

CAN FINE-GRAINED SENTIMENT ANALYSIS BENEFIT FROM COREFERENCE RESOLUTION?

Orphée De Clercq and Véronique Hoste
CLIN 28
26 January 2018

THIS PRESENTATION:

Anaphora or coreference resolution plays an important role in opinion mining to resolve the relationship between the mentioned entities in a given text and across texts.

(Liu, 2012, Feldman 2013)

BUT

It is not clear to what extent coreference resolution actually boosts sentiment analysis performance, if at all..

FINE-GRAINED SENTIMENT ANALYSIS

HOTEL

BLUE

Blue

●●●●● 65 Reviews

| #7 of 14 Hotels in Nijmegen

📍 Oranjesingel 14-20, 6511 NT Nijmegen, The Netherlands



Reviewer X

👍 44 🗳️ 16

●●●●●

Reviewed February 17, 2017 📱 via mobile

What a great hotel

Really liked this place. Great vibe, very friendly and relaxed service. Room I had was fantastic. Fast, free wi-Fi. You have to have breakfast about 50 metres down the road but that was also great. Has a nice bar with good range of drinks, beers and wines.

[Show less](#)

Below the sentence level → ABSA

FINE-GRAINED SENTIMENT ANALYSIS

*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 (Pontiki et al. 2016).*

1. What is the reviewer talking about? Which aspects?
2. How does the reviewer feel about these aspects?

FINE-GRAINED SENTIMENT ANALYSIS

1. What is the reviewer talking about? Which aspects?

- ❖ Aspect Extraction
- ❖ Aspect Categorization

2. How does the reviewer feel about these aspects?

- ❖ Aspect Sentiment Classification (*POS*|*NEG*|*NEUT*)

FINE-GRAINED SENTIMENT ANALYSIS

HOTEL

Blue

Blue

65 Reviews

#7 of 14 Hotels in Nijmegen

Oranjesingel 14-20, 6511 NT Nijmegen, The Netherlands



Reviewer X

44 16



Reviewed February 17, 2017 via mobile

What a great hotel

Really liked this place. Great vibe, very friendly and relaxed service. Room I had was fantastic. Fast, free wi-Fi. You have to have breakfast about 50 metres down the road but that was also great. Has a nice bar with good range of drinks, beers and wines.

Show less

Hotel



Ambience



Food



Service



Drinks_Variety



THIS PRESENTATION

1. What is the reviewer talking about? Which aspects?

❖ Aspect Extraction

❖ Aspect Categorization

➔ added value of coref?

2. How does the reviewer feel about these aspects?

❖ Aspect Sentiment Classification (*POS*|*NEG*|*NEUT*)

OVERVIEW

- 1) Introduction
- 2) Dataset and Annotation
- 3) Experimental Setup
- 4) Results
- 5) Conclusion & Future Work

OVERVIEW

1) Introduction

2) Dataset and Annotation

3) Experimental Setup

4) Results

5) Conclusion & Future Work

DATASET AND ANNOTATION

Dataset of 400 Dutch restaurant reviews (De Clercq & Hoste, 2016)

- ❖ 2,297 sentences
- ❖ 32,546 tokens

Annotations:

Step 1: ABSA (SemEval guidelines 2016)

Step 2: Coreference (identity relation)

DATASET AND ANNOTATION: ABSA

Explicit targets → annotated as such

```
... <sentence id="1074868:4">↵
... <text>The food was well prepared and the service impeccable.</text>↵
... <Opinions>↵
... <Opinion target="food" category="FOOD#QUALITY" polarity="positive" from="4" to="8"/>↵
... <Opinion target="service" category="SERVICE#GENERAL" polarity="positive" from="35" to="42"/>↵
... </Opinions>↵
... </sentence>↵
... <sentence id="1074868:5">↵
... <text>I'm going back.</text>↵
... <Opinions>↵
... <Opinion target="NULL" category="RESTAURANT#GENERAL" polarity="positive" from="0" to="0"/>↵
... </Opinions>↵
... </sentence>↵
```

Implicit targets → annotated as "NULL"

DATASET AND ANNOTATION: ABSA

Main aspect	Attribute	# Total	# Implicit
Ambience	General	240	56
Drinks	Prices	23	3
	Style & Options	38	4
	Quality	68	3
Food	General	15	4
	Prices	54	19
	Style & Options	209	27
	Quality	675	123
Location	General	34	7
Restaurant	General	347	296
	Prices	43	33
	Miscellaneous	26	12
Service	General	583	186
TOTAL		2445	773

