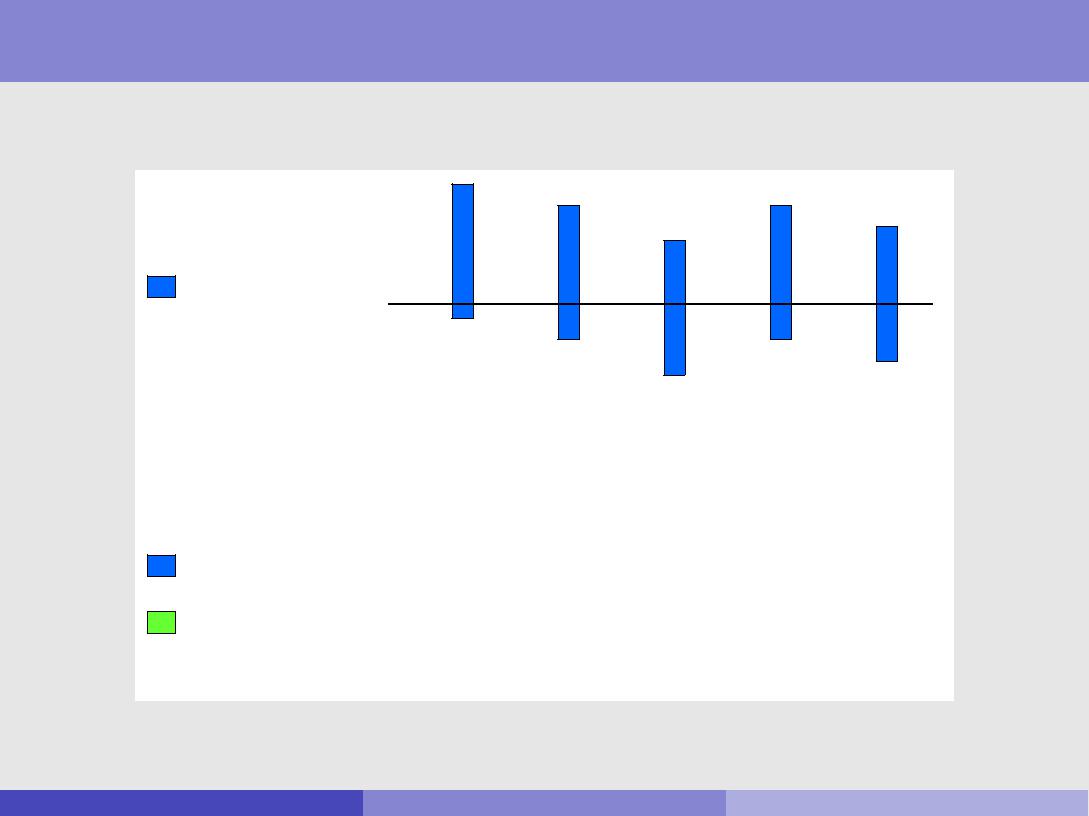
The screen is easily scratched.

I have a lot of di culty in removing nger marks from the touch screen.

Feature2: battery life

...

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 21 / 31



Sentiment Analysis

Granularity level - feature level (ctd): Visual Comparison

+

Summary of

reviews of

Cell Phone 1

\_

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Voice** | | |  | **ScreenBattery Size** | | | | | | | | | | |  | **Weight** | | | |  |
| Comparison of | + |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| reviews of |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Cell Phone 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Cell Phone 2 | \_ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

(Bing Liu)

