

Irony Detection on Amazon Customer Reviews

Presentation

Max Blunck, Steffen Knapp, Lennart Schönwandt, Daniel Wachter

01.02.2018

Softwareprojekt 17/18

Table of contents

1. Problem
2. Daten
3. Features
4. Implementation
5. Ergebnisse
6. Erkenntnisse und Probleme

Problem

Die Performance vieler NLP-Tasks wie z.B. Opinion Mining und Sentiment Analysis kann durch die zuverlässige Erkennung von Ironie und Sarkasmus in Texten (deutlich) verbessert werden.

Ironie als Kommunikationsmittel

- Ironie und Sarkasmus sind sehr häufig genutzte Ausdrucksformen
- Implizitheit und oft Kontextabhängigkeit
- Wird von menschlichen Sprechern trotzdem meist intuitiv verstanden
- Unterscheidung zwischen verbaler und situativer Ironie
- Aber: keine hinreichend linguistisch-formale Beschreibung vorhanden

Daten

Andere existierende Amazon-Datensätze

- Tsur et al. (2010):
 - Twitter und Amazon-Daten
 - 66.000 Amazon-Rezensionen
 - Nur einzelne Sätze, wenig vorannotierte Daten, sehr ähnliche Sätze
 - Nicht öffentlich verfügbar
- Reyes, Rosso (2011):
 - Rezensionen von nur fünf verschiedenen Produkten
 - Aussortierung von Rezensionen mit weniger als vier Bewertungssternen

Filatova 2012

- 437 **ironische** und 817 **nicht-ironische** Rezensionen
- Komplette Reviews, nicht nur einzelne Sätze
- Zweistufige Annotation mit Amazon Mechanical Turk
 - Data Collection
 - Data Quality Control

Data Collection

- Zielvorgabe: 1000 Paare von ironischen oder sarkastischen und normalen Rezensionen zu jeweils demselben Produkt
- Text, Label und Link
- Danach Bereinigung der Daten

Data Quality Control

- Fünf zusätzliche MTurkers bewerten die gesammelten Rezensionen neu (Mehrheitsverfahren)
- Krippendorff's alpha coefficient zur Unterscheidung zwischen zuverlässigen und unzuverlässigen Annotatoren
- Schätzen der Sternebewertung (Korrelation: 0,889)

Aufbau der Korpusinstanzen

- LABEL
- FILENAME
- STARS
- TITLE
- DATE
- AUTHOR
- PRODUCT
- REVIEW

Beispiel Ironie

<STARS>1.0 </STARS>

<TITLE>The title is the best part of the book</TITLE >

<DATE>July 25, 1998 </DATE>

<AUTHOR>A Customer </AUTHOR>

<PRODUCT>The God of Small Things (Paperback)</PRODUCT >

<REVIEW>

If you enjoy repeated phrases (e.g. "fountain in a Love-In-Tokyo"), overused metaphors, and capitalized words in the Middle Of Sentences For Effect, then this is the book for you. Otherwise, rest assured that the title really is the best part of the book.

</REVIEW>

Verteilung der Bewertungen auf sarkastische und normale Reviews

		Number of reviews with				
		1 ★	2 ★	3 ★	4 ★	5 ★
sarcastic	437	262	27	20	14	114
regular	817	64	17	35	96	605

Verteilung der Tokens

- Korpus gesamt: 21.744 voneinander verschiedene Tokens
- Nur in ironischen Reviews: 5.336 Tokens
- Nur in normalen Reviews: 9.468 Tokens
- Durchschnittliche Tokenanzahl pro Review: 271,9 Tokens

Buschmeier et al. (2014) als Baseline

An Impact Analysis of Features in a Classification Approach to Irony Detection in Product Reviews.

