Statistical Consulting

Topic-Specific Sentiment Analysis for Tweets by German MPs

Department of Statistics
Ludwig-Maximilians-Universität München

Munich, month day$^{\text{th}}$, 2021

Authors Asmik Nalmpatian
Lisa Wimmer

Project Partner Prof. Dr. Paul Thurner
Department of Political Science

Supervisors Matthias Aßenmacher, Ph.D.
Prof. Dr. Christian Heumann
Department of Statistics

Abstract

Contents

List of Abbreviations

ABSA	aspect-based sentiment analysis
BERT	Bidirectional Encoder Representations from Transformers
BOW	bag of words
DL	deep learning
LDA	latent Dirichlet allocation
ML	machine learning
MP	Member of Parliament
NLP	natural language processing
STM	structural topic model
TSSA	topic-specific sentiment analysis

List of Figures

List of Tables

1 Introduction

The advance of social media has sparked a fundamental change to public debate. Political coverage and discourse has spilled from traditional outlets over to online locations whose accessibility has lowered entrance barriers to welcome a wide audience (Bode, 2017). Social media exhibit certain properties that make them attractive to politicians seeking to broadcast their message. From a supply-side perspective it is easy to publish content: posting on social media is cheap, does not require approval of any authority, and allows for full control over the intended presentation. In an environment that spins information at enormous speed the resulting ability to react to events in real-time offers a distinct advantage over the inertia of traditional channels (Stier et al., 2018). On the receiving end, social media grant unprecedented access to target audiences. The industry's oligopolistic structure sees people from heterogeneous backgrounds convene on few global platforms to exchange their views. Research suggests that the low cost of online engagement causes political content to also circulate among users with less political affinity (Jost et al., 2018). A contrary but equally important aspect is the evolution of echo chambers. Users view content according to their perceived preferences and thus often end up in community niches populated by like-minded people. This clustering process creates groups disproportionately receptive to certain messages and has played an important role in large-scale propaganda. Echo chambers are particularly suited for the direct communication among users social media offer: they enable a dialogue between politicians and their electorate that is hard to achieve via traditional channels, and allow to deploy the power of emotion to shape opinion (Hasell and Weeks, 2016).
These opportunities have propelled internet platforms to a position at least level with the former hegemons of political debate. Twitter, in particular, has emerged as a medium for political information. In what might have been inconceivable a few years ago politicians actively convey messages to the public via tweets (van Vliet et al., 2020). The impact of this change in the political environment is complex and most certainly has positive as well as worrisome aspects. Yet, from a purely scientific point of view, activity on social media creates on its way a vast amount of publicly accessible data that benefits the research community with a constant source of information.

A question frequently posed in political analysis is the assessment of public opinion toward a particular matter. However, the textual data gathered from social media that might hold the answer command the use of specific tools subsumed under the field of *natural language processing (NLP)*. NLP has gained much traction with the rise of deep learning methods and become virtually ubiquitous, techniques ranging from simple heuristics to gigantic neural networks powering search engines and the like (Torfi et al., 2020). Statistically speaking, the above problem translates into the classification of texts into instances of certain sentiments (typically, *positive* and *negative*). Building upon the assumption that sentiment may be expressed differently in varying contexts, such *sentiment analysis* is often combined with some form of *topic modeling* (see, for example, Ficamos and Liu (2016)).
It is the goal of this project to make analysis of social media texts in a political context more easily accessible to researchers. We focus on the analysis of public sentiment in a topic-aware manner for texts collected from Twitter posts by German Members of Parliament (MPs). Our contribution is two-fold:

1. We explore how topic-specific sentiment analysis can be implemented, considering (1) standard machine learning (ML) techniques and (2) more complex deep learning (DL) models.

2. We provide extensive teaching material on both approaches, composed as a coherent online course, to educate researchers on addressing NLP problems in their own work.

The remainder of this report is organized as follows. First, we outline the project in more detail in section 2. Section 3 provides some theoretical context for topic modeling and sentiment analysis from which we derive what we call *topic-specific sentiment analysis (TSSA)*. We proceed in section 4 by sketching our data collection and cleaning process and then present our proposal for conducting TSSA, laying out for both approaches the underlying methodology and discussing the results of applying them to the data at hand. Section 5 outlines how the findings from this analysis translate to the proposed teaching material. Afterwards, in section 6, we critically assess the findings and limitations of the project, and conclude with a brief summary in section 7.

