Rule-based Opinion Target and Aspect Extraction to Acquire Affective Knowledge

Stefan Gindl\textsuperscript{1} \quad Albert Weichselbraun\textsuperscript{2} \quad Arno Scharl\textsuperscript{1}

\textsuperscript{1}Department of New Media Technology
MODUL University Vienna

\textsuperscript{2}University of Applied Sciences
HTW Chur

WWW MABSDA 2013
Outline

1. Introduction
   - Objective
   - Quick Outline

2. Methodology
   - Architecture
   - Examples

3. Results and Future Outline

4. Related Projects
   - Applications
   - Activities
Objective

Focus on opinion target and aspect extraction

“Who says what about whom/what?”

How to find out the “whom/what”?

Find out non-explicit aspects:

- opinion target $\xrightarrow{\text{good thanks to}}$ sentiment aspect, and
- opinion target $\xrightarrow{\text{bad due to}}$ sentiment aspect
Quick Outline

- Syntactic rules propagate sentiment onto targets/aspects
- Anaphora resolution for target extraction across sentences
- Application of the technique to a corpus consisting of 100,000 Amazon reviews in the electronics domain
Two major steps:

- Propagation of the sentiment value from a sentiment term onto a target. Application of a simple anaphora resolution to identify targets across sentences.
- Extraction of multi-term aspects

Preprocessing:

- Tokenization and sentence splitting with the webLyzard framework
- Dependency parsing using the Stanford parser
Architecture

- Sentences
  - Sentence splitting, tokenization
  - Stanford parser

- Graph transformation

- Syntactic rules

- Anaphora resolution

- Target and aspect extraction
Sentiment propagation in detail

Target/aspect extraction within a sentence: rules defined in [1]

- Propagate a sentiment charge from an opinionated term onto a noun target:
  **Rule 1:** $O \rightarrow O - Dep \rightarrow T$
  Example: “The phone has a good screen” ($good \rightarrow mod \rightarrow screen$)

- Propagate the sentiment charge of the first identified target onto other noun targets in the same sentence:
  **Rule 2:** $O \rightarrow O - Dep \rightarrow H \leftarrow T - Dep \leftarrow T$
  Example: “iPod is the best mp3 player” ($best \rightarrow mod \rightarrow player \leftarrow subj \leftarrow ipod$)
Sentiment propagation in detail

Cross-sentence: heuristic defined in [2]

- The first noun of the subsequent sentence is a co-referent
- Apply this rule for all subsequent sentences
Intra-sentence target extraction

Propagation graph of “The phone has a good screen.” (rule 1)
Intra-sentence target extraction

Propagation graph of “The phone has a good screen but a bad battery.” (rule 1)
Intra-sentence target extraction

Propagation graph of “The iPod is the best mp3 player.” (rule 2)
Cross-sentence target extraction

Propagation graph of “Yesterday I bought a new phone. It is the best purchase I have ever made.”
Cross-sentence target extraction

Propagation graph of “My phone is very good. It is light and has a long battery life. It looks sleek and has a crisp and bright screen. However, it is expensive and the sound quality is bad.”
Target and aspect extraction

Target extraction

- Positive targets: quality, product, price, sound, case
- Negative targets: quality, product, drive, one, thing

Aspect extraction

- Positive aspects: sound quality, **light weight**, **high quality**, digital camera, **low price**, little camera, small size, **long battery life**, ..., **spare battery**
- Negative aspects: first time, first one, new one, other reviews, few days
Future Outline

- Use aspects to identify opinionated targets where the target did not receive explicit sentiment
- Learn new extraction patterns from known target/aspect relations
- Determine applicability for different domains
- Migrate the approach to other languages
Ambiguity of sentiment terms is a problem

- Leverage context of a sentiment term to disambiguate an ambiguous sentiment term
- Create so-called “contextualized sentiment lexicons”
- Next steps: integrate external resources (SenticNet, ConceptNet)
The Sentiment Quiz, a game with a purpose

- Facebook application to create resources for sentiment analysis
  - Sentence corpus
  - Sentiment lexicons
    - Seven different languages: English, German, Italian, Portuguese, Russian, Spanish, French
    - English: 2724
    - German: 508
    - Italian: 1065
    - Portuguese: 2814
    - Russian: 2589
  - Bootstrapping to increase coverage of the sentiment lexicon
The Sentiment Quiz

Does the following term have a **negative**, **neutral** or **positive** meaning?

**Reinforcement**

Noun

Status

13

Your current score is **65 points**. Invite your friends and earn 10% of the points they make!

Spread the Word

Tell your Friends!

You will earn **10% of your friends’ points** after they accept your invitation! The calculation is recursive, so if they invite others you will even get more bonus points.

---

Gindl, Weichselbraun, Scharl (WU, MU)

Target and Aspect Extraction
Media Watch on Climate Change
(ecoresearch.net/climate)

U.S. Election Monitor 2008
(ecoresearch.net/election2008)
Interest Group on German Sentiment Analysis (IGGSA),
- MLSA corpus: Multi-Layered reference corpus for German Sentiment Analysis
- Shared Task: Competition of tools for holder/target extraction co-located with KONVENS 2014
  - sites.google.com/site/iggsahome

Practice And THeory of Opinion mining and Sentiment Analysis (PATHOS 2013),
- Submission deadline: July 12, 2013
  - sites.google.com/site/pathosworkshop
Thank you for your attention!