Features

f1: Bag-of-N-grams (Wortformen)

- mit CountVectorizer (sklearn)

f2: Bag-of-POS-Bigrams

- Tagging durch den NLTK-Tagger (Penn Treebank)
- ca. 1340 verschiedene POS-Bigramme

f8: Bag-of-N-grams (Lemmata)

- mit CountVectorizer (sklearn)

Tsur et al. (2010)

- Unterscheidung zwischen high-frequency words (HFWs) and content words (CWs)
- Threshold: 1.000/1.000.000 Wörter
- Punctuationen sind HFWs
- Beispiel:
"CW is about as CW as a CW ."
→ neues Korpus erstellen
- f1: Bag-of-N-grams

Davidov et al. (2010)

- Kontrast zwischen dem Sentiment des Reviews & des Star-Ratings
- Negatives Star-Rating + Positiver Sentiment - und umgekehrt- könnte auf Ironie hinweisen
- Sentiment Analysis mit TextBlob¹ (Python Library)

¹<http://textblob.readthedocs.io/en/dev/>

Davidov et al. (2010)

- Untersuchen der Punctuation
 - Relevante Satzzeichen
 - Satzzeichen als Rudeltiere
- Kapitalisierung
 - Dynamisches Herausfiltern von Akronymen?

Riloff et al. (2013)

Beispiel:

Absolutely **adore** it **when my bus is late**.

[+ **positives** Verb] [- **negative** Situation]

positiver Sentiment:

- positives Verb = [["R",] "V"]

Negative Situation:

- Uni-POS-List = ["V"]
- Bi-POS-List = ["VV", "VR", "RV", "TV", "VN", "VN", "VN", "VP", "V"]
- Tri-POS-List = ["VVV", "VVR", "VRV", "VVR", "VRR", "RVV", "VNR", "VIN", "VTV", "VIP"]

- Die Star-Ratings (1-5) der Reviews bilden den Featurevektor

Implementation

Programmstruktur

- Modulare Aufteilung nach Features
- Hauptprogramm zum Zusammenführen der Features & Machine Learning

Tools

- Feature Extraktion
 - POS-tagger: NLTK
 - Sentiment Analysis: TextBlob
- scikit-learn für:
 - Parameter Tuning (GridSearch)
 - Cross Validation
 - Training & Testing

Ergebnisse

10-fold Cross Validation Scores:

Alle Features, auf gesamten Korpus

	Precision	Recall	F1-Score
SVM	72.4	69.1	70.2
Dec. Trees	79.1	66.4	72.0
Naive Bayes	62.3	73.2	67.2
Logistic Reg.	80.3	70.4	74.7

10-fold Cross Validation Scores:

Alle Features, auf gesamten Korpus

	Precision	Recall	F1-Score	
SVM	72.4	69.1	70.2	71.3 ²
Dec. Trees	79.1	66.4	72.0	72.2
Naive Bayes	62.3	73.2	67.2	65.0
Logistic Reg.	80.3	70.4	74.7	74.4

²Baseline Buschmeier (2014)

Scores auf Testdaten:

Alle Features

	Precision	Recall	F1-Score
SVM	74.7	77.1	75.9
Dec. Trees	88.0	68.8	77.2
Naive Bayes	80.0	16.7	27.6
Logistic Reg.	81.3	63.5	71.3

Scores³ auf Testdaten:

Einzelne Featuresets

Feature	SVM	Dec. Trees	N. Bayes	Log. Reg.
F1: Bag-of-ngrams	66.7	33.3	56.6	59.5
F2: Bag-of-POS	57.1	39.0	54.5	53.1
F3: Surface-Patterns	55.8	38.3	51.7	57.1
F4: Sent/Rating	65.1	65.1	0.0	0.0
F5: Punctuation	0.0	42.3	0.0	0.0
F6: Contrast-Feat.	0.0	0.7	0.0	0.0
F7: Stars	71.2	71.2	0.0	69.1

³F1-Scores

Beste Konfiguration:

Bag-of-Bigrams +

Surface-Patterns +

Sentiment/rating-contrast +

Star-rating

	Precision	Recall	F1-Score
SVM	81.3	77.1	79.1
Dec. Trees	80.7	69.8	74.9
Naive Bayes	85.7	6.2	11.7
Logistic Reg.	84.1	60.4	70.3

Beste Konfiguration ohne Metadaten (Sterne):

