PA164 Natural Language Learning

Lecture 09: Sentiment Analysis

Vít Nováček

Faculty of Informatics, Masaryk University

Autumn, 2023

MUNI

Outline

1. Introduction
2. Lexicon-Based Approaches
3. Classical ML Approaches
4. Deep Learning Approaches
5. Hybrid Approaches
6. Comparing the Approaches
7. Useful References

What Is Sentiment Analysis (SA)?

- A field of research falling under text mining
- Focused on computational treatment of:
  - opinions,
  - sentiments
- ...or, in general, subjectivity, in text
- Most works focused simply on sentiment and its polarity
  - *The movie's absolutely mindblowing!* $\sim$ +0.99
  - *Eh, the food there? Not great, not terrible...* $\sim$ 0
  - *Awful, simply awful.* $\sim$ -0.95
  - Relatively easy to formalise and study
- Levels of sentiment
  - Document (overall sentiment of a text)
  - Sentence (intermediate level)
  - Aspect (sentiment associated with different aspects of an entity, such as the battery life or screen quality of a phone)

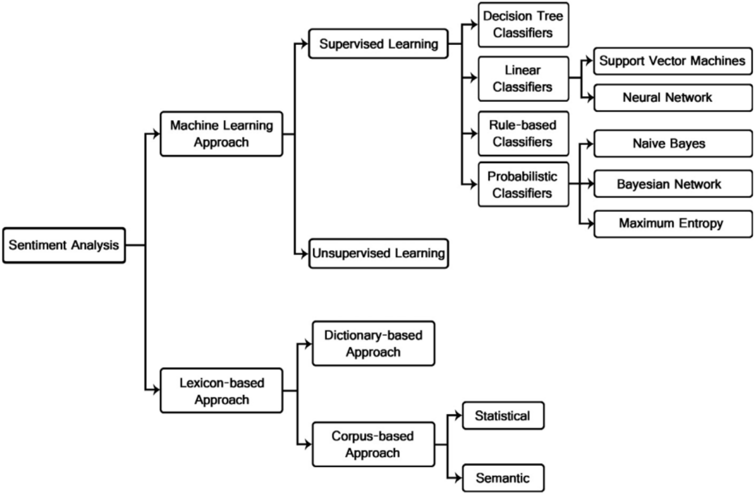
Why Bother?

- Challenging research problem involving many NLP aspects
- High practical impact as well, though
- Main sources of data and investigated use cases:
  - reviews (product development and marketing)
  - stock markets (market analysis and forecasting, business review analysis)
  - news (public opinion mining and forecasting, business review analysis)
  - political debates (public opinion mining and forecasting)
  - social media (all of the above, crime detection and prevention, disaster response, disease outbreaks, . . .)

Main Challenges

- Complexity of the problem
  - To get the sentiment right...
  - ...one would would need to get the meaning right first
  - That means dealing not only with lexical issues...
  - ...but also syntax and those pesky semantic features like anaphora, metaphor or irony
    - ★ C.f., *Wonderful, I can't get enough of this president in a fishtank!*
- Lack of sufficiently comprehensive supporting resources
  - Building annotated resources for supervised learning is expensive and often outright intractable
- Limited language cover
  - Partly related to the previous challenge
  - But it's also about lack of tools, community or simply interest

Overview of Approaches



[1] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams engineering journal 5.4 (2014): 1093-1113.

Outline

1. Introduction
2. Lexicon-Based Approaches
3. Classical ML Approaches
4. Deep Learning Approaches
5. Hybrid Approaches
6. Comparing the Approaches
7. Useful References

Rationale of the Lexicon-Based Approaches

- Sentiment lexicon—opinion words, phrases and idioms mapped to their polarity
  - Words: *bad*: -0.75, *good*: +0.75, *lovely*: +0.85
  - Phrases: *handsome boy*: +0.8, *crooked witch*: -0.9
  - Idioms: *That fella's driving me mad*: -0.8, *I'm over the moon*: +0.8
- Sentiment of a chunk of text can then be determined by:
  - Looking up the text constituents in the sentiment lexicon
  - Optional compositional analysis of the sentiment of higher-level structures (e.g., sentences)
  - Aggregation of the sentiment
- Two main types:
  1. Dictionary-based
  2. Corpus-based

