Experimente im Natural Language Processing

HS Experimentelles Arbeiten in der Sprachverarbeitung

Nils Reiter
nils.reiter@uni-koeln.de

27. Oktober 2022

Section 1

Two Papers

Today

- Preoțiuc-Pietro et al. (2019): "Automatically Identifying Complaints in Social Media"
- Panchendrarajan et al. (2016): "Implicit Aspect Detection in Restaurant Reviews using Cooccurence of Words"

- Which one did you like better?
- Which one was easier to understand?

Today

- Preoţiuc-Pietro et al. (2019): “Automatically Identifying Complaints in Social Media”
- Panchendrarajan et al. (2016): “Implicit Aspect Detection in Restaurant Reviews using Cooccurence of Words”

- Which one did you like better?
- Which one was easier to understand?

- Very typical NLP papers
  - 8-9 pages, densely written
  - Structure: Abstract – Introduction – Related work – Data description/analysis – Experimental part – Conclusions – References

Today

- Preoţiuc-Pietro et al. (2019): "Automatically Identifying Complaints in Social Media"
- Panchendrarajan et al. (2016): "Implicit Aspect Detection in Restaurant Reviews using Cooccurence of Words"

- Which one did you like better?
- Which one was easier to understand?

- Very typical NLP papers
  - 8-9 pages, densely written
  - Structure: Abstract – Introduction – Related work – Data description/analysis – Experimental part – Conclusions – References
- My opinion: Preoţiuc-Pietro et al. (2019) 'better' than Panchendrarajan et al. (2016)

Reading up on Details, Techniques, Methods

Abbreviation	Reference
MS99	Christopher D. Manning/Hinrich Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts and London, England: MIT Press
JM19	Dan Jurafsky/James H. Martin (2019). *Speech and Language Processing*. 3rd ed. Draft of October 16, 2019. Prentice Hall

Table: References to text books

Comprehension Questions

- 10-fold cross validation
- ROC AUC
- Maximum entropy classification
- Cohen's Kappa

The Tasks

Preoţiuc-Pietro et al. (2019)

- Target concept: Complaints
- Binary classification of tweets
- A tweet is positive, if it contains at least one complaint speech act
- No context dependency

The Tasks

Preoțiuc-Pietro et al. (2019)

- Target concept: Complaints
- Binary classification of tweets
- A tweet is positive, if it contains at least one complaint speech act
- No context dependency

Panchendrarajan et al. (2016)

- Target concept: Mentioned and reviewed aspects
- Multi-label classification of sentences
    - Not explicitly stated by the authors
- No context dependency

The Tasks

Preoţiuc-Pietro et al. (2019)

- Target concept: Complaints
- Binary classification of tweets
- A tweet is positive, if it contains at least one complaint speech act
- No context dependency

Panchendrarajan et al. (2016)

- Target concept: Mentioned and reviewed aspects
- Multi-label classification of sentences
  - Not explicitly stated by the authors
- No context dependency

Reminder: Classification

- Organize items into previously defined classes MS99, 192,575
- Multi-class: More than two classes (i.e., more than binary)
- Multi-label: Each item can get more than one label

The Data Sets

Preoțiuc-Pietro et al. (2019)

- 3449 English tweets, no retweets
  - 1971 to which support accounts replied
  - 739 @-replies
  - 739 other tweets

The Data Sets

Preoțiuc-Pietro et al. (2019)

- 3449 English tweets, no retweets
    - 1971 to which support accounts replied
    - 739 @-replies
    - 739 other tweets
- Preprocessing
    - Replace all usernames
    - Replace all URLs
    - Extract unigrams

The Data Sets

Preoțiuc-Pietro et al. (2019)

- 3449 English tweets, no retweets
  - 1971 to which support accounts replied
  - 739 @-replies
  - 739 other tweets
- Preprocessing
  - Replace all usernames
  - Replace all URLs
  - Extract unigrams
- Annotation
  - Two independent annotators
  - Agreement $\kappa = 0.731$ (Cohen, 1960)

The Data Sets

Preoţiuc-Pietro et al. (2019)

- 3449 English tweets, no retweets
  - 1971 to which support accounts replied
  - 739 @-replies
  - 739 other tweets
- Preprocessing
  - Replace all usernames
  - Replace all URLs
  - Extract unigrams
- Annotation
  - Two independent annotators
  - Agreement $\kappa = 0.731$ (Cohen, 1960)

Panchendrarajan et al. (2016)

- 1000 restaurant reviews from Yelp

The Data Sets

Preoţiuc-Pietro et al. (2019)

- 3449 English tweets, no retweets
  - 1971 to which support accounts replied
  - 739 @-replies
  - 739 other tweets
- Preprocessing
  - Replace all usernames
  - Replace all URLs
  - Extract unigrams
- Annotation
  - Two independent annotators
  - Agreement $\kappa = 0.731$ (Cohen, 1960)

Panchendrarajan et al. (2016)

- 1000 restaurant reviews from Yelp
- Annotation (p. 135)
  - Two independent annotators on 3 samples of 100 reviews
  - Sentence-wise annotation
  - Agreement $\kappa = 0.834$ (Cohen, 1960)

The Data Sets

Preoțiuc-Pietro et al. (2019)

- 3449 English tweets, no retweets
  - 1971 to which support accounts replied
  - 739 @-replies
  - 739 other tweets
- Preprocessing
  - Replace all usernames
  - Replace all URLs
  - Extract unigrams
- Annotation
  - Two independent annotators
  - Agreement $\kappa = 0.731$ (Cohen, 1960)