2 Project Outline

Before we outline the scope of our work it should be noted that parts of it are based on a predecessor project. Schulze and Wiegrebe (2020) studied how German MPs' Twitter data can be modeled with a *structural topic model* (*STM*, Roberts et al. (2013)), and have since engaged in follow-up research on the STM (Schulze et al., 2021). Much of the data procurement and topic modeling process is adopted from their work. The project at hand encompasses two subsequent steps. In a pioneering mode we first explore the overall feasibility of TSSA with both basic and more advanced statistical techniques from the NLP toolbox, and then, based on our findings, propose a collection of material to support fellow researchers in conducting similar studies.

Topic-specific sentiment analysis. Regardless of how the downstream task is solved, the first challenge to address is data collection. The idea here is to retrieve information from the web in an automated and resource-efficient manner that results in a suitable data structure. Afterwards, we pursue two fundamentally different approaches toward performing TSSA.

1. The first approach applies standard ML tools which require the input data to be of tabular form. Obviously, texts are complex constructs and not arbitrary sequences of interchangeable characters that can simply be cast into tabled variables (section 4.1.3 will address the challenges arising from Twitter data in more detail). It is nevertheless possible, as general research and also or own results suggest, to obtain fairly good performance with this reduction of complexity.

2. State-of-the-art approaches, by contrast, avoid such blunt simplification and attempt to teach the entire concept of language to machines. This comes at the expense of large computational requirements but achieves promising results in a variety of NLP tasks. We therefore build a deep bidirectional Transformer architecture (BERT, Devlin et al. (2019)) as a second approach and examine whether the additional complexity is justified by better performance.

The basic approach is implemented in `R` (R Core Team, 2021) and thus easily integrated with statistical education at LMU. For the BERT solution we resort to `Python` (van Rossum and Drake, 2011) which is all but standard for deep (NLP) modeling.

Knowledge transfer. Based on the results of this exploratory analysis we propose teaching material devised to support research in similar applications. The acquired collection is organized as a coherent and self-contained tutorial composed of basic theory, code demonstrations, and exercises (including solutions). While the course materials are primarily aligned to solve the TSSA task, the covered components are certainly also instructive for other types of applications. We have made the material available for both live teaching purposes and self-study on a public website. First experiences from a live workshop held in April/May 2021 for researchers from the Department of Political Science at LMU will be discussed in section 6.

3 General Theoretical Context

3.1 Terminology

In this section we briefly review general theoretical concepts; the actual methods we employ are described in chapter 4. Throughout the report we will make use of the following terminology:

Word. Words w are sequences of characters and represent the smallest unit of text we consider.
Vocabulary. The aggregate of unique terms present in a collection of text constitutes a vocabulary of length $V \in \mathbb{N}$ from which a one-hot encoding for words can be derived: for the v-th instance of the vocabulary, $v \in \{1, 2, \ldots, V\}$, this is a length-$V$ vector with all but the v-th entry, which is one, equaling zero. Note that pre-processing (discussed in chapter 4) might result in a vocabulary that is smaller than the total number of distinct words occurring across all texts.
Document. Documents $d \in \{1, 2, \ldots, D\}$, $D \in \mathbb{N}$, are generally understood to be sequences of $N_d \in \mathbb{N}$ words, and, in our case, tweets.
Corpus. Lastly, the set of all D documents considered makes up a corpus.

3.2 Theoretical Concepts

3.2.1 Topic Modeling

Idea. Recall that the ultimate goal is the classification of tweets into groups signaling a specific sentiment. It is reasonable to assume that sentiment, and the way of expressing it, is susceptible to context, which suggests potential gains from clustering tweets prior to sentiment analysis (see, for example, Ficamos and Liu (2016), Bhatia and Padmanabhan (2018), or Jang et al. (2021)).
Grouping texts into semantic clusters, or topics, is generally referred to as *topic modeling* and typically an unsupervised learning task. The idea is to uncover latent structures in a corpus along which documents can be characterized. Topic modeling is essentially a means of dimensionality reduction: text analysis requires text to be cast to numerical representation, the simplest form of which is to represent documents by counts of vocabulary instances. The dimension of the resulting document-term matrix increases exponentially in the number of documents and words contained in them, making an urgent case for compressing this dimensionality (Vayansky and Kumar, 2020).
Topic modeling results in two types of output: one that links words with their propensity of occurring within a topic $k \in \{1, 2, \ldots, K\}$, $K \in \mathbb{N}$, and one stating the extent to which documents discuss each topic. This projection of texts into a K-dimensional latent space ($K \ll V$) is a purely mathematical operation and the assessment of interpretability is up to human judgment. Usually the resulting topics are then examined with respect to their most characteristic terms, according to an appropriate measure, in the attempt to find a meaningful description. In particular, K is a hyperparameter that must be specified a priori (Aggarwal, 2018).

