Fine-grained Sentiment Analysis

An Overview of the Task and its Main Challenges

Orphée De Clercq

HUSO 2016 - EMOSEDE

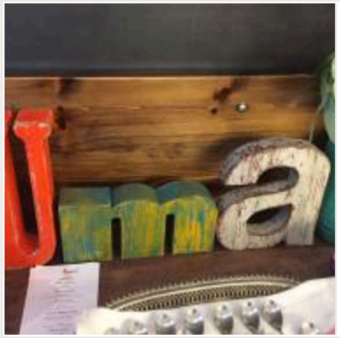
November 15, 2016

GHENT UNIVERSITY

language and translation technology team

tripadvisor

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

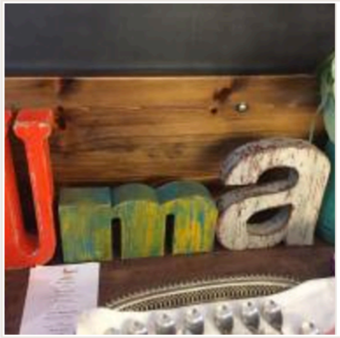
"We are far from the typical **spanish** but..." 23/11/2015
"When visiting there I always have **tapas**..." 15/05/2016

££££ | Map | Visitor photos (310)

Cuisines: Mediterranean | Spanish | Colombian | International tripadvisor

Restaurants in Barcelona

AFFORDABLE FOOD
FRIENDLY SERVICE



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

Sentiment analysis

°Early 2000s:

Wiebe (2000)

Pang et al. (2002)

…

➔ newswire text

Rise of Web 2.0 applications

2010-2016:

- 20,000 Google Scholar
- 731 papers in WoS

➔ user-generated content

Donald Trump Wins

He's the worst of America. A disgrace. Scumbag.
Vertaling bekijken

Thank u for giving up your fabulous life to help us! Lots of love from WA state ❤❤❤

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 = POS | NEG | 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).

ABSA

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)

ABSA

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

Task definition

There exist many reference works (Pang & Lee, 2008, Liu 2012, Liu 2015):

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

Reviewer X

Level 2 Contributor

5 reviews

3 restaurant reviews

2 helpful votes

"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

Report

1. Entity extraction + categorization

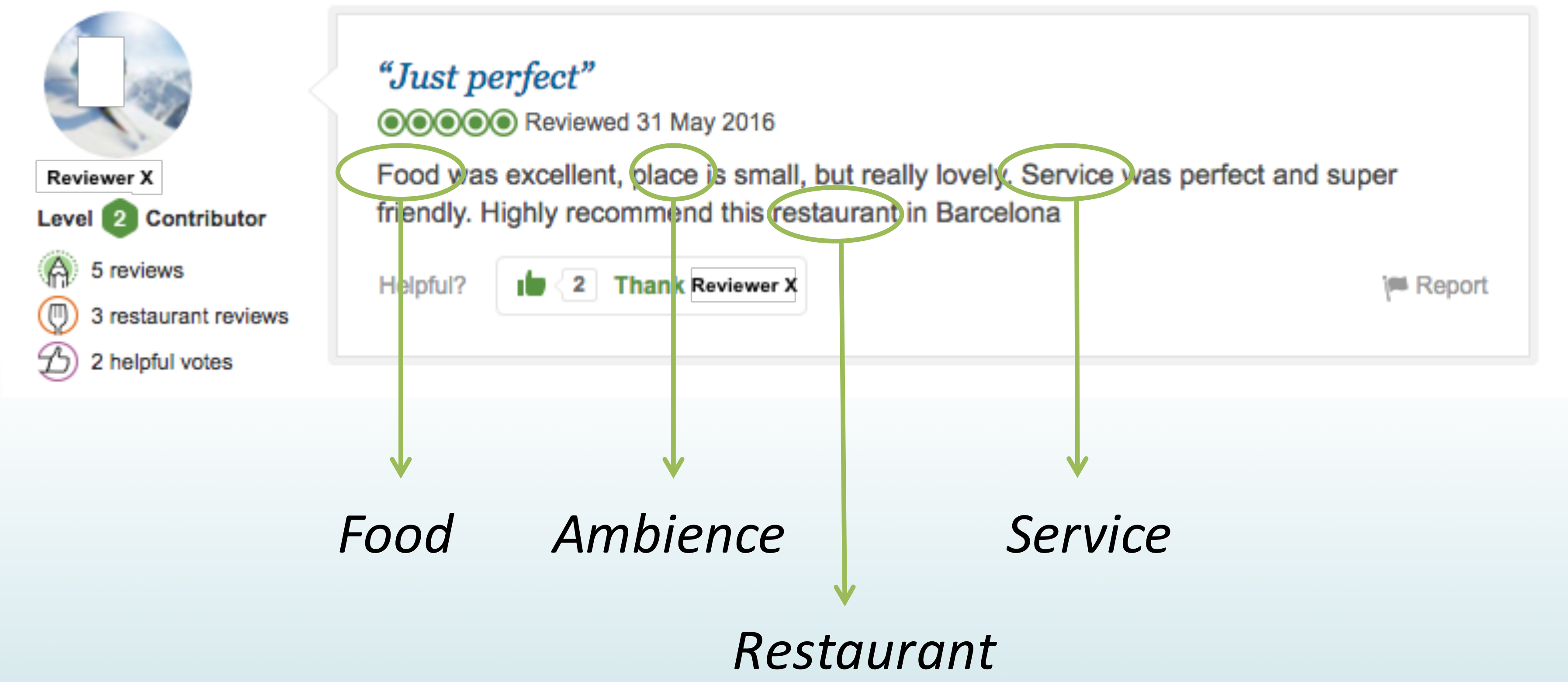
Extract all entity expressions in a document collection, and categorize or group synonymous entity expressions into entity clusters.

