Lekker, yummy, délicieux

Fine-grained sentiment analysis of customer reviews

Orphée De Clercq

3rd Belgium NLP Meetup

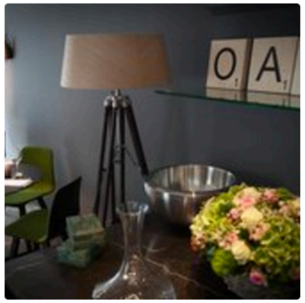
May 18, 2017

GHENT UNIVERSITY

language and translation technology team

tripadvisor

Restaurants in Gent



Oak Restaurant

315 beoordelingen

nummer 1 van 723 Restaurants in Gent

€€€€, Belgisch, Europees, Geschikt voor vegetariërs, Glutenvrije opties, V...

"heerlijke lunch in ontspannen sfee ..." 07/05/2017

"Verfijning in een ruige buurt." 03/05/2017

Vrijmoed

355 beoordelingen

nummer 2 van 723 Restaurants in Gent

€€€€, Belgisch, Europees, Geschikt voor vegetariërs, Veganistische opties,

"Fantastische ervaring" 12/05/2017

"Aangename lunch" 10/05/2017

Soup'R

176 beoordelingen

nummer 3 van 723 Restaurants in Gent

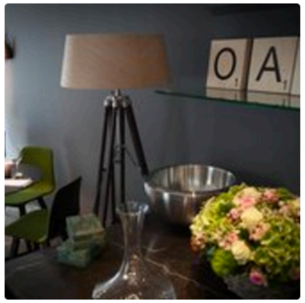
€, Belgisch, Europees, Soepen, Geschikt voor vegetariërs, Veganistische ...

"Uitstekende kwaliteit en aanbod aa ..." 14/05/2017

"Fijn tentje!" 07/04/2017

tripadvisor

Restaurants in Gent



Oak Restaurant

315 beoordelingen

nummer 1 van 723 Restaurants in Gent

€€€€, Belgisch, Europees, Geschikt voor vegetariërs, Glutenvrije opties, V...

"heerlijke lunch in ontspannen sfee ..." 07/05/2017

"Verfijning in een ruige buurt." 03/05/2017

AFFORDABLE FOOD
FRIENDLY SERVICE

Vrijmoed

355 beoordelingen

nummer 2 van 723 Restaurants in Gent

€€€€, Belgisch, Europees, Geschikt voor vegetariërs, Veganistische opties,

"Fantastische ervaring" 12/05/2017

"Aangename lunch" 10/05/2017

Soup'R

176 beoordelingen

nummer 3 van 723 Restaurants in Gent

€, Belgisch, Europees, Soepen, Geschikt voor vegetariërs, Veganistische ...

"Uitstekende kwaliteit en aanbod aa ..." 14/05/2017

"Fijn tentje!" 07/04/2017

Sentiment analysis

°Early 2000s:

Wiebe (2000)

Pang et al. (2002)

…

➔ newswire text

Rise of Web 2.0 applications

2010-2017:

- 21,000 Google Scholar
- 1,143 papers in WoS

➔ user-generated content

Trump revealed highly classified information to Russian diplomats

Replying to @realDonaldTrump

Bullshit. You committed treason and we will not rest until you are in jail. #Blabbergate #clownpresident #Russians

I love trump, so this is a 100% lie even though they have confirmed sources #Blabbergate

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 level = POS | NEG | NEUT

→ 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).

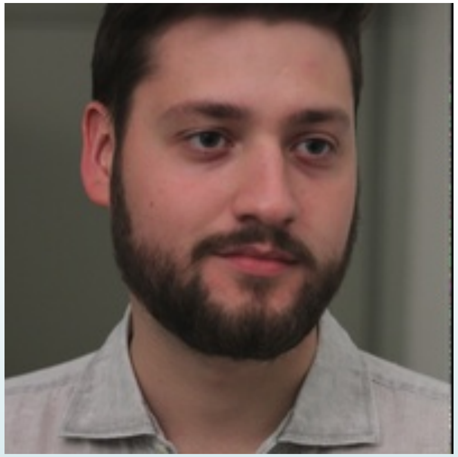
ABSA

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
- 2017: valorisation project and continuing research (various domains, languages)

ABSA

The best research = team research



Overview

① Introduction

② ABSA Task Definition

③ Datasets and Annotation

④ Subtasks

- S1: Aspect Extraction
- S2: Aspect Categorization
- S3: Aspect Polarity Classification

⑤ Challenges

⑥ Conclusion

Overview

① Introduction

② ABSA Task Definition

③ Datasets and Annotation

④ Subtasks

- S1: Aspect Extraction
- S2: Aspect Categorization
- S3: Aspect Polarity Classification