Approaches. Topic modeling approaches roughly decompose into deterministic and probabilistic, or generative, approaches. The former are based on factorizing the document-term matrix $M \in \mathbb{R}^{D \times V}$ (or a weighted version that takes into account prior probabilities of term occurrence) into two low-rank matrices, $U \in \mathbb{R}^{D \times K}$ and $W^T \in \mathbb{R}^{K \times V}$, whose product approximates M loss-minimally. Probably the most prominent methods from this category are *latent semantic analysis*, which performs singular value decomposition and thus projects the data into a subspace spanned by M's principal eigenvectors, and *non-negative matrix factorization*, a constrained version that often yields better interpretability (Aggarwal, 2018).
Non-probabilistic models suffer from limitations in inference and out-of-sample extension, which is why generative approaches, addressing these issues, have become widely popular. Generative models hail from the Bayesian paradigm. Loosely speaking, they seek to reverse-engineer the

imaginative process of document generation: first, for each document d in a corpus we draw a length-K vector of topic proportions from some distribution; then assign each word position in $\{1, 2, \dots, N_d\}$ to a topic with probabilities according to the sampled topic proportions, and then draw a word from the distribution associated with this topic (Vayansky and Kumar, 2020). *Latent Dirichlet allocation (LDA)* by Blei et al. (2003), employing Dirichlet and multinomial distributions, pioneered this approach to topic modeling. We revisit LDA in section **??** as it also provides the foundation for the STM.

3.2.2 Sentiment Analysis

Idea. Sentiment analysis refers to the process of establishing the emotive nature of the opinion an author expresses in their text. The types of sentiment that are of interest vary across applications and may be arbitrarily fine-grained. We focus on the simplest form and attempt to determine whether tweets convey positive or negative sentiment, also called *polarities*. In this, sentiment analysis is a standard classification problem (Medhat et al., 2014).
Sentiment analysis occurs at different levels of a document. Depending on its scope we can broadly discern document-level, sentence-level and aspect-level sentiment analysis. The document and sentence level techniques are not fundamentally distinct; rather, document-level analysis follows after sentence-level analysis (it is also possible to drill down even further) and aggregates the sentiments expressed by sentences for the entire document. Aspect-based sentiment analysis (ABSA) has a slightly different angle. It studies sentiment regarding aspects of topical entities (for instance, *carbon taxes* as an aspect of *environmental policy*), meaning sentiment targets are identified with respect to contents and not merely by sentence delimiters. This requires solving the sub-tasks of aspect extraction and aspect sentiment classification (Aggarwal, 2018). ABSA therefore has connections to topic modeling. We will discuss our view on the relation of topic modeling, sentiment analysis and ABSA in the subsequent section.

Approaches. There are two principal ways of approaching sentiment analysis. The first is rule-based and avoids statistical modeling altogether, instead classifying documents by summing the number of terms associated with each sentiment and assigning the class with the highest count. Prior polarities are typically taken from large dictionaries (Sidarenka, 2019). In our binary task this corresponds to identifying whether tweets contain more words with positive or negative connotation. We do incorporate this type of analysis but use the respective numbers of positive- and negative-polarity terms as an intermediate input rather than the sole grounds for classification.
Instead, we perceive sentiment analysis as a classic instance of supervised ML. Under the assumption that the data can be categorized into $g \in \mathbb{N}$ discrete classes we predict for each document its associated class from a set of features. More formally, we find a model $f : \mathcal{X} \to \mathbb{R}^g$, $\mathcal{X} \subseteq \mathbb{R}^p$ for $p \in \mathbb{N}$, that maps from the space of input features into g-dimensional Euclidean space. Each observation is assigned a vector of continuous class scores or probabilities (depending on the classifier). This mapping must be learned from a set of labeled training data, which marks a fundamental difference to the topic modeling task. The actual class labels $y \in \mathcal{Y}$ are then found by thresholding or an *argmax* operation on the score/probability vectors (Bishop, 2006). In our case the set of labels, with $g = 2$, is typically encoded as $\mathcal{Y} = \{0, 1\}$ or $\mathcal{Y} = \{-1, 1\}$.
Once the data are available in appropriate form, we can, in principle, use any type of learner suited to classification. We specifically consider random forests and regularized logistic regression for the standard ML solution and turn BERT into a classifier by fine-tuning it to sentiment analysis; details are given in section 4.