ENTITY

CATEGORY

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

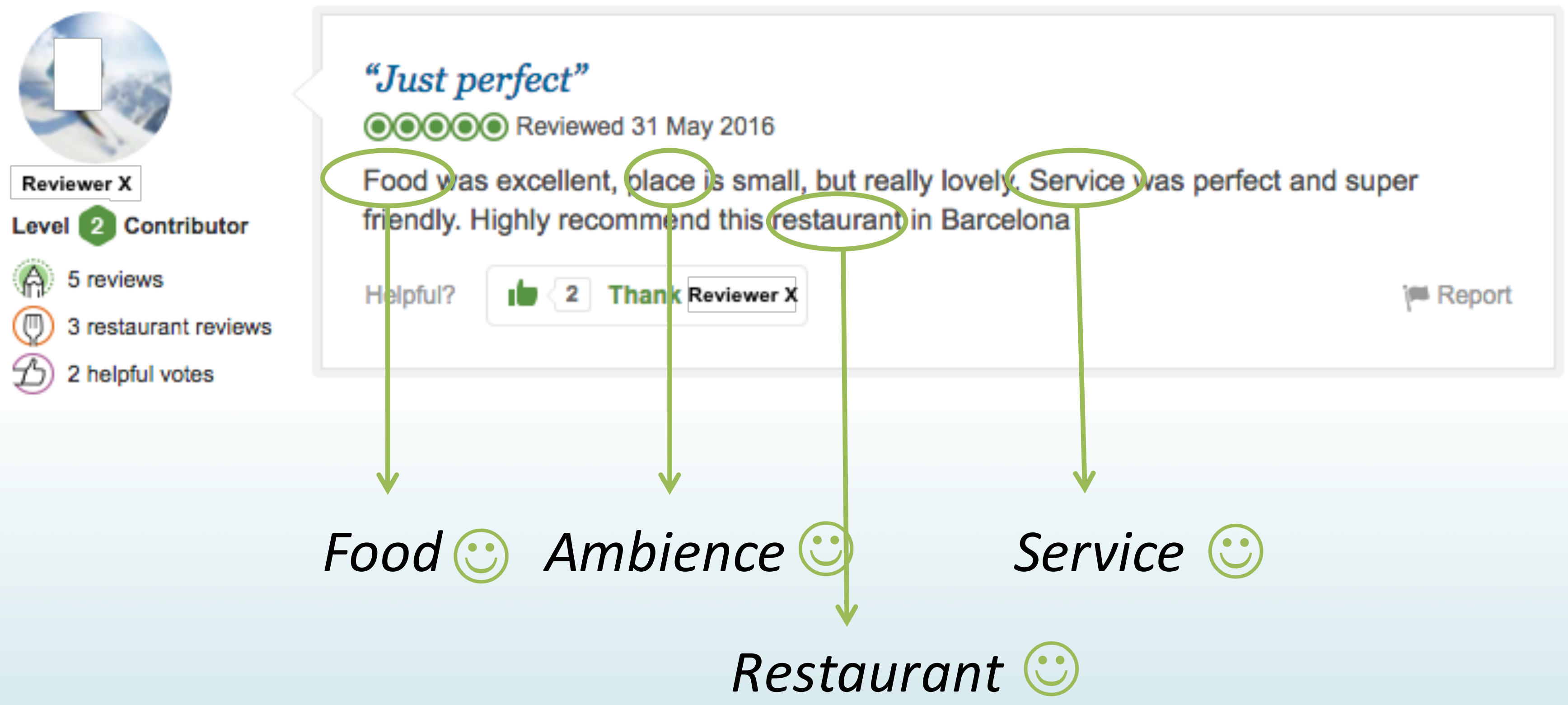
Extract opinion holders for opinions from text or structured data and categorize them.

