Fine-grained Sentiment Analysis
An Overview of the Task and its Main Challenges

Orphée De Clercq
HUSO 2016 - EMOSEDE
November 15, 2016
Restaurants in Barcelona

**Uma**

#1 of 7,575 Restaurants in Barcelona

4.50/5, 502 reviews

"... and dumpling/ceviche) cuisines as w..." 28/09/2016
"... was the oxtail meat item which was ..." 18/02/2016

££££ | Map | Visitor photos (656)

Cuisines: International, Mediterranean, Fusion, European, Spanish

**Blavis**

#2 of 7,575 Restaurants in Barcelona

4.38/5, 387 reviews

"... go early by Spanish eating times as..." 08/07/2016
"... the eggplant, oxtail, shrimp, tacos..." 04/04/2016

£ - £££ | Map | Visitor photos (180)

Cuisines: Mediterranean, European, Spanish

**Spoonik**

#3 of 7,575 Restaurants in Barcelona

4.49/5, 497 reviews

"We are far from the typical Spanish but..." 23/11/2015
"When visiting there I always have tapas..." 15/05/2016

££££ | Map | Visitor photos (310)

Cuisines: Mediterranean, Spanish, Colombian, International
Restaurants in Barcelona

Uma
#1 of 7,575 Restaurants in Barcelona
502 reviews
"... and dumpling/ceviche) cuisines as w..." 28/09/2016
"... was the oxtail meat item which was ..." 18/02/2016
££££ | Map | Visitor photos (656)
Cuisines: International, Mediterranean, Fusion, European, Spanish

Blavis
#2 of 7,575 Restaurants in Barcelona
387 reviews
"... go early by Spanish eating times as..." 08/07/2016
"... the eggplant, oxtail, shrimp, tacos..." 04/04/2016
£ | Map | Visitor photos (180)
Cuisines: Mediterranean, European, Spanish

Spoonik
#3 of 7,575 Restaurants in Barcelona
497 reviews
"We are far from the typical spanish but..." 23/11/2015
"When visiting there I always have tapas..." 15/05/2016
££ | Map | Visitor photos (310)
Cuisines: Mediterranean, Spanish, Colombian, International

AFFORDABLE FOOD
FRIENDLY SERVICE
Sentiment analysis

**Early 2000s:**
- Wiebe (2000)
- Pang et al. (2002)

**Rise of Web 2.0 applications 2010-2016:**
- 20,000 Google Scholar
- 731 papers in WoS

⇒ **newswire text**

⇒ **user-generated content**
Sentiment analysis

• Opinion polls, surveys

• Sentiment analysis on UGC:
  ➢ To track how a brand is perceived by consumers (Zabin & Jefferies, 2008)
  ➢ For market (Sprenger et al., 2014), election prediction (Bermingham & Smeaton, 2011)
  ➢ To determine the sentiment of financial bloggers towards companies and their stocks (O’Hare et al., 2009)
  ➢ By individuals who need advice on purchasing the right product or service (Dabrowski et al., 2010)
  ➢ By nonprofit organizations, e.g., for the detection of suicidal messages (Desmet, 2014)
  ➢ ...
Sentiment analysis

*Coarse-grained*: document or sentence $= \text{POS} | \text{NEG} | \text{NEUTRAL}$

→ Does not allow to discover what people like and dislike exactly.

→ Not only interested in general sentiment about a certain product, but also in their opinions about specific features, parts or attributes of that product.

*Fine-grained*: “almost all real-life sentiment analysis systems in industry are based on this level of analysis” (Liu, 2015, p. 10).
Aspect-based (or feature-based) sentiment analysis systems focus on the detection of all sentiment expressions within a given document and the concepts and aspects (or features) to which they refer.

- Van Hee et al. (2014): Coarse-grained SA on Twitter
- De Clercq et al. (2015): ABSA (English resto)
- De Clercq (2015): SemEval ABSA (Dutch resto)
- De Clercq and Hoste (2016): ABSA (Dutch resto, smartphones)
- Pontiki et al. (2016): SemEval ABSA 8 languages, 4 domains
- 2016-2017: valorisation project (various domains, languages)
The best research = team research
Overview

1. Introduction
2. Task Definition
3. Datasets and Annotation
4. Subtasks
   - Aspect Term Extraction
   - Aspect Term Categorization
   - Aspect Term Polarity Classification
5. Challenges
6. Conclusion
Overview

① Introduction

② Task Definition

③ Datasets and Annotation

④ Subtasks

➤ Aspect Term Extraction

➤ Aspect Term Categorization

➤ Aspect Term Polarity Classification

⑤ Challenges

⑥ Conclusion
Task definition


Definition of an opinion by Liu (2012):