3.2.3 Topic-Specific Sentiment Analysis

With the concepts of topic modeling and sentiment analysis we define a combined task we call *topic-specific sentiment analysis*. There are various ways in which such a combination is conceivable and we ourselves pursue different approaches, which is why we use TSSA as an overarching term. We see two types of solution to the TSSA problem: either perform topic extraction and sentiment analysis jointly, or regard it as a cascading task where sentiment analysis is subsequent to topic modeling. Most publications to date share the second view; recent work, however, has highlighted the advantages of joint modeling (e.g., Ngyuen and Shirai (2018), Tian et al. (2021), Wang et al. (2021)). While theoretical arguments are strong, we decide against simultaneous methods mostly due to their complexity which is at odds with our goal of making text analysis accessible to a broader audience. Also we would like to retain the option of performing both tasks in a stand-alone fashion. The latter has become all the more important in the course of our study as we experience some difficulties in modeling topics for tweets.
We make the general assumption of each document pertaining to exactly one topic. This seems plausible in the light of tweets' brevity (Twitter currently allows for 280 characters, up from 140 originally) and is supported by other research (Zuo et al. (2016), Qiang et al. (2019)). We then explore different options in our proposal. For the standard ML classification we use topic modeling, somewhat implicitly, as a means of feature generation for sentiment analysis. The clusters resulting from grouping tweets by topic serve as environments for computing topic-specific embeddings that are afterwards fed to the classifier. In the DL solution we experiment with sentiments that directly refer to aspects by fine-tuning BERT to an ABSA task. However, we observe in both approaches that the topical component does not aid the sentiment classification task and even tends to worsen results. We will discuss this finding in section 6.

4 Analytical Proposal

4.1 Data

4.1.1 Data Collection

The subject of our analysis are tweets by members of the German parliament (*Bundestag*) issued after the last federal election in September 2017[1]. Twitter makes these publicly accessible via its official API and the number of retrievable tweets per user can be exploited generously, so data supply is almost unrestricted. However, with sentiment classification as ultimate goal of analysis, we face a major bottleneck in the need for labeled data. Lacking the resources for large-scale annotation we did the labeling by hand. The resulting data set, from a vast amount of available data, consists of 1,215 observations, imposing some practical limits on the analytical scope.

Web scraping. For data collection from the Web we rely on the scraping procedure developed in the predecessor project, with minor modifications. The process entails four steps: first, gather MPs' names and basic information (such as party affiliation and electoral district) from the official Bundestag website; second, find Twitter account names (using individual party websites as additional sources); third, acquire socioeconomic information for the time of the last federal election on a per-district level (available at the official website of the federal returning officer); and, lastly, scrape actual tweets along with some additional variables like the number of retweets. We use a `Python` code base and mainly employ selenium webdrivers as well as the `BeatifulSoup` library (Richardson, 2007) for parsing HTML content and the `tweepy` library (Roesslein, 2020)

[1]The 2017 Bundestag is comprised of 709 seats and seven political parties: the right-wing AfD, the Christian Democrats (CDU/CSU), the Liberals (FDP), the Greens, the Left Party, and the Social Democrats (SPD). CDU/CSU and SPD as ruling parties co-exist in a grand coalition.

for accessing the official Twitter API. For more details on the procedure and the large data base assembled in the predecessor project please refer to Schulze and Wiegrebe (2020); the code is fully submitted in our electronic appendix and a somewhat more compact demo may be found among the teaching material.

Data labeling. In the data annotation phase we extracted a set with some tens of thousands of observations according to the above process and manually selected what we deem informative examples. For these we assigned polarities, i.e., predicates *positive* or *negative*, and also topic descriptions required for BERT's ABSA task. We noted in the process that a large number of tweets do not appear to carry sentiment at all. The resulting 1,215 training observations, originated by a total of 256 MPs, date from the period of October 2017 to January 2021. In figure 1 we detect both periodical fluctuations in the number of tweets over time and a general upward-sloping trend.

