Implicit Feature Extraction for Sentiment Analysis in Consumer Reviews

Kim Schouten Flavius Frasincar

Erasmus University Rotterdam
The Netherlands

Introduction

- Features are actually aspects

Introduction

- Features are actually aspects
- Aspects denote specific characteristics of the product or service being reviewed

Introduction

- Features are actually aspects
- Aspects denote specific characteristics of the product or service being reviewed
- Aspect-level sentiment analysis allows for a fine-grained overview of a product or service, which is more useful than one overall score

Introduction

- Features are actually aspects
- Aspects denote specific characteristics of the product or service being reviewed
- Aspect-level sentiment analysis allows for a fine-grained overview of a product or service, which is more useful than one overall score
- This research is limited to finding aspects (no actual sentiment analysis) in consumer reviews

Introduction - What are implicit aspects?

Introduction - What are implicit aspects?

Implicit aspects can be defined as aspects that are implied by the text, rather than literally mentioned

Introduction - What are implicit aspects?

Implicit aspects can be defined as aspects that are implied by the text, rather than literally mentioned

Examples

- *“I can’t see a thing when it’s sunny.”*
- *“The phone lasts all day.”*

Introduction - Human processing of implict aspects

- The implicit aspect is inferred based on the words in the sentence

Introduction - Human processing of implict aspects

- The implicit aspect is inferred based on the words in the sentence
- Mapping from the words in the sentence to the invisible implicit aspect(s)

Introduction - Human processing of implict aspects

- The implicit aspect is inferred based on the words in the sentence
- Mapping from the words in the sentence to the invisible implicit aspect(s)
- This mapping is shared across all users of the language

Introduction - Human processing of implict aspects

- The implicit aspect is inferred based on the words in the sentence
- Mapping from the words in the sentence to the invisible implicit aspect(s)
- This mapping is shared across all users of the language
- Hence, usually only well-known aspects or broad categories are implied

Examples

- Price, size, weight, service, etc.

Proposed Method

- Find the mapping between words in the sentence and implicit features

Proposed Method

- Find the mapping between words in the sentence and implicit features
- Count the number of co-occurrences between the implicit aspects and the words in the sentence

Proposed Method

- Find the mapping between words in the sentence and implicit features
- Count the number of co-occurrences between the implicit aspects and the words in the sentence
- For an unlabeled sentence, the implicit feature that co-occurs most often with the words in the sentence is chosen...

Proposed Method

- Find the mapping between words in the sentence and implicit features
- Count the number of co-occurrences between the implicit aspects and the words in the sentence
- For an unlabeled sentence, the implicit feature that co-occurs most often with the words in the sentence is chosen...
- ...if it exceeds a certain threshold

Training Algorithm - Counting

Initialize set of word lemmas with frequencies O
Initialize set of implicit features F
Initialize co-occurrence matrix C
for sentence $s \in$ training data do
    for word $w \in s$ do
        $O(w) = O(w) + 1$
    end for
    for implicit feature $f \in s$ do
        add f to F
        for word $w \in s$ do
            $C(w, f) = C(w, f) + 1$
        end for
    end for
end for

Training Algorithm - Threshold optimization

threshold$=0$
$bestF_1 = 0$
for $t = 0$ to 1 step 0.001 do
    Process training data
    Compute F_1
    if $F_1 > bestF_1$ then $threshold = t$
    end if
end for

Processing Algorithm

for sentence $s \in$ test data do
currentBestImplicitFeature $=$ *empty*
scoreOfCurrentBestImplicitFeature $= 0$
for implicit feature $f \in F$ do
$score = 0$
for word $w \in s$ do
if $O(w) > 0$ then
$score = score + C(w, f)/O(w)$
end if
end for
$score = score/$ length(s)
if $score > scoreOfCurrentBestImplicitFeature$ then
$currentBestImplicitFeature = f$
$scoreOfCurrentBestImplicitFeature = score$
end if
end for
if $scoreOfCurrentBestImplicitFeature > threshold$ then
Assign $currentBestImplicitFeature$ to s
end if
end for

Formula Notation

$$score_i = \frac{1}{v} \sum_{j=1}^{v} \frac{c_{i,j}}{o_j}, \tag{1}$$

where

- i is the ith aspect in the list of possible aspects for which the *score* is computed
- v is the number of words in the sentence
- j represents the jth word in the sentence
- $c_{i,j}$ is the co-occurrence frequency of aspect i and lemma j in the data set
- o_j is the frequency of lemma j in the data set

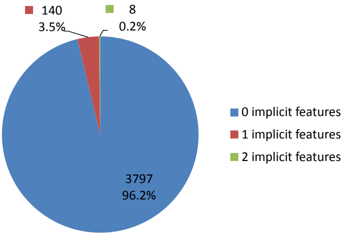
Known Limitations

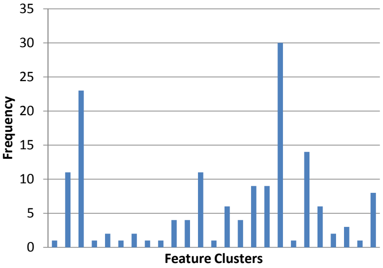
- Only one implicit aspect is chosen per sentence
- Sufficient amount of labelled training data is required