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 22 / 31

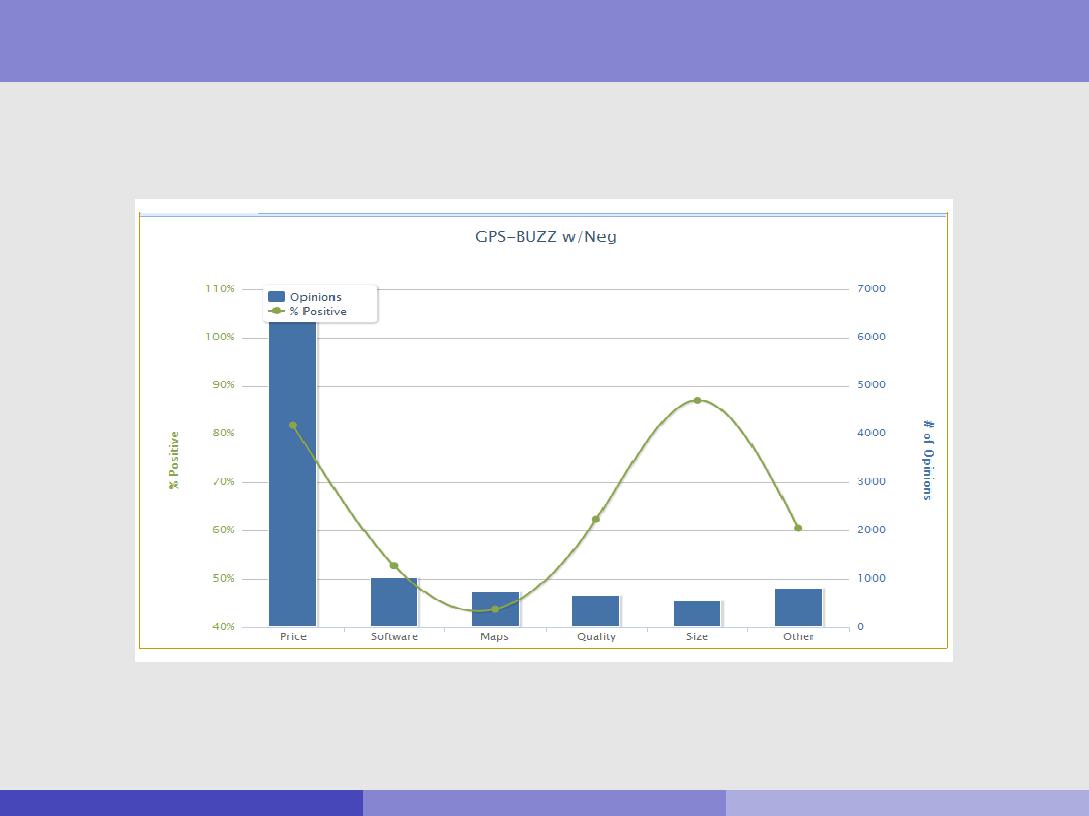


Sentiment Analysis

Granularity level - feature level (ctd): Proportion of opinions + examples

(Bing Liu)