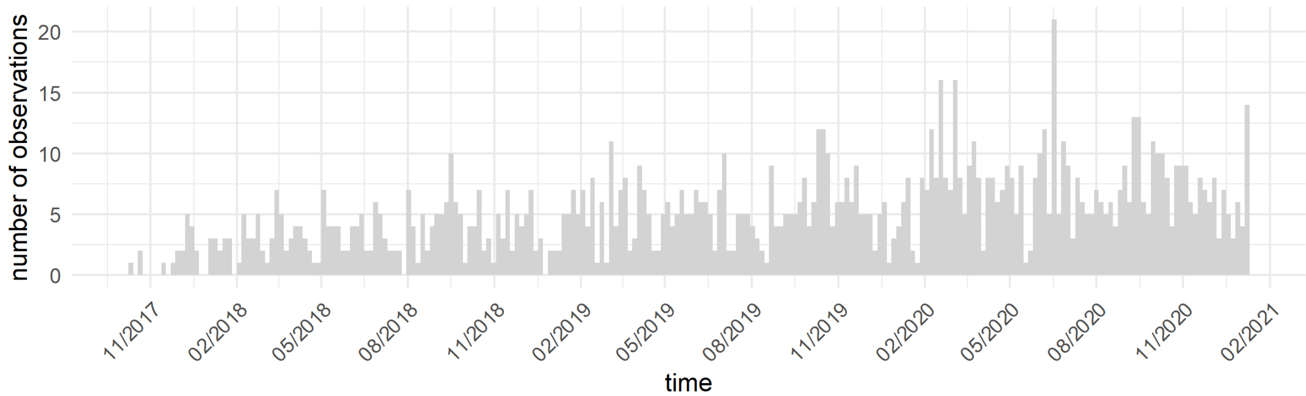

Figure 1: Observations over time.

An exemplary extract from our training data with some of the most important variables is shown in table 1. Exactly which features enter sentiment classification is documented in the subsequent chapters.

username	party	created_at	text	followers	unemployment_rate	label
karl_lauterbach	spd	2019-12-01 09:44:00	"Die Wahl ..."	337001	8.5	negative
Martin_Hess_AfD	afd	2018-08-17 07:15:00	"Vor den ..."	6574	3.5	negative
BriHasselmann	gruene	2019-09-25 15:35:00	"Ich finde ..."	20299	8.6	positive
danielakolbe	spd	2020-05-12 06:05:00	"Aber verpflichtend ..."	8158	8.3	negative
JuergenBraunAfD	afd	2020-08-13 22:05:00	"Panik-Latif + ..."	3188	3.4	negative

Table 1: Training data extract for selected variables.

Figure 2 depicts the number of observations per party both for our labeled training data and the larger sample from which the training data have been selected (containing just over 31,000 tweets). We notice two things. First, in either case, the share of tweets (blue) does not mirror the share of seats in the Bundestag (gray); most notably, the Christian Democrats tweet rather little, whereas the Greens are disproportionately active on Twitter. Second, the right-wing AfD and the Greens are over-represented in our training data at the expense of the other groups. This is simply because these two parties, in our personal experience from the annotation process, more often issue tweets that are strongly opinionated.

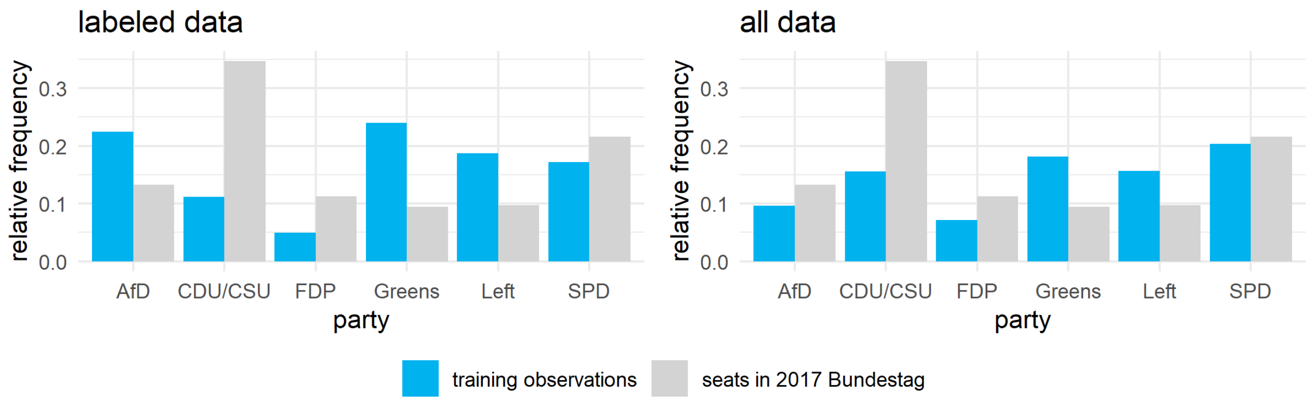

Figure 2: Observations per party in labeled training data (*left*) and entire scraped data example (*right*), both depicted against seat distribution in current parliament.