⑤ Challenges

⑥ Conclusion

ABSA Task Definition

1. What are they talking about?
2. How do they feel about these aspects?

ABSA Task Definition

1. What are they talking about?
   - Aspect Extraction (*pizza margherita, place, waiter*)
   - Aspect Categorization (*Food, Ambience, Service*)

2. How do they feel about these aspects?
   - Aspect Sentiment Classification (*POS*|*NEG*|*NEUT*)

➔ 3 subtasks = SemEval Task Description

(Pontiki et al., 2014, 2015, 2016)

Overview

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

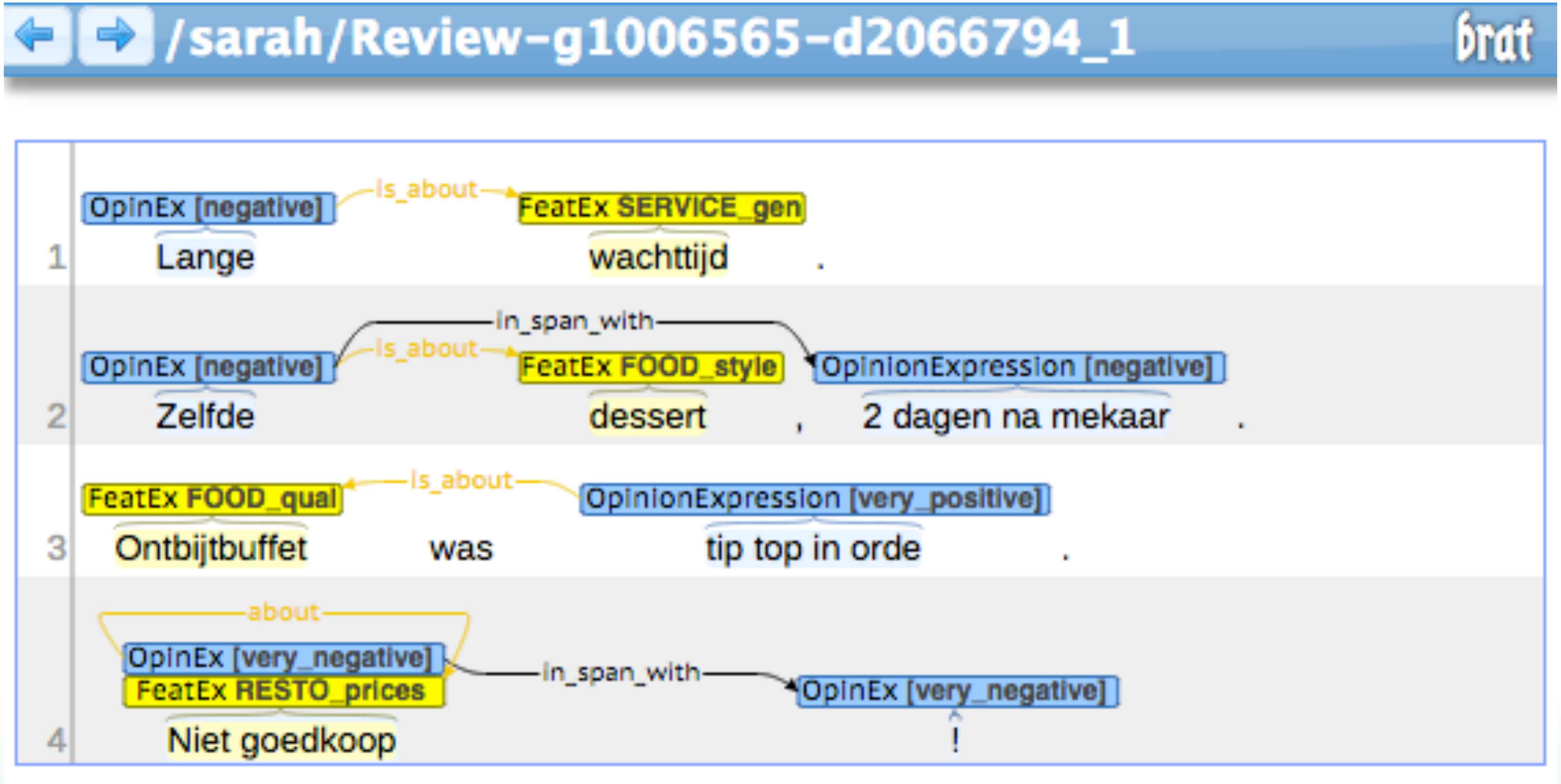
➔ Three runs of the task (2014, 2015 & **2016**)

➔ 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

➔ Annotated using the same guidelines (http://goo.gl/wOf1dX)

Annotation



/sarah/Review-g1006565-d2066794_1
brat
1 OpinEx [negative] Is_about FeatEx SERVICE_gen
Lange wachttijd .
2 OpinEx [negative] in_span_with 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 splits 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
   - S1: Aspect Extraction
   - S2: Aspect Categorization
   - S3: Aspect Polarity Classification
5. Challenges
6. Conclusion

Pipeline for Dutch: overview

ASPECT 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

S1: Aspect Extraction

Extract all aspect expressions of the entities.