“An opinion is a quintuple, \((e_i; a_{ij}; s_{ijkl}; h_k; t_l)\), where \(e_i\) is the name of an entity, \(a_{ij}\) is an aspect of \(e_i\), \(s_{ijkl}\) is the sentiment on aspect \(a_{ij}\) of entity \(e_i\), \(h_k\) is the opinion holder, and \(t_l\) is the time when the opinion is expressed by \(h_k\). The sentiment \(s_{ijk}\) is positive, negative, or neutral, or expressed with different strength/intensity levels.” (pp. 19-20)

⇒ Automatically deriving quintuples = five different tasks
Task definition

Uma

502 Reviews | #1 of 7,575 Restaurants in Barcelona

“Just perfect”
Reviewed 31 May 2016

Food was excellent, place is small, but really lovely. Service was perfect and super friendly. Highly recommend this restaurant in Barcelona

Helpful? 2 Thank Reviewer X
1. Entity extraction + categorization

Extract all entity expressions in a document collection, and categorize or group synonymous entity expressions into entity clusters.
2. Aspect extraction + categorization

Extract all aspect expressions of the entities, and categorize these aspect expressions into clusters. These aspects can be both explicit and implicit.
3. Opinion holder extraction + categorization

Extract opinion holders for opinions from text or structured data and categorize them.
4. Time extraction + standardization

Extract the times when opinions are given and standardize different time formats
5. Aspect sentiment classification

Determine whether an opinion on an aspect is positive, negative or neutral, or assign a numeric sentiment rating to the aspect.
Derived quintuples:

- (Uma, Food, positive, Reviewer X, May-31-2016)
- (Uma, Ambience, positive, Reviewer X, May-31-2016)
- (Uma, Service, positive, Reviewer X, May-31-2016)
- (Uma, Restaurant, positive, Reviewer X, May-31-2016)
Task definition: customer reviews

- (Uma, Food, positive, Reviewer X, May-31-2016)
- (Uma, Ambience, positive, Reviewer X, May-31-2016)
- (Uma, Service, positive, Reviewer X, May-31-2016)
- (Uma, Restaurant, positive, Reviewer X, May-31-2016)

ABS A of customer reviews:
- Aspect Extraction
- Aspect Categorization
- Aspect sentiment classification

SemEval Task Description
(Pontiki et al., 2014, 2015, 2016)
Overview

① Introduction

② Task Definition

③ Datasets and Annotation

④ Subtasks
  ➢ Aspect Term Extraction
  ➢ Aspect Term Categorization
  ➢ Aspect Term Polarity Classification

⑤ Challenges

⑥ Conclusion
Customer reviews

Previous research

Movie reviews (Thet et al. 2010), electronic products (Hu and Liu 2004, Brody and Elhadad 2010), restaurants (Ganu et al. 2009).

➡️ Difficult to compare

SemEval shared task

Online data competition: everyone works on the same data.

➡️ Better to compare
➡️ State of the art
SemEval benchmark data

- Three runs of the task (2014, 2015 & 2016)
- Lots of data in different domains & languages

<table>
<thead>
<tr>
<th>Domain</th>
<th>Subdomain</th>
<th>Language</th>
<th>#Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronics</td>
<td>Camera</td>
<td>Chinese</td>
<td>8040</td>
</tr>
<tr>
<td></td>
<td>Laptops</td>
<td>English</td>
<td>3308</td>
</tr>
<tr>
<td></td>
<td>Phones</td>
<td>Chinese</td>
<td>9521</td>
</tr>
<tr>
<td></td>
<td>Phones</td>
<td>Dutch</td>
<td>1697</td>
</tr>
<tr>
<td>Hotels</td>
<td></td>
<td>Arabic</td>
<td>6029</td>
</tr>
<tr>
<td>Restaurants</td>
<td></td>
<td>Dutch</td>
<td>2297</td>
</tr>
<tr>
<td></td>
<td></td>
<td>English</td>
<td>2676</td>
</tr>
<tr>
<td></td>
<td></td>
<td>French</td>
<td>2429</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Russian</td>
<td>4699</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spanish</td>
<td>2951</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Turkish</td>
<td>1248</td>
</tr>
<tr>
<td>Telecom</td>
<td></td>
<td>Turkish</td>
<td>3310</td>
</tr>
</tbody>
</table>
Annotation

Guidelines are available online: http://goo.gl/wOf1dX

Three steps:

I. All explicit and implicit targets - the word or words referring to a specific entity or aspect - are annotated.
II. These targets are assigned to domain-specific predefined clusters of aspect categories.
III. Sentiment expressed towards every aspect is indicated.
Lange wachttijd.