Lastly, when we inspect the class label distribution in the training data, an imbalance favoring the negative class becomes immediately visible: some 72% of tweets have been marked as negative. This reflects our general impression that most tweets which do carry sentiment express negative opinions and might be partly appropriated to the fact that the majority of authors belong to opposition parties.

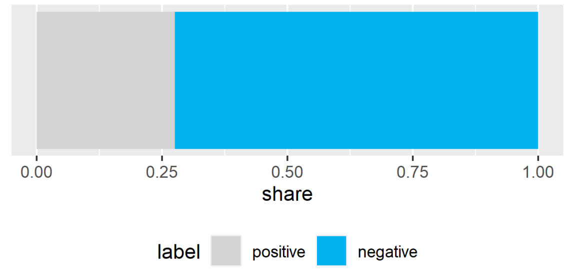

Figure 3: Distribution of class labels.

4.1.2 Data Pre-Processing

The standard ML approach requires a lot more feature engineering than BERT does, but some general pre-processing steps are applied for both. In an initial step all tweets in non-German language are excluded from the data. We proceed with basic text cleaning, namely transcription of German umlauts and ligature s into standard-Latin characters and removal of non-informative symbols (such as those induced by ampersand conversion). The next block of operations is specific to Twitter data and includes the identification, separate storage and subsequent removal of special characters such as hashtags, emojis and user tags. By this we ensure the data are available for explicit analysis but do not introduce noise in the text. We finish the pre-processing procedure by assigning a unique identifier to each tweet.

4.1.3 Challenges

Text data come with many idiosyncrasies to begin with: language is highly diverse, irregular, and subject to constant change. Contextual dependencies and complex constructs such as colloquialisms or sarcasm pose serious obstacles in NLP, besides which profanities like spelling or translation mistakes must be handled (Mohammad, 2017). Some particular properties of the data at hand add to these challenges.

Language-specific. Not surprisingly, most work in NLP is concerned with the analysis of English documents. Although German is not a low-resource language and attracts its own share of research, many analyses and tools are predominantly tailored to English. German grammar is another aspect that needs to be considered. Syntax is heterogeneous, and inflections due to cases and genera result in many variations of lexical lemmata (Rauh, 2018).

Twitter-specific. Tweets' brevity is arguably the most critical issue for analysis. The limit of 280 characters means that words rarely appear more than once and we cannot expect many indicators of topics or sentiments in each document. It also prompts the use of abbreviations. On a similar note, we observe that Twitter posts often refer to certain events or topical entities without explicitly mentioning them, which is probably both due to the character limit and the real-time character of publications. The message may be clear for an informed human annotator then but will be hard to grasp for machines. Furthermore, tweets tend to be of rather informal style and use language that appears almost exclusively in social media, enlarging the vocabulary the classifier must understand.

Context-specific. The context of our data lessens the degree of informality somewhat; the issued documents are mostly political statements and as such more akin to written texts from other sources. Still, the political domain introduces new vocabulary yet again and makes the transfer of knowledge from other contexts harder. Lastly, as mentioned before, we find many tweets to be solely informative and detect an imbalance toward negative sentiment in those that do convey opinion.

4.2 Standard Machine Learning Solution

4.2.1 Feature Extraction

Static feature extraction forms the first block of the standard ML procedure and provides tabular data which then enter an automated ML (AutoML) pipeline. We refer to these features as static because they can be calculated prior to any training process, depending solely on single observations. Dynamic features, by contrast, are computed globally across several observations, which requires strict separation between train and test sphere in their computation.

Static features

The static part of features is based on the so-called *bag-of-words (BOW)* assumption that leads to texts being treated as arbitrary collections of vocabulary instances. In particular, information about grammar and word order is discarded in BOW approaches. This is obviously a strong simplification but hard to avoid entirely with standard classifiers (Cambria et al., 2017).
The static feature extraction steps rely heavily on `R`'s `quanteda` package (Benoit et al., 2021) for organizing documents in corpus objects, tokenizing texts and performing look-ups with dictionaries. We use lists of stopwords in multiple places to exclude uninformative tokens such as determiners or auxiliary verbs. These are compilations from various sources, including `quanteda`'s built-in list, open-source data available on GitHub and some manually appended, domain-specific terms. In addition, we repeatedly apply *stemming* in order to reduce words to their root form (for instance, cutting back both *"works"* and *"working"* to *"work"*).
All of these operations are aimed at compressing document representation from the total of unique original terms to more generalized tokens that co-occur across texts This way we create features that are actually shared by multiple observations and thus help classifiers infer feature-target relations with the ability to generalize. For the static part we exploit insights from other studies (e.g., Baly et al. (2017), Correa et al. (2017), Jabreel and Moreno (2017), Sidarenka (2019)) and include the following features:

1. **Lexicon-based polarity counts.** These comprise two sources of prior polarity, namely words and emojis. We use the aggregate of three large German polarity-term dictionaries, *GlobalPolarityClues* (Waltinger, 2010), *SentiWS* made available online by Leipzig university and a collection by Rauh (2018), for word polarities. The resulting lexicon distinguishes weak

and strong sentiment. In each tweet, if a word is part of the dictionary and judged to have either positive or negative connotation (weak or strong), it raises the respective polarity count by one. Emojis are treated analogously, albeit without the discrimination between degrees of sentiment, using an emoji polarity list accumulated by Kralj Novak et al. (2015).

2. **Twitter variables.** We make use of the additional information provided by the Twitter API and note for each tweet the number of likes and retweets as well as the number of users tagged. For hashtags we resort to the same naive measure since we find them to be so heterogeneous across documents in our training set that no meaningful information can be extracted. This certainly marks an opportunity for future improvement.

3. **Syntactic features.** This group attempts to mitigate the simplification introduced by the BOW assumption to some extent.

Dynamic features

Foo

Structural Topic Model. Foo

Word Embeddings. Foo

4.2.2 Sentiment Classifiers

Random Forests. foo

Regularized Logistic Regression. foo

4.2.3 Automated Machine Learning Pipeline

The design of the AutoML pipeline as a graph learner ensures that training and test sets remain completely walled off from one another at each resampling step.

4.2.4 Results

4.3 Deep Learning Solution

4.3.1 Methodology

Deep transfer learning

foo

BERT

foo

4.3.2 Results

5 Knowledge Transfer

5.1 Static Material

5.2 Live Teaching

6 Discussion

6.1 Analytical Proposal

6.2 Knowledge Transfer

7 Conclusion

foo

A Appendix

B Electronic Appendix

Data, code and figures are provided in electronic form.

References

Aggarwal, C. C. (2018). *Machine Learning for Text*, Springer.

Baly, R., Badaro, G., Hamdi, A., Moukalled, R., Aoun, R., El-Khoury, G., El-Sallab, A., Hajj, H., Habash, N., Shaban, K. B. and El-Hajj, W. (2017). Omam at semeval-2017 task 4: Evaluation of english state-of-the-art sentiment analysis models for arabic and a new topic-based model, *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*, p. 603–610.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., Lowe, W. and Müller, C. (2021). *quanteda: Quantitative Analysis of Textual Data*. R package version 3.0.0.
URL: *https://CRAN.R-project.org/package=quanteda*

Bhatia, S. and Padmanabhan, D. (2018). Topic-specific sentiment analysis can help identify political ideology, *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 79–84.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Springer.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation, *Journal of Machine Learning Research* **3**: 993–1022.

Bode, L. (2017). Gateway political behaviors: The frequency and consequences of low-cost political engagement on social media, *Social Media + Society* **3**.

Cambria, E., Das, D., Sivaji and Feraco, B. A. (2017). Challenges in sentiment analysis, *in* E. Cambria, D. Das, Sivaji and B. A. Feraco (eds), *Affective Computing and Sentiment Analysis*, Springer.

Correa, E. A., Marinho, V. Q. and dos Santos, L. B. (2017). Nilc-usp at semeval-2017 task 4: A multi-view ensemble for twitter sentiment analysis, *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*, p. 611–615.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding.

Ficamos, P. and Liu, Y. (2016). A topic based approach for sentiment analysis on twitter data, *International Journal of Advanced Computer Science and Applications* **7**.

Hasell, A. and Weeks, B. E. (2016). Partisan provocation: The role of partisan news use and emotional responses in political information sharing in social media, *Human Communication Research* **42**: 641–661.

Jabreel, M. and Moreno, A. (2017). Sitaka at semeval-2017 task 4: Sentiment analysis in twitter based on a rich set of features, *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*, p. 694–699.