Dictionary-Based Lexicon Approaches

- Manual definition of seed lexicon terms
- Bootstrapping of the full-fledged lexicon
  - Automatic extension of the seed by synonyms (via WordNet or thesaurus)
- Can't determine context- or domain-dependent sentiments, though

Corpus-Based Lexicon Approaches

- Solves the main limitation of the dictionary-based approaches
- Uses collocations and syntactic patterns in a corpus to extend the seed lexicon
- Two overall sub-approaches:
  1. Statistical
     - ★ Various types of distributional analysis (phrases sharing similar sentiment are assumed to co-occur in corpora)
     - ★ Approaches based on semantic spaces (topological analysis of meaning in embedding spaces of sorts)
  2. Semantic
     - ★ Based on the assumption that semantically similar words have similar sentiments
     - ★ Often deploys WordNet or a similar resource to bootstrap the lexicon from the annotated seeds
     - ★ Can utilise other semantic relationships than synonymy and similarity, too (e.g., hypero-hyponymy or antonymy)

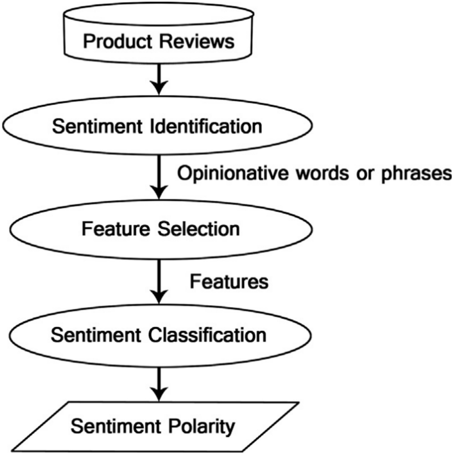
Syntactic and Semantic Analysis in Lexicon Approaches

- Syntactic extensions:
  - Deep (such as dependency) parsing of sentences
  - Propagation and/or composition of sentiment along the syntactic relations
- Semantic extensions:
  - Utilising discourse information, often rhetorical relations
    - Examples: Contrast, Support, Correction, Result, Continuation
  - Rhetorical structure theory
    - Identifying rhetorically meaningful sub-units of text
    - Corresponding to argumentative structure of the implied communication
    - Support vs. opposition can be mapped to positive vs. negative sentiments

Outline

1. Introduction
2. Lexicon-Based Approaches
3. Classical ML Approaches
4. Deep Learning Approaches
5. Hybrid Approaches
6. Comparing the Approaches
7. Useful References

Rationale of the Machine Learning Approaches



[1] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams engineering journal 5.4 (2014): 1093-1113.

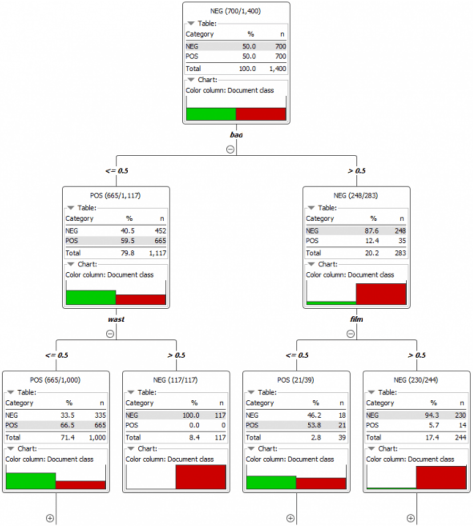
Probabilistic Classifiers

- Naive Bayes
  - Simple to design
  - Easy to train
  - Imbalanced accuracy, though (positive sentiments may be classified more accurately than negative ones)
- Bayesian network
  - More sophisticated model, catering for complex feature dependencies
  - Difficult to train, though
  - Then again, can work well in semi-supervised settings
- Maximum entropy
  - Encodes labeled features sets as vectors
  - Learns weights that can be used for aggregating the encoded vectors to determine label likelihoods
  - Used for cross-language model development

Linear Classifiers

- SVMs
  - Notably, approaches based on SVMs can take into account meta-aspects of sentiment
    - Review quality, subjectivity or author credibility
  - Can outperform sophisticated neural models in some settings
- Neural networks
  - Initially not terribly successful
  - Prohibitive training times, sensitivity to unbalanced data
  - Under the right settings, even the first models did outperform the classical ML models, though

Decision Trees



[2] Thiel, Kilian, Rudnitckaia, Lada. "Sentiment Analysis Tutorial." A KNIME.com blog post (https://www.knime.com/blog/sentiment-analysis) (2021).