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 23 / 31

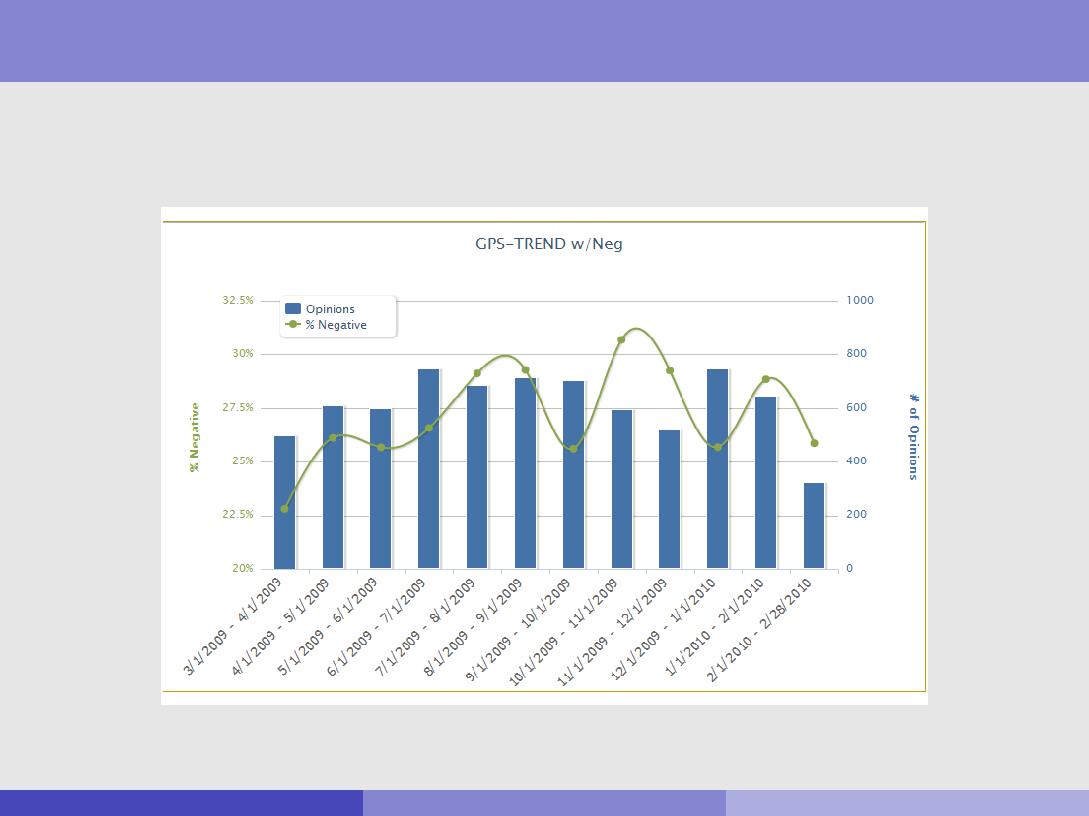


Sentiment Analysis

Granularity level - feature level (ctd): Frequency of opinions for a feature

(Bing Liu)

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 24 / 31



Sentiment Analysis

Granularity level - feature level (ctd): Aggregate opinions over time (trends)

(Bing Liu)

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 25 / 31

Challenges for Sentiment Analysis

This past Saturday, I bought a Nokia phone and my girlfriend bought a Motorola phone with Bluetooth. We called each other when we got home. The voice on my phone was not so clear, worse than my previous phone. The battery life was long. My girlfriend was quite happy with her phone. I wanted a phone with good sound quality. So my purchase was a real disappointment. I returned the phone yesterday.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 26 / 31

Challenges for Sentiment Analysis

I've seen movies where there was practically no plot besides explosion, explosion, catchphrase, explosion. I've even seen a movie where nothing happens. But White on Rice was new on me: a collection of really wonderful and appealing characters doing completely ba ing and uncharacteristic things.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 27 / 31

Challenges for Sentiment Analysis

One has to solve a number of language processing problems:

(oj ; fjk ; soijkl ; hi ; tl )

oj : a target object: Named Entity Recognition

fjk : a feature of oj : Information Extraction

soijkl : a sentiment about fjk : Sentiment determination

hi : an opinion holder: Information (or metadata) Extraction

tl : a time: Information (or metadata) Extraction

In addition:

Co-reference resolution Relation extraction

Synonym match (\voice" = \sound quality")

None of them is a solved problem!

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 28 / 31

Main components

Identifying target objects

Named Entity Recognition: well-known tools based on gazetteers and simple context rules. E.g.: Paris, BMW and Ford.

Need good gazetteers: Web is dynamic, new products appearing everyday.

Will not work for objects like names of movies.

Bootstrap from seed gazetteers: e.g. if know that iPhone 4 is an object, can nd out that iPhone 5 is also an object.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 29 / 31

Main components

Co-reference (and synonym) resolution

Important to resolve objects and features.

E.g.: \I bought a Canon d500 camera yesterday. It looked beautiful. I took a few photos last night. They were amazing". I am happy with the device.

!

E.g.: \I bought a Canon d500 camera yesterday. The Canon d500 camera looked beautiful. I took a few photos last night. The photos were amazing". I am happy with the camera.

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 30 / 31

Extra reading

Bing Liu and Lei Zhang (2012). A survey on opinion mining and sentiment analysis. Kluwer Academic Publishers:

<http://www.cs.uic.edu/~lzhang3/paper/opinion_survey.pdf>

Rob Gaizauskas (U. She eld) [COM3110/4115/6150: Text Processing](#page1) 31 / 31