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.



Task definition

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~~)

META-DATA

➔ABSA 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

Domain	*Subdomain*	*Language*	*#Sentences*
Electronics	Camera	Chinese	8040
	Laptops	English	3308
	Phones	Chinese	9521
	Phones	Dutch	1697
Hotels		Arabic	6029
Restaurants		Dutch	2297
		English	2676
		French	2429
		Russian	4699
		Spanish	2951
		Turkish	1248
Telecom		Turkish	3310

Annotation

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

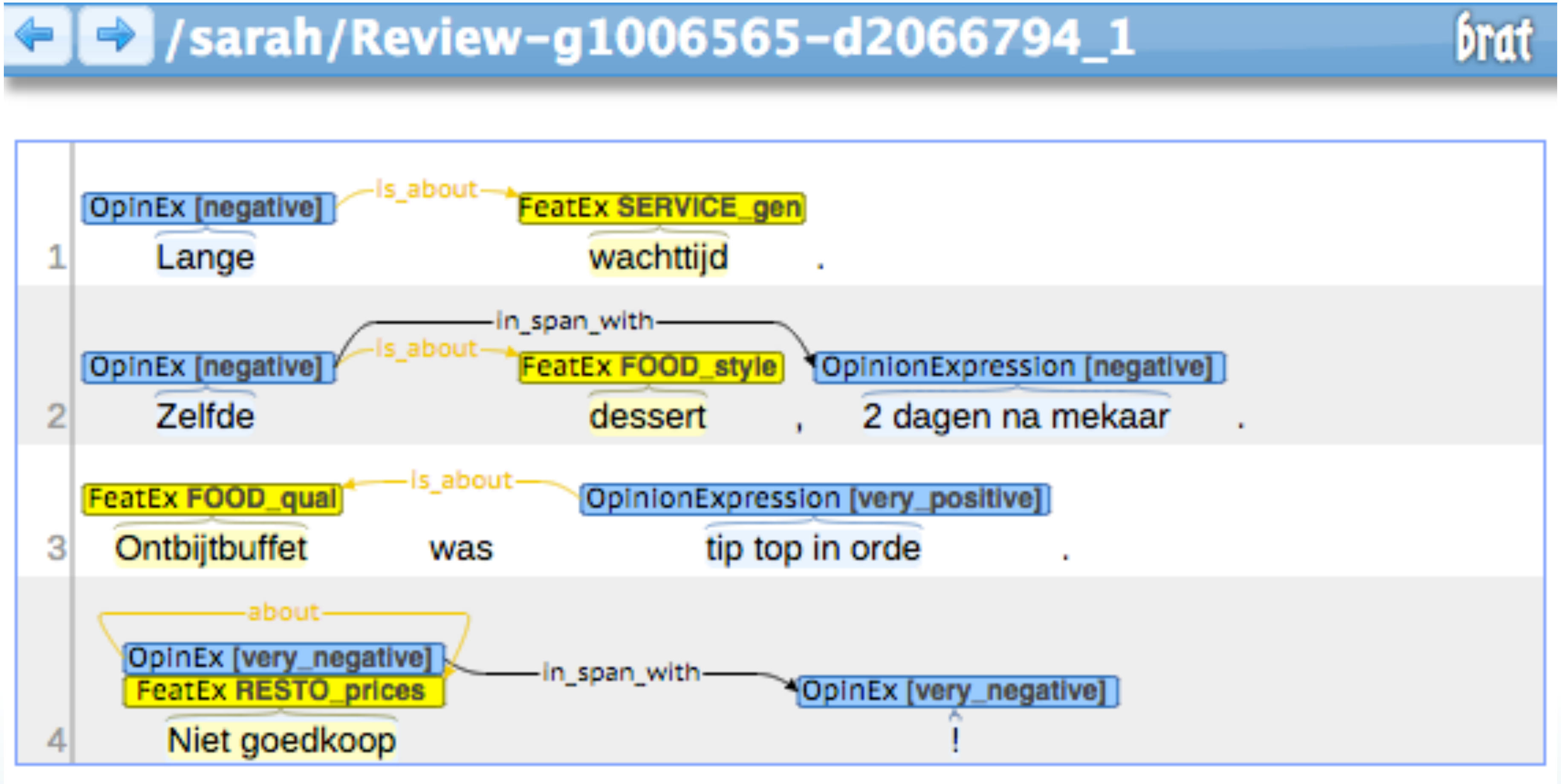
Three steps:

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.

Annotation



/sarah/Review-g1006565-d2066794_1
brat
1
OpinEx [negative]
is_about
FeatEx SERVICE_gen
Lange
wachttijd
.
2
in_span_with
OpinEx [negative]
is_about
FeatEx FOOD_style
OpinionExpression [negative]
Zelfde
dessert
,
2 dagen na mekaar
.
3
FeatEx FOOD_qual
is_about
OpinionExpression [very_positive]
Ontbijtbuffet
was
tip top in orde
.
4
about
OpinEx [very_negative]
FeatEx RESTO_prices
in_span_with
OpinEx [very_negative]
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

ASPECT TERM EXTRACTION

Subjectivity Heuristic

Term Extraction with TExSIS

- Preprocessing (LeTs)
- Termhood Unithood
- Additional Filtering

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

ASPECT CATEGORY CLASSIFICATION

Features for category classification

Lexical
- Bag-of-words

Semantic
- Cornetto
- DBpedia
- Semantic roles

paella → FOOD_quality
waiter → SERVICE_general

ASPECT POLARITY CLASSIFICATION

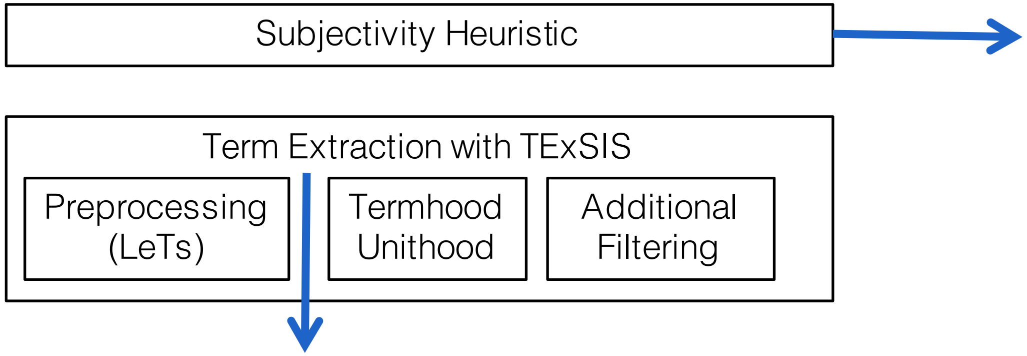
Features for polarity classification

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

FOOD_quality
SERVICE_general

Aspect Term Extraction

Extract all aspect expressions of the entities.