Subjectivity Heuristic

Only when subjective!
Lexicons
- Pattern (De Smedt & Daelemans, 2012)
- Duoman (Jijkoun & Hofman,2009)

Term Extraction with TExSIS

Preprocessing (LeTs) | Termhood Unithood | Additional Filtering

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)

+ some additional linguistic filtering

S1: Aspect 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

S1: 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

→ Inspired + developed CRF classifier for Dutch = better results

S2: Aspect Categorization

Categorize all extracted aspect expressions

ASPECT CATEGORY CLASSIFICATION

Features for category classification

Lexical
- Bag-of-words

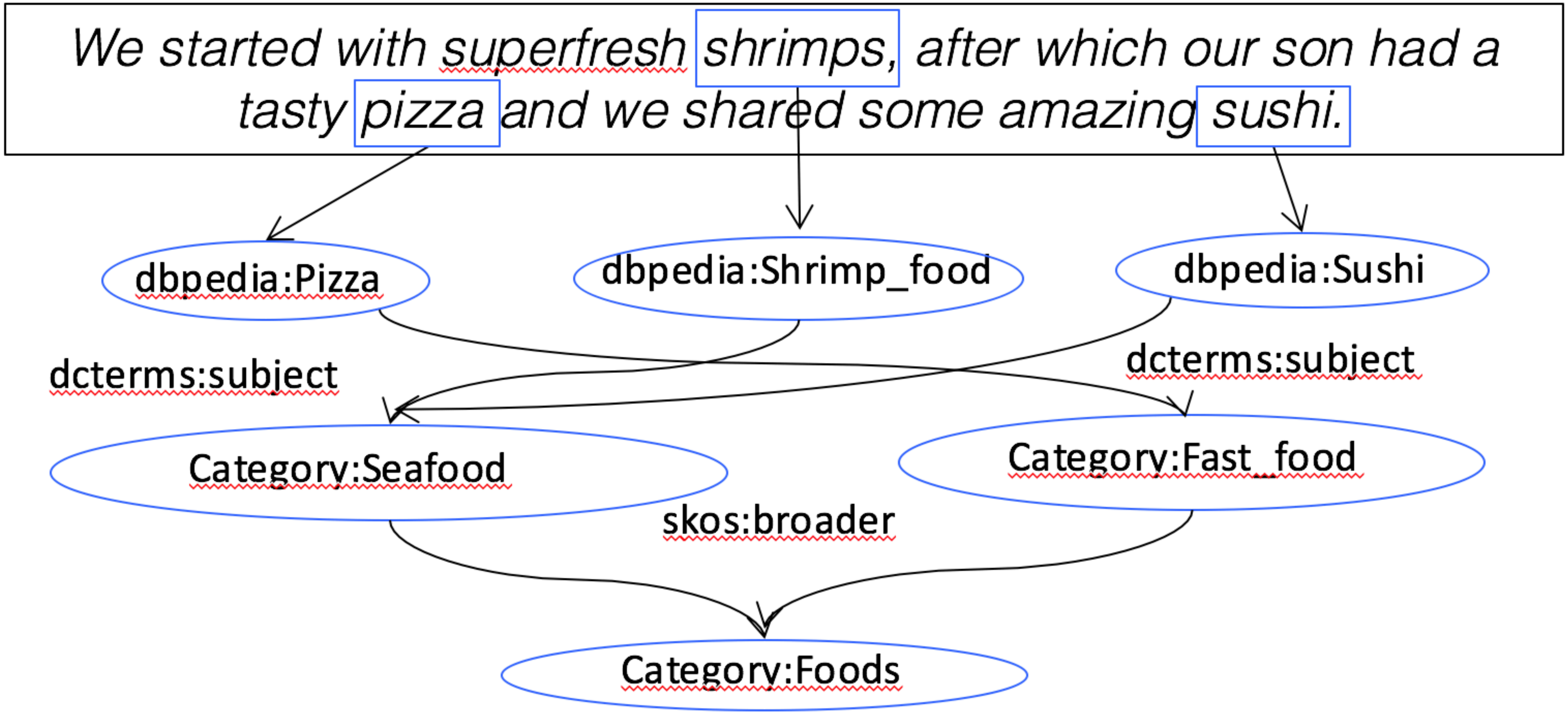
Semantic
- Cornetto
- DBpedia
- Semantic roles