Zelfde dessert, 2 dagen na mekaar.

Ontbijtbuffet was tip top in orde.

Niet goedkoop.
Experimental data

Train and test split have been created for all SemEval datasets

- Focus on Dutch (restaurant reviews)
  - 300 reviews for training (development)
  - 100 reviews for testing (held-out)

- Explain the pipeline we developed
- State of the art approaches and results on English (restaurant reviews)
Overview

① Introduction
② Task Definition
③ Datasets and Annotation
④ Subtasks
  ➢ Aspect Term Extraction
  ➢ Aspect Term Categorization
  ➢ Aspect Term Polarity Classification
⑤ Challenges
⑥ Conclusion
Pipeline for Dutch: overview

Tasty **paella**, but rude **waiter**.

**ASPECT TERM EXTRACTION**
- Subjectivity Heuristic
- Term Extraction with TExSIS
  - Preprocessing (LeTs)
  - Termhood
  - Unithood
  - Additional Filtering

**ASPECT CATEGORY CLASSIFICATION**
- Features for category classification
  - **Lexical**
    - Bag-of-words
  - **Semantic**
    - Cornetto
    - DBpedia
    - Semantic roles

**ASPECT POLARITY CLASSIFICATION**
- Features for polarity classification
  - **Lexical**
    - Token and character n-grams
    - Sentiment lexicons
    - Word-shape

**FOOD_quality**
**SERVICE_general**
Aspect Term Extraction

**Extract** all aspect expressions of the entities.

- Subjectivity Heuristic
- Term Extraction with TExSIS
  - Preprocessing (LeTs)
  - Termhood
  - Unithood
  - Additional Filtering

Only when subjective!

Lexicons
- Pattern (De Smedt & Daelemans, 2012)
- Duoman (Jijkoun & Hofman, 2009)

TExSIS = hybrid system combining linguistic and statistical information (Macken et al., 2013)

**Linguistic** = which words?
- Preprocessing using LeTs (Van de Kauter et al., 2013)
- PoS patterns (i.e. nouns, noun phrases)

**Statistical** = are they terms?
- Termhood, unithood measures (LL, c-value)

Additional filtering...
After a [good appetizer] our [mother] ordered a [pizza margherita], which was divine!

...Additional filtering

- Subjectivity (based on same lexicons)
- Semantic
  - DBPedia (Mendes et al. 2011): tag terms with DBPedia Spotlight and look for categories.
After a good [appetizer] our mother ordered a [pizza margherita], which was divine!
Aspect Term Extraction

Results

Training data split in devtrain (250) and devtest (50)

Best setting on held-out test set (100).

Evaluation metrics: precision, recall and F-1

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TExSIS</td>
<td>24.78</td>
<td>39.61</td>
<td>30.48</td>
</tr>
<tr>
<td>TExSIS + subj</td>
<td>29.15</td>
<td>66.18</td>
<td>40.47</td>
</tr>
<tr>
<td>TExSIS + subj + sem</td>
<td>37.85</td>
<td>59.42</td>
<td>46.24</td>
</tr>
<tr>
<td>Held-out</td>
<td>35.87</td>
<td>58.18</td>
<td>44.38</td>
</tr>
</tbody>
</table>
Aspect Term Extraction

State of the art English

Supervised machine learning approaches most successful
Sequential labeling task (IOB2 annotation ~ NER)

*Toh and Su (2016) = top system*

- CRF classifier
- NE features
- Additional features from RNN (Liu, Joty & Meng, 2015)
- 72.34
Categorize all extracted aspect expressions.

**Aspect Category Classification**

Features for category classification

- **Lexical**
  - Typical bag-of-words: token unigram

- **Lexico-semantic**
  - Cornetto (in synset or hypernym/hyponym of main cats)
  - DBPedia (belong to unique categories)

- **Semantic roles**
  - Term evokes semantic role, which role
    *(The food *tasted* good vs The food just *cost* too much)*

Classification task

- Predefined categories
- Multiclass problem:
  - Main categories
  - Subcategories
Aspect Term Categorization

Results

Ten-fold cross validation on training data. LibSVM

*Round 1*: gradually adding more features

*Round 2*: joint optimization, feature groups vs individual features

Best results on held-out test

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Round 1</th>
<th>Round 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>bow</strong></td>
<td>53.28</td>
<td>54.69</td>
</tr>
<tr>
<td><strong>bow + lexsem</strong></td>
<td>60.72</td>
<td>62.94</td>
</tr>
<tr>
<td><strong>bow + srl</strong></td>
<td>54.80</td>
<td>56.16</td>
</tr>
<tr>
<td><strong>bow + lexsem + srl</strong></td>
<td>60.01</td>
<td>62.89</td>
</tr>
<tr>
<td><strong>Held-out</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Aspect Term Categorization