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…

Aspect Term Extraction

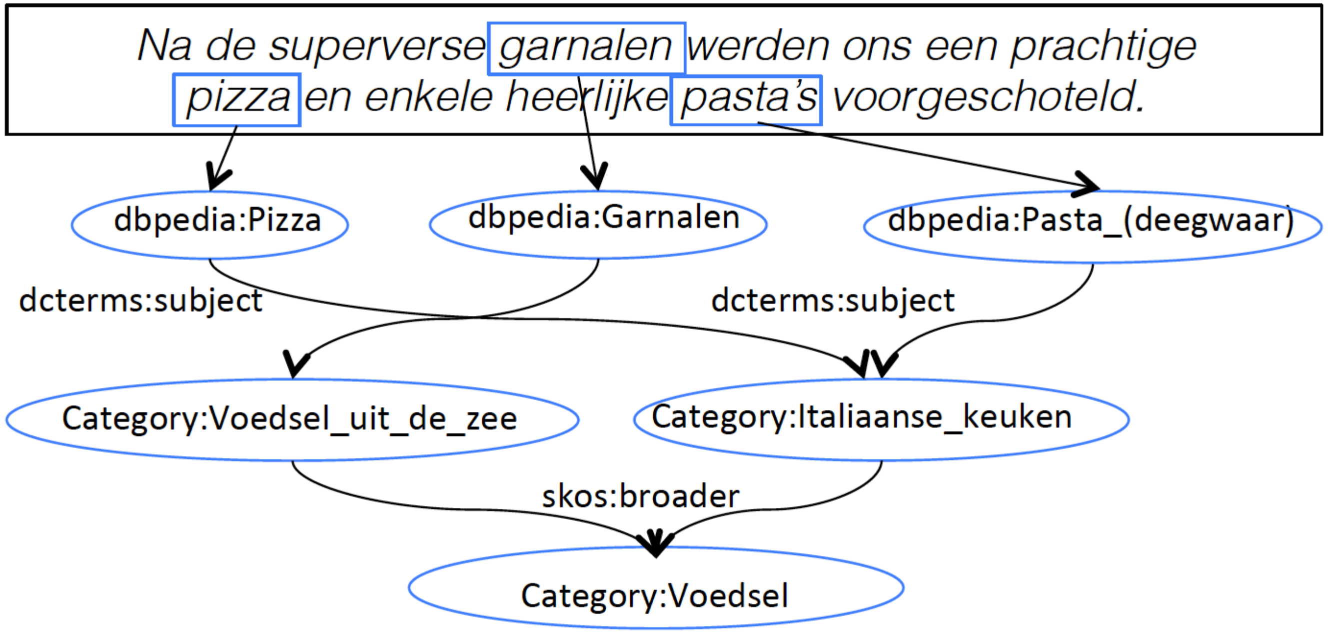
TExSIS output:

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

…Additional filtering

- Subjectivity (based on same lexicons)
- Semantic
  - Cornetto (Vossen et al. 2013): synsets look for hypernym-synonym links.
  - DBPedia (Mendes et al. 2011): tag terms with DBPedia Spotlight and look for categories.

Aspect Term Extraction



Additional filtering output:

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

	Precision	Recall	F-1
TExSIS	24.78	39.61	30.48
TExSIS + subj	29.15	**66.18**	40.47
TExSIS + subj + sem	**37.85**	59.42	**46.24**
Held-out	35.87	58.18	44.38

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

Aspect Term Categorization

Categorize all extracted aspect expressions.

ASPECT CATEGORY CLASSIFICATION

Features for category classification

Lexical
- Bag-of-words

Semantic
- Cornetto
- DBpedia
- Semantic roles

Classification task
- Predefined categories
- Multiclass problem:
  - Main categories
  - Subcategories

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*)

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

Accuracy

	Round 1	*Round 2*	
bow	53.28	54.69	
		Joint optimization	
		featgroups	indfeats
bow + lexsem	**60.72**	**62.94**	63.16
bow + srl	54.80	56.16	56.70
bow + lexsem + srl	60.01	62.89	**63.27**
Held-out			**66.42**

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 (cooooool!!!!!), 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

	Default	Joint optimization
All features	76.40	**79.06**
Held-out		**81.23**

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

ABSA

➔ 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 (*cooool!!*)

