

# Opinion Mining

A Generic Approach to Generate Opinion Lists of Phrases for Opinion Mining Applications (WISDOM-12)

A Phrase-Based Opinion List for the German Language (PATHOS-12)

Sven Rill



Herbstschule Information Retrieval, Dagstuhl - October 3, 2012

# Table of Contents

---

- 1 Introduction
- 2 Opinion List Generation
- 3 Results
- 4 Future Work

# Opinion Mining

---

## United Breaks Guitars

`http://www.davecarrollmusic.com/music/ubg/`

`http://www.youtube.com/watch?v=5YGc4z0qozo`

# Opinion Mining

## Videostatistiken

### Aufrufe und Fundquellen



# Opinion Mining

---

4 von 4 Kunden fanden die folgende Rezension hilfreich

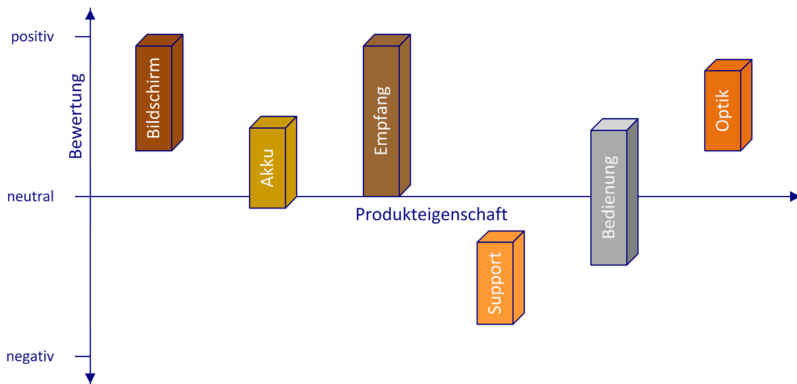
★★★★★ **Einfach der Hammer**, 2. Januar 2011

Von **franz** (österreich) - [Alle meine Rezensionen ansehen](#)

**Rezension bezieht sich auf: Apple iPhone 4 16GB (Wireless Phone Accessory)**

Alles was das Herz begehrt, eine gute kamera die tolle bilder macht, ein xenon blitz wäre aber besser gewesen. einfache bedienung, viele viele tolle apps, davon viele auch kostenlos, itunes store am handy, man kann musik kaufen wo immer man ist. ein display das besser gar nicht sein könnte, von einem empfangsproblem habe ich bis heute noch nichts bemerkt, auch von angeblich gelben flecken am display keine spur. internet ist gut zu bedienen. zubehör: kopfhörer sind voll in ordnung, ladekabel welches zugleich das usb kabel ist, 2 stück apple sticker. was mir besonders gut gefällt ist die verbindung wi-fi. das design ist unschlagbar und ein richtiger eye catcher. wenn man es einmal hat dann will man nichts anderes mehr. Fazit: 5 sterne - sehr empfehlenswert, wer etwas besonderes will?!....iphone4

# Opinion Mining



# Resources for Opinion Mining

---

## Opinion Mining Applications

- Need lexical resources  
(lists of opinion bearing words + opinion values)
- Opinion Lists mainly exist for English language<sup>1</sup>
- Lack of resources for the German language<sup>2</sup>
- Difficult: treatment of phrases containing negation and/or valence shifters → sentiment composition

---

<sup>1</sup>SentiWordNet, Semantic Orientations of Words, Subjectivity Lexicon

<sup>2</sup>Polarity Lexicon (PL), GermanPolarityClues (GPC)

# Basic Idea

## Idea

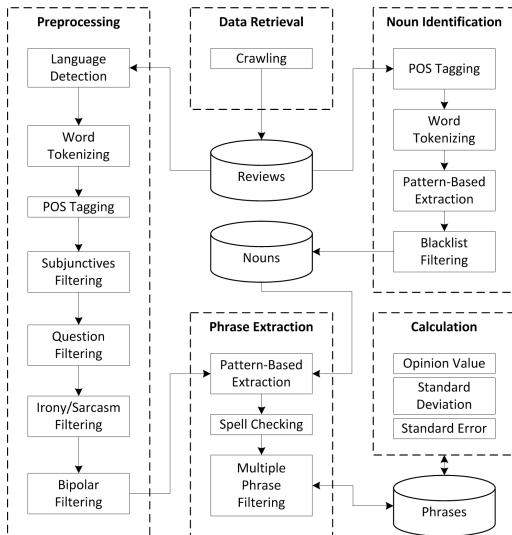
- Exploit correlation between review title and star rating because ...
  - Review title summarizes customer's opinion
  - Star rating summarizes customer's opinion
- Usage of any review platform possible

### Examples: Review star ratings and titles

★★★★★	Dieses Buch ist <u>brilliant!</u>	<i>This book is <u>brilliant!</u></i>
★★★★★	<u>Bestes Handy!</u>	<i><u>Best mobile phone!</u></i>
★★★	Der Film ist <u>nicht so gut.</u>	<i>The movie is <u>not so good.</u></i>
★	<u>Einfach nur schrecklich.</u>	<i><u>Just awful.</u></i>



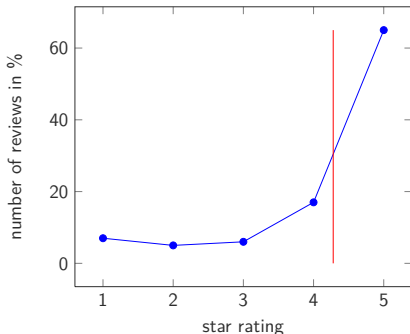
# Overview Opinion List Generation



# Opinion Value Correction (1)

Problem: Wrong value for some single adjectives/nouns.

