Sentiment Analysis and Opinion Mining in Product Reviews
Fine-grained Analysis and Cross-Linguality

Feb 23rd, 2015

Roman Klinger
roman.klinger@ims.uni-stuttgart.de
http://www.romanklinger.de/

IMS, University of Stuttgart and CITEC, Bielefeld University

Talk at the conference
“From Big Data to Smart Knowledge – Text and Data Mining in Science and Economy”
Motivation

Task Definition

Joint Model

Cross-Linguality
Motivation

Task Definition

Joint Model

Cross-Linguality
Motivation

Decision making is individually based on...!
- Size
- Battery
- Operating System
- Sensors
- ...

Goal

Extract aspects from reviews with evaluation.
Take Away

- Closer look on product reviews and their content
- Fine-grained sentiment analysis: **Aspects and Evaluation**
- Probabilistic model with joint inference has higher performance than pipeline model
- Building models for languages without manual annotation is possible: **Annotation projection and instance filtering**
Outline

1. Motivation
2. Task Definition
3. Joint Model for Aspect and Subjective Phrase Detection
4. Cross-Linguality
## Examples (I)

### Review for dog costume

**So cute!** . . . I can tell you it looks great on her and my mom and I got a total kick out of it. You have to cut your own slits for your dog’s ears behind the frill, but honestly that’s better than having ones in the wrong place, I think. Mine looks a little redder than what the picture shows, too, but it’s just too fabulous, I really don’t mind. I cannot wait to walk my dog nearer Halloween. . .

### Aspects and Subjective evaluations

- looks – great – got a total kick out of it
- slits – that’s better
- frill
- redder – don’t mind
Examples (II)

Review for trash can

Great Deal!!!! stainless steel trash can price
look performance awesome
motion sensor stopped working
trash lid
turn off button underside lid trash can seems be
very good quality looks great only complaint recommendation
place the on and off switch lid
compartment

http://www.amazon.com/dp/B0031M9H3A
Task Definition

- Extract **aspects** of a product, service, ...
- Extract **subjective phrases** which express an opinion
- Extract **relations**, that means, which subjective phrase is evaluating which aspect
Approaches to the task

Rules
- Assume frequent nouns to denote aspects
- Assume adjectives close by to denote subjective phrases to evaluate them
- e.g. M. Hu et al. (2004). “Mining and summarizing customer reviews”. In: SIGKDD

Distributional
- Topic modelling, distributional semantics
- e.g. C. Lin et al. (2009). “Joint sentiment/topic model for sentiment analysis”. In: CIKM

Supervised learning
- Based on annotated segments, learn to produce such on unseen data
- e.g. Yang et al., 2013; Jakob et al., 2010
The idea for a joint model

Given subjective phrases
I really like the **aspect** of my camera.

Given aspect phrases
I **subjective** the display.

Given nothing?
I **subjective** the **aspect**.

⇒ Build model which uses interactions between these phrases.
Hypothesis

Pipeline model

- Step 1: Detect phrases
  I really like the display.

- Step 2: Detect relations
  I really like the display.

Issues

- Relation cannot be extracted when phrases were not detected
- Information of relation is not taken into account for extracting phrases

How could a joint model be built?
Joint Model

- Probabilistic model
- Inference via Markov Chain Monte Carlo Sampling
- Training similar to perceptron
- Model based on lexical and grammatical information

Inference steps
- Propose subjective phrase for each token
- Propose aspect with relation
- Change span length
- ...
Results

What is the potential impact?

- Given subjective ⇒ predict aspects
- Predict subjective ⇒ predict aspects
- Given aspect ⇒ predict subjective
- Predict aspect ⇒ predict subjective
- Predict aspect ⇔ predict subjective
Multiple Languages

Summarization so far
We have a model trained on annotated data.

Research question
How to infer a model for other languages without annotating?

Teaser: Work in progress
Cross-Lingual Training

**Motivation**

There are certainly better machines, but offers the best price-performance ratio.

**Possible Issues**

- Translation quality might be suboptimal
- Alignment quality might be suboptimal
- Does that limit the quality of a model?

**Idea**

- Only use projected data with high quality
- Estimate quality of translation
Cross-Lingual Training with Filtering: de → en

![Graphs showing F1 measures for different conditions.](image-url)
Motivation

Task Definition

Joint Model

Cross-Linguality

Take Away

- Closer look on product reviews and their content
- Fine-grained sentiment analysis: Aspects and Evaluation
- Probabilistic model with joint inference has higher performance than pipeline model
- Building models for languages without manual annotation is possible: Annotation projection and instance filtering
Sentiment Analysis and Opinion Mining in Product Reviews
Fine-grained Analysis and Cross-Linguality

Feb 23rd, 2015

Roman Klinger
roman.klinger@ims.uni-stuttgart.de
http://www.romanklinger.de/

IMS, University of Stuttgart and CITEC, Bielefeld University

Talk at the conference
“From Big Data to Smart Knowledge – Text and Data Mining in Science and Economy”