Panchendrarajan et al. (2016)

- 1000 restaurant reviews from Yelp
- Annotation (p. 135)
  - Two independent annotators on 3 samples of 100 reviews
  - Sentence-wise annotation
  - Agreement $\kappa = 0.834$ (Cohen, 1960)
- Highly skewed distribution (Most sentences do not contain implicit aspects)

Experimental Setup

Preoţiuc-Pietro et al. (2019)

- 10-fold cross validation JM19, 69
- Parameters: 3-fold CV in inner loop

Experimental Setup

Preoţiuc-Pietro et al. (2019)

- 10-fold cross validation JM19, 69
- Parameters: 3-fold CV in inner loop
- Evaluation
  - Mean accuracy
  - F1 (macro-average)
  - ROC AUC MS99, 270
    - (ROC = receiver operating characteristic curve / AUC = area under curve)

Experimental Setup

Preoţiuc-Pietro et al. (2019)

- 10-fold cross validation JM19, 69
- Parameters: 3-fold CV in inner loop
- Evaluation
  - Mean accuracy
  - F1 (macro-average)
  - ROC AUC MS99, 270
    - (ROC = receiver operating characteristic curve / AUC = area under curve)

Panchendrarajan et al. (2016)

- 10-fold cross validation JM19, 69
- Additional 400 reviews used for testing M1

Experimental Setup

Preoţiuc-Pietro et al. (2019)

- 10-fold cross validation JM19, 69
- Parameters: 3-fold CV in inner loop
- Evaluation
  - Mean accuracy
  - F1 (macro-average)
  - ROC AUC MS99, 270
    - (ROC = receiver operating characteristic curve / AUC = area under curve)

Panchendrarajan et al. (2016)

- 10-fold cross validation JM19, 69
- Additional 400 reviews used for testing M1
- Evaluation
  - Precision/recall/F1 MS99, 267 ff.

Preprocessing

Processing steps before actual task solving

Preoțiuc-Pietro et al. (2019)

- Part of speech
- Sentiment
- Request detection
- Politeness
- Time expressions

Preprocessing

Processing steps before actual task solving

Preoțiuc-Pietro et al. (2019)

- Part of speech
- Sentiment
- Request detection
- Politeness
- Time expressions
- Word2vec

Preprocessing

Processing steps before actual task solving

Preoţiuc-Pietro et al. (2019)

- Part of speech
- Sentiment
- Request detection
- Politeness
- Time expressions
- Word2vec
- Rule-based ad-hoc systems
  - Intensifiers
  - Pronoun types
  - LIWC

Preprocessing

Processing steps before actual task solving

Preoțiuc-Pietro et al. (2019)

- Part of speech
- Sentiment
- Request detection
- Politeness
- Time expressions
- Word2vec
- Rule-based ad-hoc systems
  - Intensifiers
  - Pronoun types
  - LIWC

Panchendrarajan et al. (2016)

- Dependency relations
  - Which one?

Preprocessing

Processing steps before actual task solving

Preoțiuc-Pietro et al. (2019)

- Part of speech
- Sentiment
- Request detection
- Politeness
- Time expressions
- Word2vec
- Rule-based ad-hoc systems
  - Intensifiers
  - Pronoun types
  - LIWC

Panchendrarajan et al. (2016)

- Dependency relations
  - Which one?

Pre-Processing

- No global definition of what counts as pre-processing
- Context-dependent

Methods

Preoțiuc-Pietro et al. (2019)

- Baseline: Most frequent class

Methods

Preoțiuc-Pietro et al. (2019)

- Baseline: Most frequent class
- Logistic regression with manually specified features JM19, 75 ff.

Methods

Preoţiuc-Pietro et al. (2019)

- Baseline: Most frequent class
- Logistic regression with manually specified features JM19, 75 ff.
- Neural networks with one-hot-encoded word vectors as input
  - MLP: Feedforward neural network JM19, 129 ff.
  - LSTM: Sequential classifier (word by word) JM19, 184 ff.

Methods

Preoţiuc-Pietro et al. (2019)

- Baseline: Most frequent class
- Logistic regression with manually specified features JM19, 75 ff.
- Neural networks with one-hot-encoded word vectors as input
  - MLP: Feedforward neural network JM19, 129 ff.
  - LSTM: Sequential classifier (word by word) JM19, 184 ff.

Panchendrarajan et al. (2016)

- M1 (for explicit aspects): Maximum entropy classifier with n-grams as features ($2 \leq n \leq 5$) = Logistic regression JM19, 75 ff.

Methods

Preoțiuc-Pietro et al. (2019)

- Baseline: Most frequent class
- Logistic regression with manually specified features JM19, 75 ff.
- Neural networks with one-hot-encoded word vectors as input
  - MLP: Feedforward neural network JM19, 129 ff.
  - LSTM: Sequential classifier (word by word) JM19, 184 ff.

Panchendrarajan et al. (2016)

- M1 (for explicit aspects): Maximum entropy classifier with n-grams as features ($2 \leq n \leq 5$) $=$ Logistic regression JM19, 75 ff.
- M2 (for implicit aspects)
  - Training: Collect dictionary (called 'model' by the authors)
  - Testing
    1. Generate candidates, based on score A_i (Eq. 1)
    2. Remove candidates according to rules (Fig. 1)
  - Modification 1 and 2 (p. 133)

Summary

- Typical NLP papers: Focus in methods
- Complaints
  - Very clear
  - Classical machine learning wins
- Reviews
  - Implicit aspects in restaurant reviews
  - Machine learning and rules on top