BUT

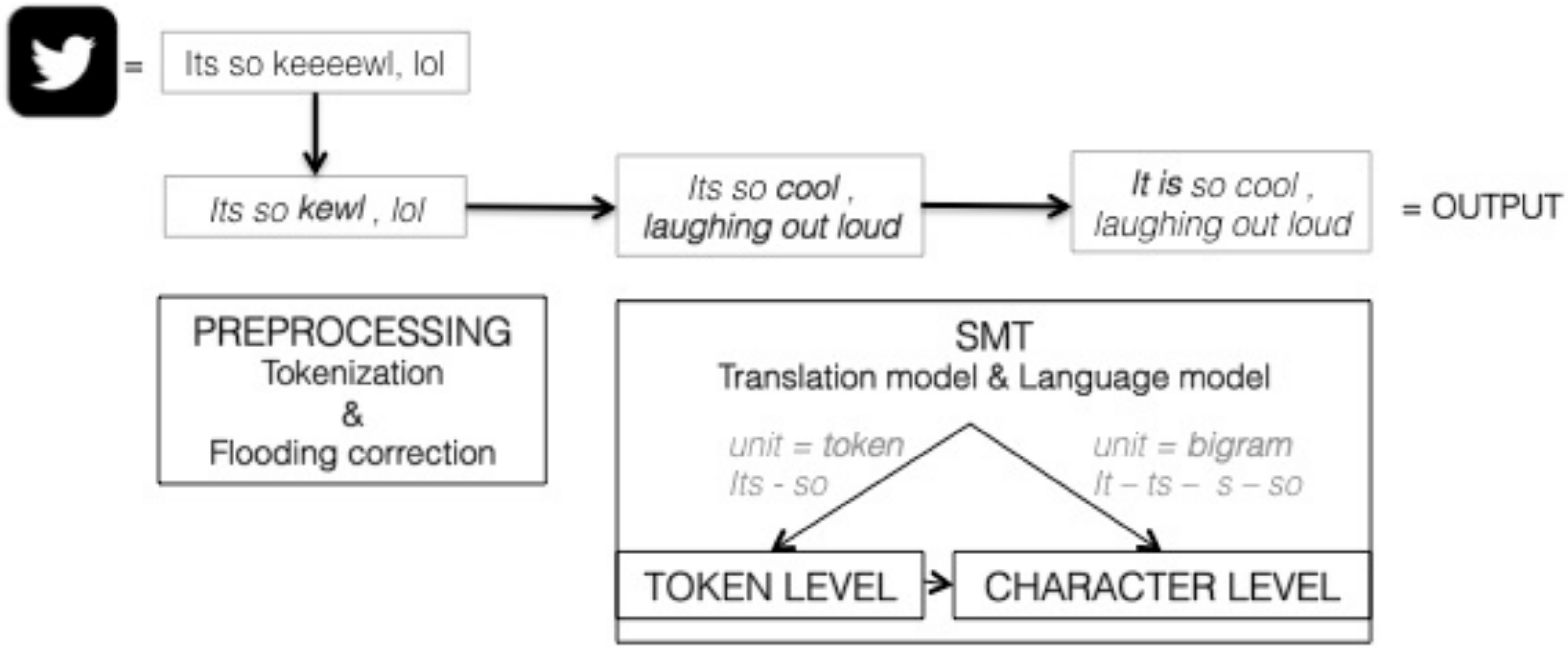
- Full of misspellings, grammatical errors, abbreviations, … ➔ hinder standard NLP tools.

Its so keeeeewl, lol

➔ polarity classification: importance of lexical features

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.

Requires deep understanding

"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

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

Bermingham, A., & Smeaton, A. F. (2011). On using twitter to monitor political sentiment and predict election results. *Psychology*, 2-10.

Brody, S. & Elhadad, N. (2010). An unsupervised aspect-sentiment model for online reviews, Proceedings of NAACL-2010, pp. 804-812.

Brun, C., Perez, J. & Roux, C. (2016). XRCE at SemEval-2016 Task 5: Feedbacked Ensemble Modeling on Syntactico-Semantic Knowledge for Aspect Based Sentiment Analysis, Proceedings of SemEval-2016, pp. 277–281.

Cambria, E., Schuller, B., Xia, Y. & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis, *IEEE Intelligent Systems*, vol. 28, no. 2, 2013, pp. 15–21.

Dabrowski, M., Acton, T., Jarzebowski, P. & O'Riain, S. (2010). Improving customer decisions using product reviews - CROM - Car Review Opinion Miner, in Proceedings of WEBIST-2010, pp. 354–357.

De Clercq, O., Van de Kauter, M., Lefever, E., & Hoste, V. (2015). Applying hybrid terminology extraction to aspect-based sentiment analysis. Proceedings of SemEval 2015, pp. 719–724.

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

De Clercq, O., & Hoste, V. (2016). Rude waiter but mouthwatering pastries! An exploratory study into Dutch aspect-based sentimesnt analysis, Proceedings of LREC 2016, pp. 2910–2917.

De Smedt, T. and Daelemans, W. (2012). Vreselijk mooi! Terribly beautiful: a subjectivity lexicon for Dutch adjectives, Proceedings of LREC-2012 , pp. 3568-3572.

Desmet, B. (2014) “Finding the online cry for help: automatic text classification for suicide prevention,” PhD, Ghent University.

Ganu, G., Elhadad, N. & Marian, A. (2009). Beyond the stars: improving rating predictions using review text content, Proceedings of WebDB-2009, pp. 1-6.