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)

S2: Aspect Categorization



S2: Aspect 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*)

S2: Aspect 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**

S2: Aspect 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

S3: 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

S3: 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**

S3: 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

Challenges

Domain adaptation

Multilingual society

User-generated content

Creative language use

Requires deep understanding of natural language

Challenges

Domain adaptation

Multilingual society

User-generated content

Creative language use

Requires deep understanding of natural language

Domain adaptation

Valorisation project (= ABSA on customer feedback)

- ➢ 3 different domains: retail, banking, HR (= Dutch)
- ➢ Annotated training data (ca. 1,000 verbatims)

- Subtask 1 = domain-independent
- Subtask 2 = mapping between domains
- Subtask 3 = partly domain-independent

➔ “There’s no data like more data”

Domain adaptation

Focus on consumer reviews or customer feedback

- 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, ...

Multilingual society

Master’s thesis: English – Dutch – French (SemEval)

~ linguistic analysis

- Aspect terms:
  - All languages mostly use nouns to refer to explicit aspects or
  - MWUs
    - salade composée, grilled potato, gefrituurde pijlinktvis

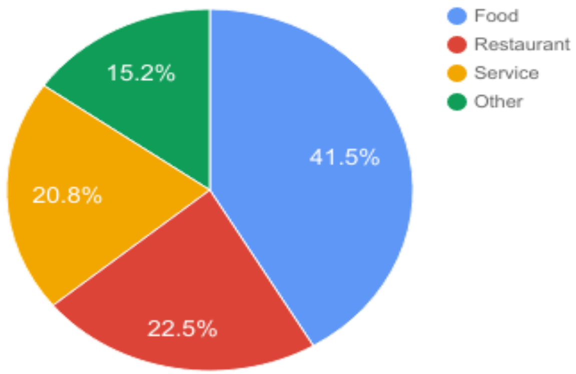
  ➔ English, on average more MWUs

Multilingual society

Master's thesis: English – Dutch – French (resto)

~ linguistic analysis

- Aspect categories:

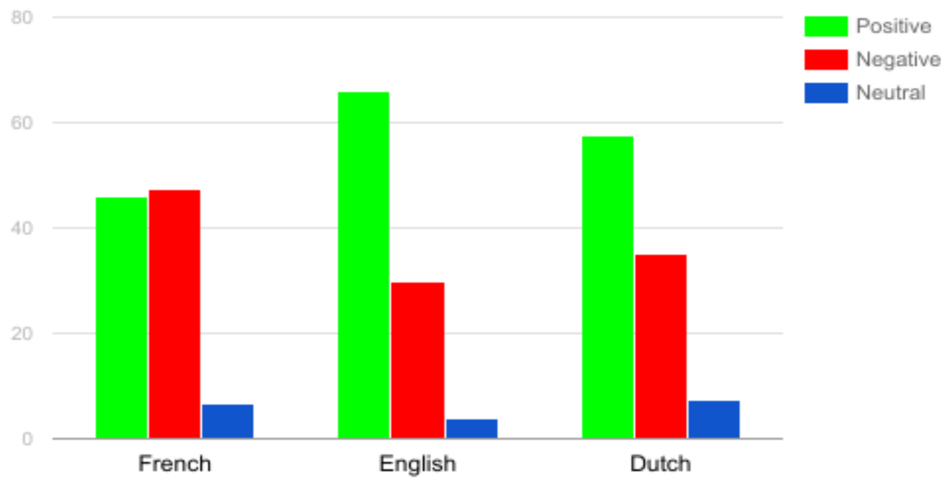


Multilingual society

Master's thesis: English – Dutch – French (resto)

~ linguistic analysis

- Aspect categories:



"Lekker, yummy, affreux"

Multilingual society

Master's thesis: English – Dutch – French (resto)

~ linguistic analysis

- Lexicon-based approach (subtask 1 & 2)
- Deadline = 30 May

What about other languages?

Can we apply the same methodology?

Language-independent system?

Challenges

Domain adaptation

Multilingual society

User-generated content

Its so keeeeewl, lol

Creative language use

Requires deep understanding of natural language

Challenges

Domain adaptation

Multilingual society

User-generated content

Creative language use

Going to the dentist for a root canal. Yay, can’t wait!!!!

Requires deep understanding of natural language

Challenges

Domain adaptation

Multilingual society

User-generated content

Creative language use

Requires deep understanding of natural language

- Explicit vs. implicit sentiment
- Coreference resolution
- …

Overview

Conclusion

What is aspect-based sentiment analysis?

- ABSA 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/

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.

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. (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).

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.

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.

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.

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.