Bag-of-words +

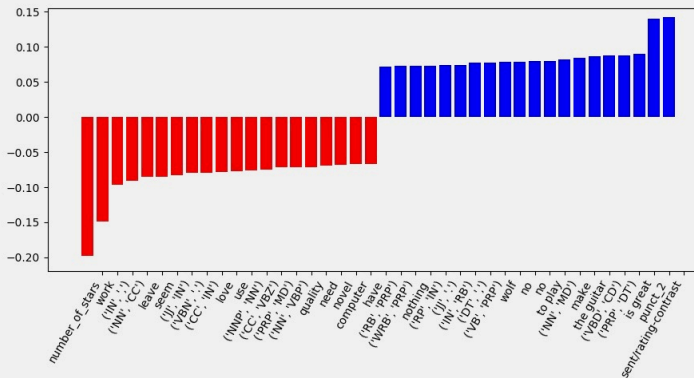
Bag-of-POS +

Surface-Patterns

	Precision	Recall	F1-Score
SVM	69.2	65.6	67.4
Dec. Trees	55.3	27.1	36.4
Naive Bayes	63.1	55.2	58.9
Logistic Reg.	78.2	63.5	70.1

Top Features

SVM



Fehlerhafte Klassifizierung

3 of 3 people found the following review helpful:

★★★★★ **Best face lotion out there... Period.**, July 24, 2006

By [L. F. Bretts](#)

If you are looking for a lotion that won't leave your face feeling greasy, then look no further. It is THE best out there.

Help other customers find the most helpful reviews

Was this review helpful to you?

[Report this](#) | [Permalink](#)

Review Details

Item not available

Reviewer



[L. F. Bretts](#)
REAL NAME™

Location: San Diego, CA

New Reviewer Rank: 3,787,798

Classic Reviewer Rank: 1,236,697

Prediction: Ironic

Label: not Ironic

Fehlerhafte Klassifizierung

15 of 19 people found the following review helpful:

☆☆☆☆ Taste..., January 14, 2010

By [enoezark "k-1"](#)

This review is from: **Communion Wafers Box of 1000**

With a little peanut butter and jelly, these things make a great snack on the go.

Help other customers find the most helpful reviews

Was this review helpful to you?

[Report this](#) | [Permalink](#)

Review Details

Item



[Communion Wafers Box of 1000](#)

☆☆☆☆ (28 customer reviews)

5 star:	<div></div>	(11)
4 star:	<div></div>	(2)
3 star:	<div></div>	(7)
2 star:	<div></div>	(4)
1 star:	<div></div>	(4)

Prediction: Ironic

Label: not Ironic

Fehlerhafte Klassifizierung

17 of 17 people found the following review helpful:

★★★★★ **OMG, so great**, October 3, 2009

By **Robert D. Walton "Wolf Heart"**

This review is from: **Handerpants (Misc.)**

I mean, I always wanted my crotch and my hands to have more in common, now they do!

Help other customers find the most helpful reviews

Was this review helpful to you?

[Report this](#) | [Permalink](#)




Review Details

Item



Handerpants

★★★★☆ (4 customer reviews)

5 star:  (2)
4 star:  (1)
3 star:  (0)
2 star:  (0)
1 star:  (1)

\$7.99

Prediction: not Ironic

Label: Ironic

Erkenntnisse und Probleme

Fragen + Diskussion

References i

- D. Davidov, O. Tsur, and A. Rappoport. 2010. Semisupervised recognition of sarcastic sentences in Twitter and Amazon. In Proceeding of the 23rd international conference on Computational Linguistics. July.
- Elena Filatova. 2012. Irony and sarcasm: Corpus generation and analysis using crowdsourcing. In Language Resources and Evaluation Conference. LREC2012.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 2004). pages 168–177. Seattle, WA, USA. August.
- Buschmeier et al. 2014. An Impact Analysis of Features in a Classification Approach to Irony Detection in Product Reviews.

- Antonio Reyes. Paolo Rosso. 2011. Mining Subjective Knowledge from Customer Reviews: A Specific Case of Irony Detection.
- Ellen Riloff. Ashequl Qadir. Prafulla Surve. Lalindra De Silva. Nathan Gilbert. Ruihong Huang. 2013. Sarcasm as Contrast between a Positive Sentiment and Negative Situation.