Rule-Based Classifiers

- Related to decision trees
- Inferring rules from the data
  - LHS is a propositional DNF of features (present in a text chunk), RHS is a label (sentiment)
  - Optimising support (how many matching instances in data) and confidence (conditional probability of the RHS, given the LHS)
- Quite like associative rule mining

Weakly Supervised Approaches

- Deal with the problem of the lack of labels
- Possible weakly supervised solutions in the field of SA:
  - Example-level
    - Label sentences based on sentiment key-words present in them
    - Use sentence similarity measure(s) to propagate the labels to unlabeled examples
  - Feature-level
    - An initial supervised classifier using sentiment lexicon
    - Using that classifier to constrain predictions on unlabeled data
- Unsupervised solutions exist, too
  - Typically based on some distributional similarity measure between words and polarity prototypes

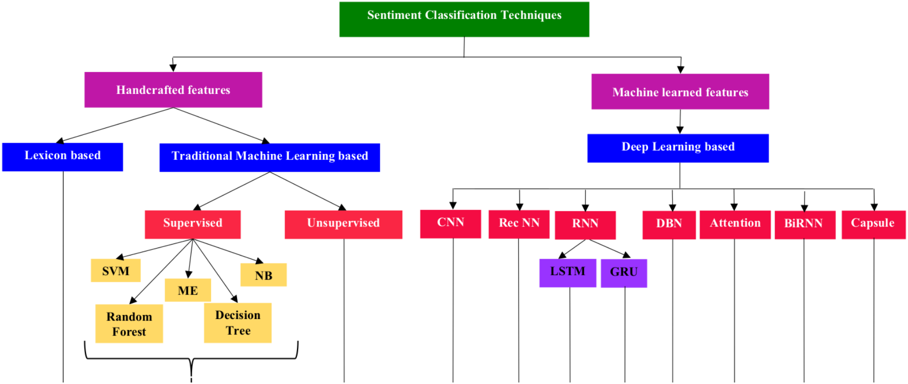
Meta-Classifiers

- Working with one model only may be challenging
  - Unbalanced data (often intrinsically—for instance, news outlets prefer negative stories)
  - Dynamic nature of data (e.g., social network feeds)
  - Unclear relationships between formal training and evaluation procedures and "real world" relevance of the models
- Ensemble approaches working with multiple data sets and/or models can at least partly overcome these challenges

Outline

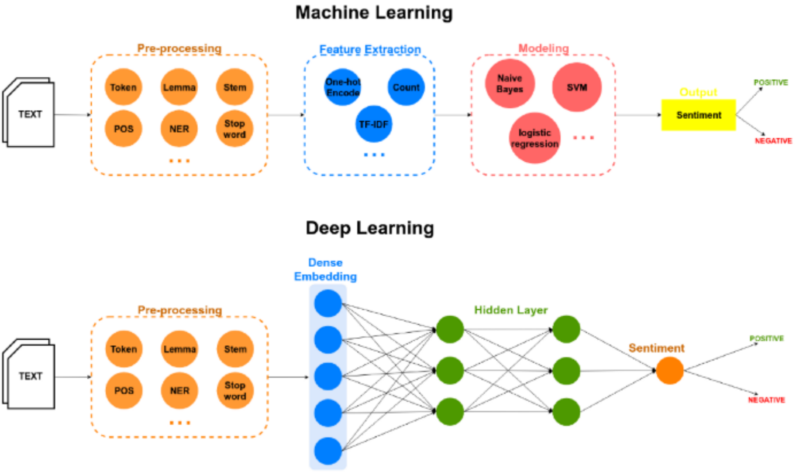
1. Introduction
2. Lexicon-Based Approaches
3. Classical ML Approaches
4. Deep Learning Approaches
5. Hybrid Approaches
6. Comparing the Approaches
7. Useful References

Taxonomy of Machine Learning Approaches Revisited



[3] Yadav, Ashima, and Dinesh Kumar Vishwakarma. "Sentiment analysis using deep learning architectures: a review." Artificial Intelligence Review 53.6 (2020): 4335-4385.