Ghosh, A., Li, G., Veale, T., Rosso, P, Shutova, E., Barnden, J. & Reyes, A. (2015). Semeval-2015 task 11: Sentiment analysis of figurative language in twitter, in Proceedings of SemEval 2015, pp. 470–478.

Hu, M. and Liu, B.(2004). Mining and summarizing customer reviews, Proceedings of KDD-2004, pp. 168-177.

Jijkoun, V. and Hofmann, K. (2009). Generating a non-English subjectivity lexicon: Relations that matter, Proceedings EACL-2009, pp. 398-405.

Liu, B. (2012). Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167.

Liu, B. (2015) Sentiment Analysis - Mining Opinions, Sentiments, and Emotions. Cambridge University Press.

Liu, P., Joty, S. & Meng, H. (2015). Fine-grained opinion mining with recurrent neural networks and word embeddings, Proceedings EMNLP-2015, pp. 1433–1443.

O'Hare, N. ,Davy, M., Bermingham, A. , Ferguson, P. Sheridan, P. Gurrin, C. & Smeaton, A.F. (2009). Topic-dependent sentiment analysis of financial blogs, Proceedings of TSA-2009, pp. 9–16.

Macken, L., Lefever, E. & Hoste, V. (2013). TExSIS: Bilingual Terminology Extraction from Parallel Corpora Using Chunk-based Alignment, *Terminology* 19 (1), pp. 1-30.

Mendes, P. N., Jakob, M., Garca-Silva, A. & Bizer, C. (2011). DBpedia Spotlight: Shedding light on the web of documents, Proceedings of the 7th International Conference on Semantic Systems (I-Semantics-2011) , pp. 1-8.

Pang, B., Lee, L. & Vaithyanathan, S (2002). Thumbs up?: Sentiment classification using machine learning techniques, Proceedings of EMNLP-2002,pp. 79-86.

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos I. & Manandhar, S. (2014). Semeval-2014 task 4: Aspect based sentiment analysis, Proceedings of SemEval-2014, pp. 27–35.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S. & Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis, Proceedings of SemEval-2015, pp. 486–495.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S.M. & Eryiğit, G. (2016). Semeval-2016 task 5: Aspect based sentiment analysis, Proceedings of SemEval-2016, pp. 19–30.

Severyn, A. & Moschitti, A. (2015) UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification, Proceedings of SemEval-2015, pp. 464–469.

Sprenger, T. O., Tumasjan, A. ,Sandner P. G. & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs, *European Financial Management*, vol. 20, no. 5, pp. 926–957.

Thet, T. T., Na, J.-C. and Khoo, C. S. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards, *Journal of Information Science* 36 (6), 823-848.

Toh, Z. & Su, J. (2016). NLANGP at SemEval-2016 Task 5: Improving Aspect Based Sentiment Analysis using Neural Network Features, Proceedings of SemEval 2016, pp. 282–288.

Van de Kauter, M., Coorman, G., Lefever, E., Desmet, B., Macken, L. & Hoste, V. (2013), LeTs Preprocess: The multilingual LT3 linguistic preprocessing toolkit, *Computational Linguistics in the Netherlands Journal* 3 , 103- 120.

Van Hee, C., Van de Kauter, M., De Clercq, O., Lefever, E., & Hoste, V. (2014). LT3: sentiment classification in user-generated content using a rich feature set, Proceedings of SemEval 2014, pp. 406–410.

Van Hee, C., Lefever, E., & Hoste, V. (2016). Exploring the realization of irony in Twitter data, Proceedings LREC 2016, pp. 1795–1799.

Van Hee, C., Lefever, E., & Hoste, V. (2016). Monday morning are my fave :) #not. Exploring the automatic recognition of irony in English tweets. To appear at COLING 2016.

Vossen, P. (1998) EuroWordNet: a multilingual database with lexical semantic networks for European Languages , Kluwer, Dordrecht.

Zabin, J. & Jefferies, A. (2008). Social media monitoring and analysis: Generating consumer insights from online conversation, Aberdeen Group Benchmark Report, Aberdeen Group, Tech. Rep.

Survey paper in HUSO Proceedings:

De Clercq, O. (2016). The Many Aspects of Fine-Grained Sentiment Analysis: An Overview of the Task and its Main Challenges, Proceedings of the International Conference on Human and Social Analytics (HUSO).