Proseminar Sentimentanalyse:

*Introduction*

Winter Semester 2017/2018
Instructor: Michael Wiegand
Spoken Language Systems
Saarland University, Germany
Before we start:

Do not forget to register for the course (link now available via course webpage; this is not the LSF-registration)

The deadline is 24th October (tomorrow!)
Outline

- What is Sentiment Analysis?
- Administrative Matters
Outline

• What is Sentiment Analysis?
• Administrative Matters
What is Sentiment Analysis?

- Sentiment Analysis deals with the (automatic) detection and analysis of texts that express opinions, evaluations, emotions, and uncertainty.
- *I think* the decision was right.
- *This movie is absolutely* *boring*.
- *I am* *frustrated*.
- *This idea of yours* *could* work.
Subtasks in Sentiment Analysis

• What is subjective language?
• How can we automatically determine the polarity of a text?
• How can we automatically extract the entities that are involved in an opinion expressed in text?
• Beyond that, take also into consideration:
  • Emotions
  • Uncertainty
  • …
Subjectivity Detection

- Automatically identify texts that convey subjectivity:
  - *I hate this book.*
  - *The bus leaves at 9:30h.*
- Are there particular words or phrases that convey subjectivity?
  - *awful, okay, superb, expect, thought*
- Are there certain syntactic clues that convey subjectivity?
  - *Adjectives are more indicative than other parts of speech.*
Polarity Classification

- Automatically determine the polarity of texts:
  - *Peter hates pancakes*.
  - *Mary loves pizza*.
- One may also determine the degree of polarity:
  - *nice << great << excellent*
Polarity Classification

• Need to know which words carry polarity.
  • good, clever, fantastic, excellent
  • bad, idiot, horrible, shocking

• Need to consider context:
  • Negation:
    Peter does not hate pancakes.
  • Irony:
    This is the greatest operating system ever; it crashes every 5 minutes on my laptop!
Opinion Holder/Target Extraction

• Holder/source extraction of the opinion:
  • *In a recent interview, [Noam Chomsky] praised Snowden.*
  • *This is total nonsense. [speaker]*

• Target extraction of the opinion:
  • *In a recent interview, Noam Chomsky praised [Snowden].*
  • *[This] is total nonsense.*
Other Parameters to Consider

- Granularity:
  - document
  - sentence
  - phrase or word
- Domain:
  - Train and test on the same domain (*in-domain classification*).
  - Train and test in different domains (*cross-domain classification*).
- Language:
  - Train and test on the same language (*monolingual setting*).
  - Train and test on different languages (*cross-lingual setting*).
Why do we want to do Sentiment Analysis?

• General aim of Natural Language Processing is to find algorithms to compute the meaning of text.

• Sentiment analysis can be regarded as one building block towards that goal.
Why do we want to do Sentiment Analysis?
Why do we want to do Sentiment Analysis?

One of the most popular applications is automatically classifying sentiment in movie reviews.
Why do we want to do Sentiment Analysis?
Why do we want to do Sentiment Analysis?

Various companies of commercial products are keen to know the general sentiment of customers towards their products/brand.
Why do we want to do Sentiment Analysis?
Why do we want to do Sentiment Analysis?

Another application is monitoring popular opinion on political issues (e.g. prior to a general election).
Why do we want to do Sentiment Analysis?
Why do we want to do Sentiment Analysis?

- Many hotel reviews are fakes (e.g. praising a bad hotel).
- Use Sentiment Analysis to filter this *opinion spam*. 
Outline

• What is Sentiment Analysis?
• Administrative Matters
Aims of this Course

• Learn something about Sentiment Analysis.
• … from a linguistic point of view.
• Learn to read and understand scientific literature.
• Learn to present a paper.
• Learn to write a term paper.
Aims of this Course

• Jump in at the deep end:
  • No textbook.
  • We will read research publications.

• Difference to textbooks:
  • Condensed language
  • Terminology may vary.
  • Basic research (*task/evaluation may be a bit abstract*)
  • There may be things you will have to look up yourself.
  • Establish connections to between different papers yourself.
How is this course organized?

• Short introduction (given by instructors):
  • this week
  • next week
• Student presentations:
  • From the third week onwards: every week one presentation
Admission Policy

• We only have space for 10 students.
• We may have more registrations.
• BSc students (from CoLi) are given priority!
• This course will be held in German!
Requirements for Participation

• You should have attended the lectures:
  • Einführung in die Computerlinguistik
  • Mathematische Grundlagen III

• Some basic knowledge on linguistics (mostly syntax and semantics).

• Some basic knowledge on machine learning and evaluation we will have a very brief recap on that next week.
Requirements for Passing the Course

• Oral presentation
• Term paper
• Reviewing a paper
• Constant active participation
• You are allowed to be absent without excuse for only one session!
• Oral exam (on demand)
Oral Presentation

- Length: up to 45 minutes \((excluding\ discussion)\)
- In German or English
- Based on one paper \(\text{(see course webpage)}\).
- You are welcome to include more related papers!
- We have a discussion after each oral presentation:
  - You will have to answer questions!
  - Make sure you really understood the problem discussed in your paper!
- If requested, I will give a few useful hints how to do a presentation.
Oral Presentation

• I offer every student an additional (individual) meeting prior to the oral presentation:
  • Discuss open questions.
  • Feedback on presentation draft

• Arrangement of appointment:
  • Send in draft at least **8 days prior** to your presentation!
Term Paper (Hausarbeit)

• Write an essay on the paper you orally presented.
• Length: 10-15 pages
• In German or English
• Link to detailed guide how to write a term paper on the course webpage (this guide is binding for this seminar!)
• Use LaTeX (template on webpage)
• Deadline for term paper: 28th February
Reviewing a Paper

• Review a paper that you do not orally present.
• You are given the main responsibility of the discussion session after the student presentation of that paper:
  • Ask questions about the paper.
  • Get involved into the discussion.
  • Show us that you have understood the paper and thought about the problem/task presented.
Oral Exams (*CoLi only!*)

- I offer oral exams for this course.
- Remember: passing 3 oral exams is one requirement for the *old* CoLi-BSc!
- Topic of exam: another paper from the seminar which you did **not** orally present.
- Length of exam: 15 minutes
- Please notify instructors in case you want to have an exam by *October 30th*. 
How is the final mark computed?

- **Weighting:**
  - 40% oral presentation
  - 40% term paper
  - 20% reviewing

- **In case oral exam is requested:**
  - 10% oral exam
  - 90% *everything else with internal weighting from above*

- **If the final „score“ is between two grades, course participation will be considered.**
Timeline

• Next session(s):
  • Recap on machine learning and evaluation
  • How to *present* a paper (on demand)

• From then onwards:
  • One student presentation per week.
  • Schedule on webpage.
What to do next?

• Register for the course as soon as possible.

• Tell us which paper you want to present.
  • Give a list of 5 papers sorted by priority
  • deadline: 26th October (Thursday)
If you have further questions

• Ask now!
• Another good point in time: at the end of each session.
• *Don’t* try to discuss complex issues via email!
• Ask for an appointment instead.