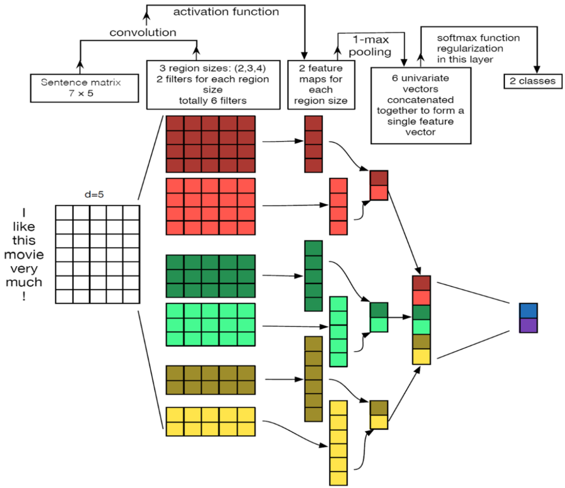
Rationale of the Deep Learning Approaches



[4] Jain, Kamal. "Sentiment Analysis using Deep Learning." A Medium blogpost

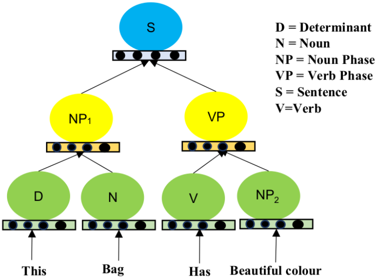
(https://medium.com/analytics-vidhya/sentiment-analysis-using-deep-learning-a416b230ca9a) (2020).

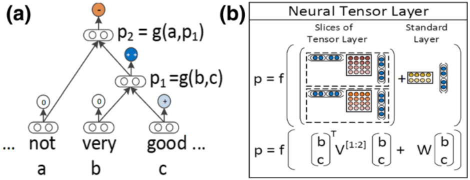
CNNs in SA – Illustrative Example



[5] Zhang, Ye, and Byron Wallace. "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification." arXiv preprint arXiv:1510.03820 (2015).

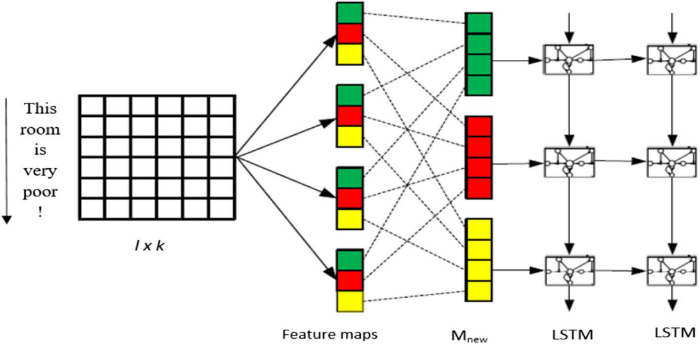
RecNN in SA – Illustrative Example





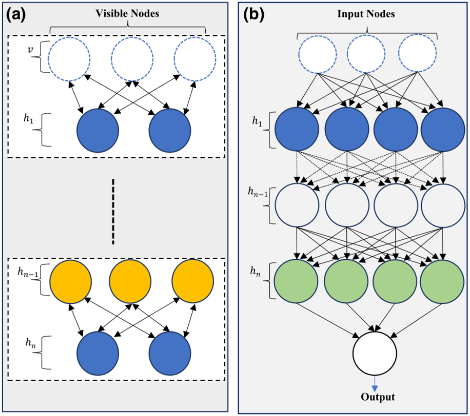
[3] Yadav, Ashima, and Dinesh Kumar Vishwakarma. "Sentiment analysis using deep learning architectures: a review." Artificial Intelligence Review 53.6 (2020): 4335-4385.