Data Analysis

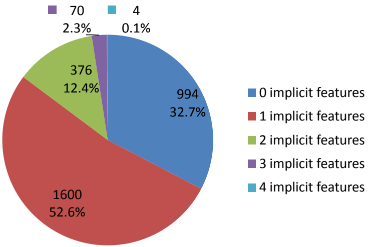
- Two data sets: product reviews and restaurant reviews
- Both contain about 3000 sentences

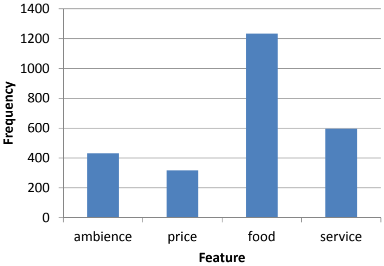
Data Analysis - Product Data





Data Analysis - Restaurant Data





Evalution

Method

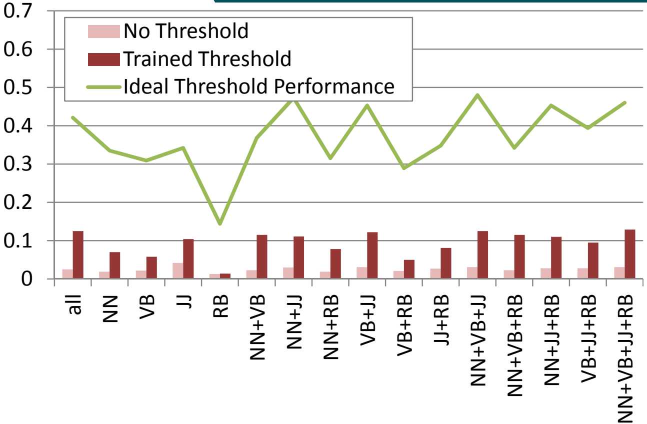
- All evaluations have been performed using 10-fold cross-validation
- Both the counting and the threshold optimization are done using training data only
- Because of previous work, we used different combinations of part-of-speech filters to control what kind of words would be contained in the co-occurrence matrix

Evalution

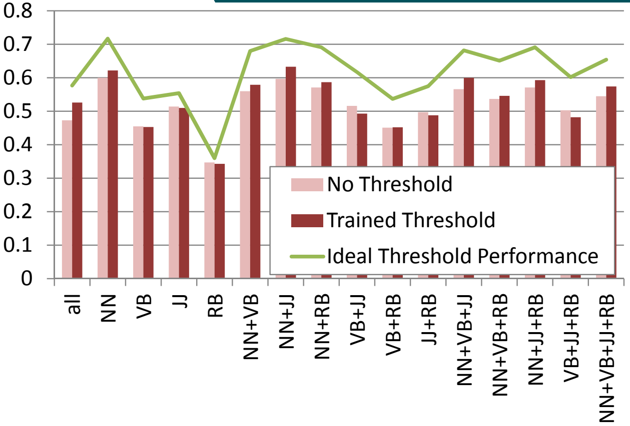
Error types

- Incorrectly state that a sentence contains some implicit aspect: *lower precision*
- Incorrectly state that a sentence does not contain an implicit aspect: *lower recall*
- Correctly state that a sentence contains an implicit aspect, but pick the wrong one: *both precision and recall will be lower*

Results - Product Data

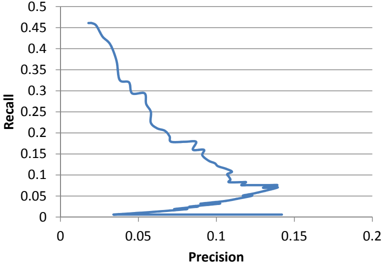


Results - Restaurant Data



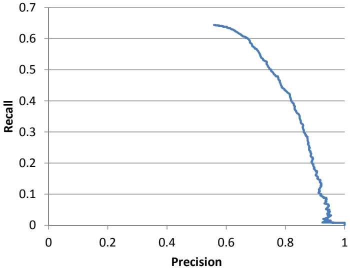
Results - Precision-Recall curves

Product Data



Using NN+VB+JJ+RB filter

Restaurant Data



Using NN+JJ filter

Results - Comparison

product review data set

method	no threshold	trained threshold	difference
Zhang & Zhu	1.2% (`all`)	1.4% (`NN+VB+JJ+RB`)	+0.2 pp.
proposed method	4.2% (`JJ`)	12.9% (`NN+VB+JJ+RB`)	+8.7 pp.
difference	+3 pp.	+11.5 pp.	

restaurant review data set

method	no threshold	trained threshold	difference
Zhang & Zhu	31.5% (`all`)	32.4% (`all`)	+0.9 pp.
proposed method	59.7% (`NN+JJ`)	63.3% (`NN+JJ`)	+3.6 pp.
difference	+28.2 pp.	+31.1 pp.	

Conclusions

- Significantly improved on existing method, although at the cost of being a supervised method
- The algorithm needs a sufficient amount of data to work properly
- The use of a threshold is beneficial, especially for the small data set

Future Work

- Allow for more than one implicit aspect per sentence
- Learn a threshold for each implicit aspect
- Move towards a more concept-level approach
  - "This phone doesn't fit in my pocket"

Questions?

Contact

schouten@ese.eur.nl

Acknowledgments

The authors are partially supported by the Dutch national program COMMIT. We also would like to thank Andrew Hagens, Gino Mangnoesing, Lotte Snoek, and Arno de Wolf for their contributions to this research.