State of the art English
Supervised machine learning approaches most successful

Toh and Su (2016) = top system
• Individual binary classifiers trained on each category (combined)
• Lexical bag of words (unigram, bigram)
• Lexical-semantic: clusters learned from large reference corpus
• Additional features from CNN (Severyn & Moschitti, 2015)
• 73.031
Aspect Polarity Classification

Determine whether opinion is POS | NEG | NEUTRAL

Features for polarity classification

Lexical
- Token and character n-grams
- Sentiment lexicons
- Word-shape

Three-way classification

Token and character n-gram features
unigram, bigram and trigram (tok) & trigram, fourgram (char)

Sentiment lexicon
- DuoMan and Pattern lexicon, matches pos, neg, neut

Word-shape
- UGC characteristics, character of punctuation flooding (coooooool!!!!!!), last token has punct, capitalized tokens
Aspect Polarity Classification

Results
Ten-fold cross validation on training data. LibSVM
Default: all features
Joint optimization: individual feature selection
Best results on held-out test set

Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Default</th>
<th>Joint optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All features</strong></td>
<td>76.40</td>
<td><strong>79.06</strong></td>
</tr>
<tr>
<td><strong>Held-out</strong></td>
<td>76.40</td>
<td><strong>81.23</strong></td>
</tr>
</tbody>
</table>
Aspect Polarity Classification

State of the art English
Supervised machine learning approaches most successful

*Brun, Perez & Roux (2016)* = top system

- Ensemble classifiers
- Syntactic parser = basic features (prepro + NER + syntax)
- Semantic component added (based on designated polarity & semantic lexicons)
- 88.126
Acceptable results for English on all three subtasks.

Dutch: subtasks 1 and 2 still quite challenging

Same true for other languages or other domains!!

Note:
In reality, these are not separate tasks ➔ error percolation

*e.g. for Dutch polarity classification, accuracy drops to 39.70*
Overview

① Introduction
② Task Definition
③ Datasets and Annotation
④ Subtasks
  ➢ Aspect Term Extraction
  ➢ Aspect Term Categorization
  ➢ Aspect Term Polarity Classification
⑤ Challenges
⑥ Conclusion
Domain adaptation

Focus on consumer reviews

• Product-oriented
• Aspect expressions: nouns or nouns phrases
• Will almost always include an opinion

In reality

• Non-opinionated text co-occurs with opinionated text (skewed)
• Verbal expressions or a variety of words can be used to refer to certain aspects. E.g. political tweets, discussion forums, ...
User-generated content

• Different from standard text.
• Highly expressive: emoticons, flooding (cooooool!!)

BUT

• Full of misspellings, grammatical errors, abbreviations, ...
  ➔ hinder standard NLP tools.

⇒ polarity classification: importance of lexical features

⇒ Its so keeeewl, lol
User-generated content

- Normalization (Van Hee et al., under review)

➡ Helps, especially for unseen data
Creative language use

It was so nice of my dad to come to my graduation party #not
Going to the dentist for a root canal. Yay, can’t wait!!!!

• Sarcasm, irony, humour and metaphor.
• NLP = difficult to interpret this

➡ Interesting research emerging. SemEval 2015 task on irony (Ghosh et al., 2015), however too much focus on hashtags. Van Hee et al. (2016) propose alternative: paper to appear at COLING 2016.
“Sentiment analysis requires a deep understanding of the explicit and implicit, regular and irregular, and syntactic and semantic language rules.” (Cambria et al., 2013)

• Explicit sentiment: seems easy but words are never used in isolation
  – Negation, modifiers (intensifiers, diminishers, ...) ➔ crucial!
• Implicit sentiment: more complex, read between the lines. Even factual statements can evoke different opinions.
• Coreference: crucial but not much research.
Overview

① Introduction
② Task Definition
③ Datasets and Annotation
④ Subtasks
  Ø Aspect Term Extraction
  Ø Aspect Term Categorization
  Ø Aspect Term Polarity Classification
⑤ Challenges
⑥ Conclusion
Conclusion

What is aspect-based sentiment analysis?

• Task definition
• Benchmark datasets (SemEval)
• State of the art approaches (customer reviews)
• Challenges
(AB)SA is far from solved

Much more to be researched
Maybe we should cooperate

Orphée De Clercq
orphee.declercq@ugent.be
https://www.lt3.ugent.be/people/orphee-de-clercq/
@OrfeeDC
References


De Clercq, O. (2015). Tipping the scales: exploring the added value of deep semantic processing on readability prediction and sentiment analysis. PhD, Ghent University.


**Survey paper in HUSO Proceedings:**