Reason: J-shaped distribution of online reviews



Mean star rating = 4.28 stars / Mean opinion value = 0.64

⇒ Correction necessary

# Opinion Value Correction (2)

---

Classification of single adjectives/nouns into two classes:

**1** *Neutrals Following J-shaped Distribution:*

Star distribution follow the J-shaped distribution

- Quantitatively:

$$S_{J_1} = 1 - \sqrt{\sum_{s=1}^5 \left(\frac{n_s}{n} - a_s\right)^2}$$

$S_{J_1} > 0.85 \Rightarrow$  following J-shaped distribution

- Qualitatively:

$S_{J_2}$ : TRUE if  $n_1 \geq n_2$  and  $n_4 \leq n_5$

**2** *Others*

---

$a_s$ : relative frequency of s-star review titles in the whole sample

# Opinion Value Correction (3)

Correction of the opinion values: Weighting the star rating frequencies  $n_s$  with the relative frequencies of the whole sample  $a_s$

$$OV_c = \left( \frac{\sum_{s=1}^5 \frac{n_s \cdot s}{a_s}}{\sum_{s=1}^5 \frac{n_s}{a_s}} - 3 \right) : 2$$

Adjective	OV	$S_{J_1}$	$S_{J_2}$	$OV_c$
schnell - <i>fast</i>	<b>0.69</b>	0.96	F	—
gut - <i>good</i>	<b>0.67</b>	0.75	F	—
hübsch - <i>nice</i>	<b>0.56</b>	0.71	T	—
jung - <i>young</i>	0.66	0.93	T	<b>0.12</b>
schwarz - <i>black</i>	0.67	0.97	T	<b>0.08</b>
andere - <i>other</i>	0.57	0.88	T	<b>0.01</b>

# Examples for Adjectives and Phrases

Opinion Values (OV) for adjectives and adjective-based phrases

<b>Adjective</b>	<b>OV</b>
großartig - <i>great</i>	0.95
gut - <i>good</i>	0,67
brauchbar - <i>useful</i>	0.38
durchschnittlich - <i>average</i>	-0.02
schlecht - <i>bad</i>	-0.56
grotzenschlecht - <i>abysmal</i>	-0.96
<b>Adjective-based phrase</b>	<b>OV</b>
extrem gut - <i>extremely good</i>	1.00
sehr gut - <i>very good</i>	0.90
nicht schlecht - <i>not bad</i>	0.38
nicht optimal - <i>not optimal</i>	-0.05
eher schwach - <i>rather weak</i>	-0.26
sehr schlecht - <i>very bad</i>	-0.83

# Examples for Nouns and Phrases

Opinion Values (OV) for nouns and noun-based phrases

<b>Noun</b>	<b>OV</b>
Weltklasse - <i>world class</i>	0.98
Meisterwerk - <i>masterpiece</i>	0.94
Zumutung - <i>impertinence</i>	-0.77
Zeitverschwendung - <i>waste of time</i>	-0.93
<b>Noun-based phrases</b>	<b>OV</b>
absolutes Muss - <i>absolute must</i>	0.97
sehr gute Qualität - <i>very good quality</i>	0.94
mangelnde Qualität - <i>lack of quality</i>	-0.55
absoluter Fehlkauf - <i>absolute bad buy</i>	-1.00

# Negation

Two effects of negation ('<phrase>' vs. 'not <phrase>')

- 1 sign(OV) changed, |OV| roughly unchanged
- 2 polarity changed from 'strong' to 'weak' or 'zero' polarity

## Examples

Phrase	$OV_P$	$OV_{\bar{P}}$
sehr hilfreich - <i>very helpful</i>	0.91	-0.69
gut - <i>good</i>	0.67	-0.64

Phrase	$OV_P$	$OV_{\bar{P}}$
perfekt - <i>perfect</i>	0.95	0.11
ganz einfach - <i>quite simple</i>	0.82	0.36
optimal - <i>optimal</i>	0.87	-0.04

# Valence Shifters

Examples of phrases based on 'gut' - 'good'

Phrase	OV	Frequency
sensationell gut - <i>sensational good</i>	1.00	27
sehr sehr gut - <i>very very good</i>	0.97	114
sehr gut - <i>very good</i>	0.90	13,002
echt gut - <i>really good</i>	0.83	264
wirklich gut - <i>really good</i>	0.79	589
<b>gut - good</b>	<b>0.67</b>	<b>31,619</b>
ziemlich gut - <i>pretty good</i>	0.62	105
nicht so gut - <i>not that good</i>	-0.08	323
nicht sehr gut - <i>not very good</i>	-0.46	13
nicht gut - <i>not good</i>	-0.64	180
leider nicht gut - <i>unfortunately not good</i>	-0.85	17



# Future Work

---

## Opinion list for German

- Manually cleaning of list
- Provision of list to the community, planned in October  
[www.iisys.de](http://www.iisys.de), group *Analytical Information Systems*,  
[www.opinion-mining.org](http://www.opinion-mining.org)

## Opinion lists for other languages

- List for English

## Future plans

- Inclusion of verb-based phrases
- Implementation of benchmarks and a quantitative evaluation

Thank you for your attention!