31.6%
implicit

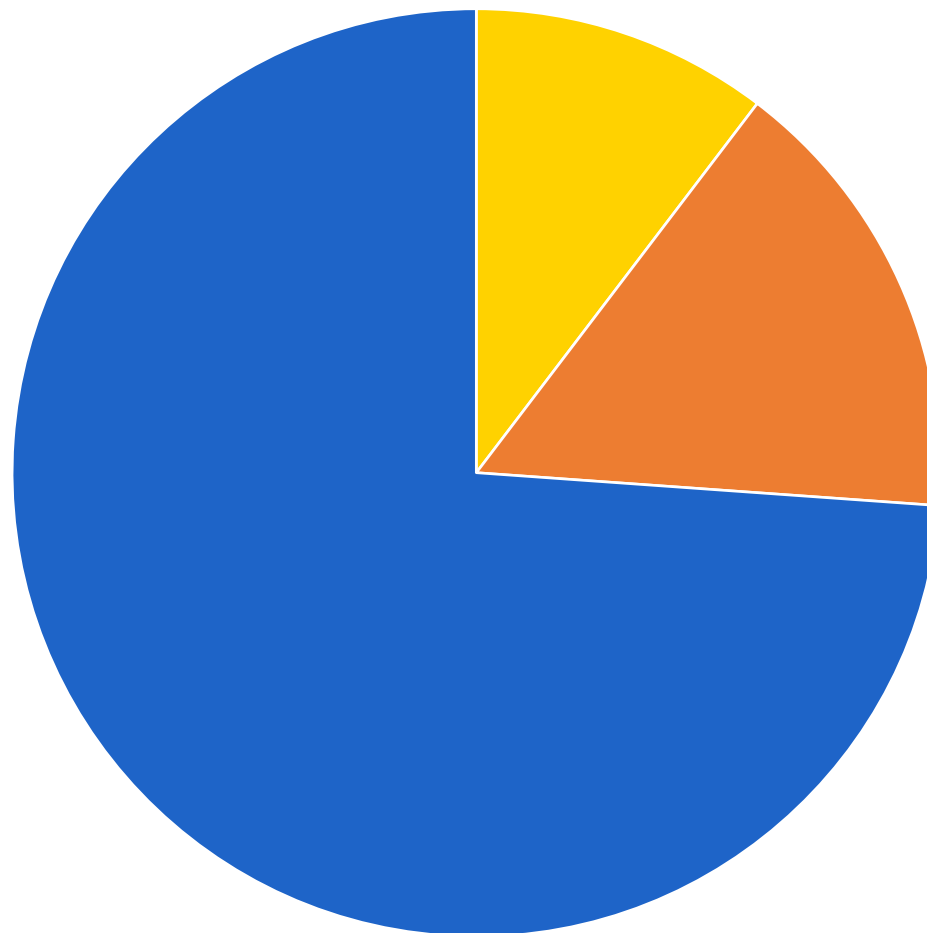
DATASET AND ANNOTATIONS: COREF

Implicit instances were annotated → three options

- (1) In La Dolce Vita we ordered a [pizza], recommended by the boss.
[It] was absolutely divine! = **COREF**
- (2) This place is incredibly tiny. [They] refuse to seat parties of 3 or more on weekends. = **EMPTY**
- (3) I'm going back. = **IMPLICIT**

DATASET AND ANNOTATIONS: COREF

- IMPLICIT (571)
- COREF (80)
- EMPTY (122)



OVERVIEW

- 1) Introduction
- 2) Dataset and Annotation
- 3) Experimental Setup**
- 4) Results
- 5) Conclusion & Future Work

EXPERIMENTAL SETUP

Aspect Categorization

- We relied on the gold standard explicit and implicit (NULL) targets
- Train and test split (~ SemEval 2016)

	TRAIN	TEST
# targets	1843	602
# implicits	563	210
# explicit	1280	392

- Supervised ML with LibSVM → multiclass classification task
- Evaluation: classification **accuracy**

EXPERIMENTAL SETUP: FEATURES

❖ Information sources (features)

- Bag-of-words token unigram features
- Semantic features:
 - Cornetto (Vossen et al. 2013)
 - Terms that co-occur in synset of target (from category)
 - Terms which are hyponym/hypernym of target in synset