LSTM+CNN in SA – Illustrative Example



[6] Huang, Qiongxia, et al. "Deep sentiment representation based on CNN and LSTM." 2017 International Conference on Green Informatics (ICGI). IEEE, 2017.

DBN in SA – Illustrative Example



[3] Yadav, Ashima, and Dinesh Kumar Vishwakarma. "Sentiment analysis using deep learning architectures: a review." Artificial Intelligence Review 53.6 (2020): 4335-4385.

Other Recent Approaches

- Attention-based networks
  - Using the attention mechanism to filter out less relevant parts of text
- Bi-directional RNNs
  - Reflecting information ahead (and not only behind) in the sequence
  - Having thus full context window
- Capsule networks
  - Remedy to drawbacks of CNNs (namely limited representation of hierarchies)
  - Dynamic routing between capsule networks (vectors of neurons) instead of max-pooling
  - Can be trained with much less information than most other architectures
- And, obviously, transformers
  - Typically via fine-tuning pre-trained models

Commonly Used Sentiment Data Sets (1/2)

Stanford large movie review (IMDB)	Yelp dataset	Stanford sentiment treebank (SSTb)	Amazon review dataset
50,000 binary labeled movie reviews, balanced	restaurant reviews, labeled on 1-5 scale	11,855 movie reviews (Rotten Toma-toes), five sentiment classes	4 types of product reviews on Amazon, binary labels

Commonly Used Sentiment Data Sets (2/2)

CMU-MOSI dataset	MOUD dataset	Getty images dataset	Twitter dataset	Twitter image dataset
multimodal dataset, 2,199 opinionated utterances	another multimodal dataset, in Spanish	588,221 labeled data points, image and text	220,000 tweets with both text and images	1,269 image tweets

Comparison of Selected Approaches (Performance)

Refs.	Dataset	Technique/Method	Accuracy (%)	# Instances in training set
Baktha and Tripathy (2017)	Amazon health product reviews	Vanilla RNN LSTM **GRU** Bi-Vanilla RNN Bi-LSTM Bi-GRU	57.30 78.10 **83.90** 58.00 79.20 81.10	Amazon: 8000
Xu et al. (2011)	Amazon mobile phone reviews	**Multi-class SVM** CRF without interdependencies CRF with interdependencies	**61.38** 60.04 66.17	–
Rain (2013)	Amazon books, media, and kindle product reviews	Decision Tree **Naïve Bayes**	79.84 (for books) **84 (for kindle)**	–
Shaikh and Deshpande (2016)	Amazon books, camera, music product reviews	Naïve Bayes (Books) Naïve Bayes (Camera) Naïve Bayes (Music)	80 (multiword level feature) 80 (single word level feature) 80 (single word level feature)	260
Al-Smadi et al. (2017)	Arabic Hotels' Reviews (Al-Smadi et al. 2016; Pontiki 2016)	For aspect-based sentiment analysis tasks Deep RNN **SVM**	For Sentiment Polarity Identification RNN:87 **SVM:95.4**	19,226

[3] Yadav, Ashima, and Dinesh Kumar Vishwakarma. "Sentiment analysis using deep learning architectures: a review." Artificial Intelligence Review 53.6 (2020): 4335-4385.

Comparison of Selected Approaches (Execution Times)