Jang, H., Rempel, E., Roth, D., Carenini, G. and Janjua, N. Z. (2021). Tracking covid-19 discourse on twitter in north america: Infodemiology study using topic modeling and aspect-based sentiment analysis, *Journal of Medical Internet Research* **23**(2).

Jost, J. T., Barbera, P., Bonneau, R., Langer, M., Metzger, M., Nagler, J., Sterling, J. and Tucker, J. A. (2018). How social media facilitates political protest: Information, motivation, and social networks, *Advances in Political Psychology* **39**(1).

Kralj Novak, P., Smailović, J., Sluban, B. and Mozetič, I. (2015). Emoji sentiment ranking 1.0, https://www.clarin.si/repository/xmlui/handle/11356/1048.

Leipzig, U. (2021). Leipzig corpora collection download page, https://wortschatz.uni-leipzig.de/en/download.

Medhat, W., Hassan, A. and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal* **5**: 1093–1113.

Mohammad, S. M. (2017). Challenges in sentiment analysis, *in* E. Cambria, D. Das, Sivaji and B. A. Feraco (eds), *A Practical Guide to Sentiment Analysis*, Springer.

Ngyuen, H. and Shirai, K. (2018). A joint model of term extraction and polarity classification for aspect-based sentiment analysis, *Proceedings of the 10th International Conference on Knowledge and Systems Engineering (KSE).*

Qiang, J., Qian, Z., Li, Y., Yuan, Y. and Wu, X. (2019). Short text topic modeling techniques, applications, and performance: A survey, *Journal of Latex Class Files* **14**(8).

R Core Team (2021). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rauh, C. (2018). Replication data for: Validating a sentiment dictionary for german political language, https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BKBXWD.

Richardson, L. (2007). *Beautiful Soup Documentation.*

Roberts, M., Stewart, B., Tingley, D. and Airoldi, E. (2013). The structural topic model and applied social science, *Advances in Neural Information Processing Systems Workshop on Topic Models*, pp. 1–20.

Roesslein, J. (2020). *Tweepy: Twitter for Python!*

Schulze, P. and Wiegrebe, S. (2020). Twitter in the parliament - a text-based analysis of german political entities, *Technical report*, Ludwig-Maximilians-Universität, Munich.

Schulze, P., Wiegrebe, S., Thurner, P. W., Heumann, C., Aßenmacher, M. and Wankmüller, S. (2021). Exploring topic-metadata relationships with the stm: A bayesian approach.

Sidarenka, U. (2019). *Sentiment Analysis of German Twitter*, PhD thesis, University of Potsdam.

Stier, S., Bleier, A., Lietz, H. and Strohmaier, M. (2018). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on facebook and twitter, *Political Communication* **35**(1): 50–74.

stopwords-iso (2021). https://github.com/stopwords-iso/stopwords-de.

Tian, Y., Yang, L., Sun, Y. and Liu, D. (2021). Cross-domain end-to-end aspect-based sentiment analysis with domain-dependent embeddings, *Comlexity* .

Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N. and Fox, E. A. (2020). Natural language processing advancements by deep learning: A survey, *CoRR* .
URL: *https://arxiv.org/abs/2003.01200*

van Rossum, G. and Drake, F. L. (2011). *The Python Language Reference Manual*, Network Theory Ltd.

van Vliet, L., Törnberg, P. and Uitermark, J. (2020). The twitter parliamentarian database: Analyzing twitter politics across 26 countries, *PLoS ONE* **15**(9).

Vayansky, I. and Kumar, S. A. (2020). A review of topic modeling methods, *Information Systems* **94**.

Waltinger, U. (2010). Germanpolarityclues: A lexical resource for german sentiment analysis, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*, electronic proceedings, Valletta, Malta.

Wang, X., Xu, G., Zhang, Z., Jin, L. and Sun, X. (2021). End-to-end aspect-based sentiment analysis with hierarchical multi-task learning, *Neurocomputing* .

Zuo, Y., Wu, J., Zhang, H., Lin, H., Wang, F., Xu, K. and Xiong, H. (2016). Topic modeling of short texts: A pseudo-document view, *Proceedings of KDD '16*.

Declaration of Authorship

We hereby declare that the report submitted is our own unaided work. All direct or indirect sources used are acknowledged as references. We are aware that the report in digital form can be examined for the use of unauthorized aid and in order to determine whether the report as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of our work with existing sources we agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future reports submitted. Further rights of reproduction and usage, however, are not granted here. This paper was not previously presented to another examination board and has not been published.