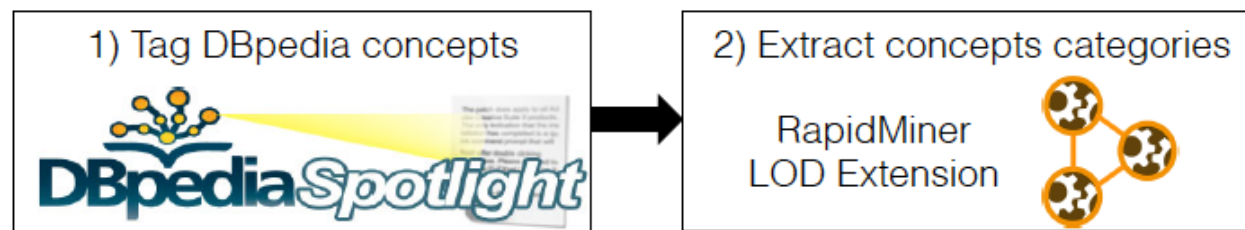
EXPERIMENTAL SETUP: FEATURES

❖ Information sources (features)

- Bag-of-words token unigram features
- Semantic features:
 - Cornetto (Vossen et al. 2013)
 - DBpedia (Lehman et al. 2013)

Tag DBpedia concepts with Spotlight (Mendes et al. 2011)

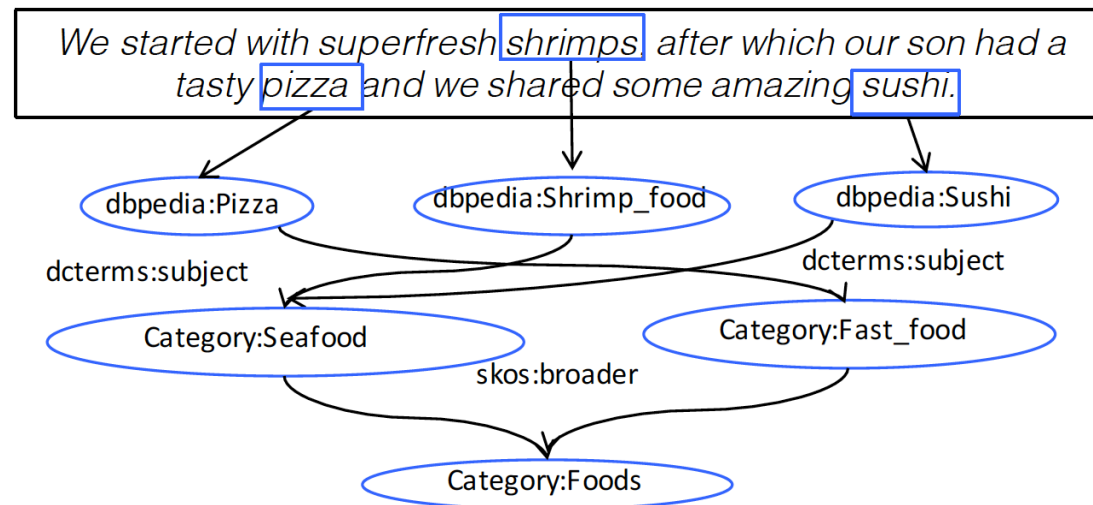
Extract concept categories with LOD Extension (Paulheim & Fürnkranz 2012)



EXPERIMENTAL SETUP: FEATURES

❖ Information sources (features)

- Bag-of-words token unigram features
- Semantic features:
 - Cornetto (Vossen et al. 2013)
 - DBpedia (Lehman et al. 2013)



EXPERIMENTAL SETUP: FEATURES

❖ Information sources (features)

- Bag-of-words token unigram features
- Semantic features:
 - Cornetto (Vossen et al. 2013)
 - DBpedia (Lehman et al. 2013)

➔ hypothesis: resolving coreferential relations prior to classification could lead to better results.

(1) In La Dolce Vita we ordered a [pizza], recommended by the boss. [It] was absolutely divine!