Refs.	Sentiment analysis task	Dataset (#Training sample instances)	Deep learning architecture	Execution or training time (s)	Platform	Highest accuracy (%)
Zhao et al. (2017)	Sentence classification	Review from Amazon on digital cameras, cell phones and laptops (600)	CNN **LSTM**	1800 18,000 (Execution time)	Nvidia GTX 980Ti GPU	87.9
Li et al. (2019)	Sentence classification	Yelp 2015 (808,052)	GRU **Bi-GRU** Sliced RNN Bi- Sliced RNN	1609 3176 218 440 (Training time)	Keras, NVIDIA GTX 1080Ti GPU	73.36
Li et al. (2017b)	Sentence classification	Online debates (24352), Restaurants (2614) and laptop reviews (5485) from SemEval 2014 and 2015	CNN LSTM MemNet **AttNet**	CNN:5 LSTM:200 MemNet:150 AttNet:200 (Training time)	TITAN X GPU	(Avg F-score) Debates: 52.23 Tweets: 35.34 Review: 55.93
Tay et al. (2017)	Aspect-based sentiment analysis	Customer reviews for laptop (1813), restaurants (3102), SemEval 2014 (3587), Tweets from SemEval 2016 (2771), Online Debates (24564)	**Memory NN** LSTM Attention- LSTM	Memory NN:6 LSTM:9 Attention- LSTM: 12 (Execution time)	NVIDIA GTX 1070 GPU	(Overall F1 score) 69.2
Yuan et al. (2018)	Multi-domain sentiment classification	Amazon multi-domain dataset (Amazon) (Blitzer et al. 2007) (1400) Sanders Twitter Sentiment Dataset (Sanders)	RNN GRU LSTM **BiLSTM + attention**	RNN:71 GRU:246 LSTM:310 LSTM with peephole connection:411 (Training time)	TensorFlow, NVIDIA K80 GPU	(Avg accuracy) Amazon: 87.69 Sanders: 86.32

[3] Yadav, Ashima, and Dinesh Kumar Vishwakarma. "Sentiment analysis using deep learning architectures: a review." Artificial Intelligence Review 53.6 (2020): 4335-4385.

Outline

1. Introduction
2. Lexicon-Based Approaches
3. Classical ML Approaches
4. Deep Learning Approaches
5. Hybrid Approaches
6. Comparing the Approaches
7. Useful References

Rationale of the Hybrid Approaches

- To get the best of all worlds, simply
- Typically, injecting precise and reliable, but small manually annotated data, plus some linguistic insights. . .
- . . . into robust data-crunching models
- Examples:
  - Rules and lexicon for rich features, then machine learning
  - RecNN (combining syntax and deep learning)
  - Transformer + capsule ANN
  - CNN + biLSTM
  - . . .

Outline

1. Introduction
2. Lexicon-Based Approaches
3. Classical ML Approaches
4. Deep Learning Approaches
5. Hybrid Approaches
6. Comparing the Approaches
7. Useful References

Pros and Cons: Lexicon-Based Approaches

- Pros:
  - Highly accurate
  - Can reflect complex linguistic domain knowledge
- Cons:
  - Expensive to design and maintain
  - Low coverage even under the most ideal circumstances

Pros and Cons: Machine Learning Approaches

- Pros:
  - Substantially higher coverage than lexicon-based approaches
  - Not necessarily sacrificing much accuracy
- Cons:
  - May involve rather arcane and expensive feature engineering
  - Typically based on bag-of-words semantics - may miss a lot due to disregarding compositionality

Pros and Cons: Deep Learning Approaches

- Pros:
  - Learn features on their own
  - Can reflect some implicit syntactic structure of the texts
  - Can learn to aggregate sentiment in higher-level language units
- Cons:
  - Tuning and training largely empirical process
  - May take long to train on large datasets
  - Simpler models (e.g., SVM) can still perform better

Outline

1. Introduction
2. Lexicon-Based Approaches
3. Classical ML Approaches
4. Deep Learning Approaches
5. Hybrid Approaches
6. Comparing the Approaches
7. Useful References

Further Readings

- Classical approaches (including “shallow” ML):
  - Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. ”Sentiment analysis algorithms and applications: A survey.” Ain Shams engineering journal 5.4 (2014): 1093-1113.
- Deep learning approaches:
  - Yadav, Ashima, and Dinesh Kumar Vishwakarma. ”Sentiment analysis using deep learning architectures: a review.” Artificial Intelligence Review 53.6 (2020): 4335-4385.
- Recent surveys:
  - Liu, Bing. Sentiment analysis and opinion mining. Springer Nature, 2022.
  - Wankhade, Mayur, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. ”A survey on sentiment analysis methods, applications, and challenges.” Artificial Intelligence Review 55.7 (2022): 5731-5780.
  - Chan, Jireh Yi-Le, et al. ”State of the art: a review of sentiment analysis based on sequential transfer learning.” Artificial Intelligence Review 56.1 (2023): 749-780.