EXPERIMENTAL SETUP

- ❖ Step 1: Perform coreference resolution prior to classification
 - Gold-standard relations
 - Predicted relations using COREA (Hoste 2005, De Clercq et al. 2011)
- ❖ Step 2: Multiclass classification: LibSVM (10-fold cv on training)
 - Round 1: default hyperparams: bow + lexsem
 - Round 2: joint optimization with Gallop (Desmet et al. 2013):
 - ➔ Feature group selection ➔ Individual feature selection (semantic)
- ❖ Step 3: Test optimal model on held-out test set

OVERVIEW

- 1) Introduction
- 2) Dataset and Annotation
- 3) Experimental Setup
- 4) Results**
- 5) Conclusion & Future Work

RESULTS: STEP 1 & 2

❖ Without coreference resolution

	Round 1	Round 2	
bow	53.28	54.69	
		Joint optimization	
		featgroups	individu al
+ sem	60.72	62.94	63.16

❖ With coreference resolution

	Round 1	Round 2	
		Joint optimization	
		featgroups	individu al
+ sem (GOLD)	61.26	62.78	63.59
+ sem (AUTO)	59.63	60.77	60.88

RESULTS: STEP 3

Best model based on training data:

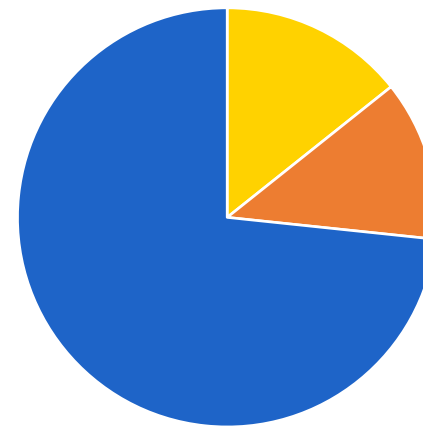
- ❖ Joint optimization: hyperparameters + individual feature selection
- ❖ Gold-standard coreference

TRAIN	TEST
63.59	66.42

→ SemEval 2016 = + 10 points

❖ Test set:

- IMPLICIT (154)
- COREF (30)
- EMPTY (26)



OVERVIEW

- 1) Introduction
- 2) Dataset and Annotation
- 3) Experimental Setup
- 4) Results
- 5) Conclusion & Future Work**

CONCLUSION

- ❖ Fine-grained sentiment analysis (ABSA) → Aspect Categorization
- ❖ Added value of resolving coreference prior to classification
 - Annotations (ABSA + truly implicit, empty, coref)
 - Joint optimization with individual feature selection gives best results:
 - Features: bag-of-words + semantic (Cornetto & DBpedia)
 - Best result on training = **63.59** accuracy
 - Result on held-out test = **66.42** accuracy
- ❖ Coreference resolution helps, however, only if **gold standard!**

FUTURE WORK

- ❖ Corroborate on other datasets:
 - Larger datasets
 - Other domains
 - Other languages (English: different results)
- ❖ Across texts → now limited within one text (one review)

References

- * De Clercq, O., I. Hendrickx, and V. Hoste (2011) Cross-domain Dutch coreference resolution. In Proceedings of RANLP-2011, pages 186–193.
- * De Clercq, O. and V. Hoste (2016) Rude waiter but mouthwatering pastries! An exploratory study into Dutch aspect-based sentiment analysis. Proceedings of LREC 2016, pages 2910-2917.
- * Desmet, B., Hoste, V., Verstraeten, D. & Verhasselt, J. (2013). Gallop Documentation. Technical Report LT3 13-03, Ghent University.
- * Feldman, R. (2013) Techniques and applications for sentiment analysis. Communications of the ACM, 56(4):82–89.
- * Hoste, V. (2005). Optimization Issues in Machine Learning of Coreference Resolution. University of Antwerp.
- * Liu, B. (2012) Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1):1–167.
- * Paulheim, H. and Fürnkranz, J. (2012) Unsupervised Generation of Data Mining Features from Linked Open Data. In Proceedings of WIMS-2012, page 31.

References

- * 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., Jimenez-Zafra, S.M. and Eryvigit, G. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of SemEval-2016, pages 19–30. *
- * Lehmann, J., Robert I., Max J., Jentzsch, A., Kontokostas, D. and Mendes, P., Hellmann, S., Morse, M., van Kleef, P., Auer, S. and Bizer, C. (2013) DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web Journal, 6:167–195.
- * Mendes, P., Jakob, M., Garcia-Silva, A. and Bizer, C. (2011) DBpedia Spotlight: Shedding light on the web of documents. In Proceedings I-Semantics-2011, pages 1–8.
- * Vossen, P., I. Maks, R. Segers, H. van der Vliet, M.F. Moens, K. Hofmann, E. Tjong Kim Sang, and M. de Rijke (2013) Cornetto: a lexical semantic database for Dutch. In Peter Spyns and Jan Odijk, editors, Essential Speech and Language Technology for Dutch, Theory and Applications of Natural Language Processing. Springer, pages 165–184.