Political Stance Detection for Danish

Rasmus Lehmann
ITU Copenhagen
Denmark
rasmus_lehmann@hotmail.com

Leon Derczynski
ITU Copenhagen
Denmark
ld@itu.dk

Abstract

The task of stance detection consists of classifying the opinion expressed within a text towards some target. This paper presents a dataset of quotes from Danish politicians, labelled for stance, and also stance detection results in this context. Two deep learning-based models are designed, implemented and optimized for political stance detection. The simplest model design, applying no conditionality, and word embeddings averaged across quotes, yields the strongest results. Furthermore, it was found that inclusion of the quote's utterer and the party affiliation of the quoted politician, greatly improved performance of the strongest model.

Dansk abstrakt: I indeværende artikel præsenteres et annoteret datasæt over citater fra danske politikere, samt to Deep Learning-baserede modeller til brug ved identifikation af holdninger i de annoterede citater. Det konkluderes at den simpleste af de to modeller opnår de bedste resultater, samt at brug af information vedrørende citaternes kontekst forbedrer modellernes resultater.

1 Introduction

As a result of digitalization, the availability of information regarding the state of politics has never been greater, interviews, debates, party programs and articles all readily available online. This can be seen as a democratic benefit, contributing to the enlightenment of the population, giving individuals a basis on which to form their opinions and place their votes. However, the large amount of information available means the time required for keeping up to date on the state of politics becomes increasingly higher. A partial solution to this problem is to convert textual data into quantitative data, representing a large amount of text in a more compact fashion. This can be achieved using Natural Language Processing (NLP), the field concerned with the automatic parsing, analysis and understanding of text. Within this field is the task of stance detection, concerned with discerning the stance in a text towards some target. Building a model which can accurately solve the task of stance detection can help generate quantitative data regarding the state of Danish politics.

The objective of this work is two-fold; creating a dataset of quotes from politicians labelled for stance, allowing statistical analysis of opinions within parties and for each politician, and building a machine learning-based stance detection model, able to determine the stances within quotes in the generated dataset.

The task of collecting data for the dataset is defined as the extraction of quotes from news articles for all political parties within the Danish parliament. Here, considerations are made regarding the objectivity of the collected data, both taking into account the subjectivity of journalists, media outlets and the researcher.

The task of data labelling will be performed using the labels *for*, *against* and *neutral*. For this task, the subjectivity of the researcher is the primary concern, in regards to the objectivity and general applicability of the dataset.

The task of stance detection is defined as the automatic detection of a stance within a given quote towards some target, using the stance classes *for* and *against* the target, or as neither *for* nor *against* the target, which we call *neutral*. The goal of this work is to create a model which can perform this task, both to be used as a tool for political analysis and to expand the generated dataset by automatic labelling of quotes, as well as to be used as a benchmark for further research within the field of NLP in Danish.

Proceedings of the 22nd Nordic Conference on Computational Linguistics (NoDaLiDa), pages 197–207
Turku, Finland, 30 September – 2 October,

2 Related Work

Stance detection has been addressed through a number of different model approaches, including probabilistic classifiers (Qazvinian et al., 2011), kernel-based classifiers (Mohammad et al., 2017; Enayet and El-Beltagy, 2017; Collins and Duffy, 2001) and ensemble learners (Zeng et al., 2016; Tutek et al., 2016). Recently, deep learning approaches have shown promise at this task. The two top performing teams of SemEval 2016 Task 6 both applied deep learning models (Zarrella and Marsh, 2016; Wei et al., 2016) as did those in RumourEval 2019 (Gorrell et al., 2019).

The task of stance detection has been applied widely within political analysis, both analyzing the stance of politicians towards a given topic (Lai et al., 2016; Skeppstedt et al., 2017), as is the task within this paper, and also to identify the stance of individuals towards some politician or policy (Aker et al., 2017; Augenstein et al., 2016; Mohammad et al., 2016; Johnson and Goldwasser, 2016; Iyyer et al., 2014). For several of these cases, the stance target has been mentioned explicitly in the data. This is not necessarily the case for the dataset generated for this paper, increasing the difficulty of the task significantly. Furthermore, all of these examples perform stance detection for English, whereas the dataset generated for this data is in Danish. This further increases the difficulty of the task, as fewer resources are available.

Enevoldsen and Hansen (2017) perform sentiment analysis in Danish using newspaper articles, using the AFINN dictionary over sentiment of Danish words (Årup Nielsen, 2011), performing ternary classification of articles using *for*, *against* and *neutral* labels. However, no research has been done within political stance detection in Danish (Kirkedal et al., 2019), and only very recently has any work been done for stance in Danish in the first place – just Lillie et al. (2019), published at the same time as this paper.

3 Data

We assembled a dataset of quotes from Danish politicians, extracted from articles from the Danish media outlet Ritzau. Considerations were made regarding the objectivity of the collected data, and seeing as Ritzau is owned by a conglomerate of media outlets from all areas of the political spectrum (Ritzau, 2019), it is assumed that articles from the media outlet will not contain bias towards any given party. A data statement (Bender and Friedman, 2018) is in the appendix.

A shortlist of possible topics to include in the dataset was attained based on an opinion poll executed by Kvalvik (2017), seeking to identify the topics most important to the Danish population, when voting in the next election. Here, the five most important topics were identified as health policy, social policy, immigration policy, crime and justice policy and finally environment and climate policy. Immigration policy was chosen as the topic to be included in the dataset, due to alternative topics being defined too broadly to easily allow a clear definition of annotation guidelines.

3.1 Choice of Politicians

To accurately represent the full spectrum of Danish legislative politics, politicians from all political parties with seats in parliament are included in the dataset. From each party, ten politicians have been chosen for inclusion in the dataset. Politicians with seats in parliament have been prioritized over those without seats. For the parties with more than ten politicians in parliament, prioritization has been made as follows:

1. Ministers
2. Party heads
3. Speakers
    (a) Speakers within the five top topics of interest to the Danish population as presented by (Kvalvik, 2017)
    (b) Speakers not within the five top topics
4. Non-speaker Members of parliament

Considerations were made regarding the gender representativity within the dataset. The metrics just described yields the gender distribution presented in Table 1. It can be observed that the approach creates a skewed gender distribution of included politicians, but the skewness is judged to be within a reasonable margin, with 58% male and 42% female politicians.

3.2 Data Labelling

The choice of labelling convention is based on that applied by Mohammad et al. (2016) in organizing SemEval-2016 Task 6, which is concerned with the detection of stance within tweets, and the creation of a dataset for this task. Three classes are defined along which quotes are labelled, the first called *for*, declaring support of a given topic, the

Party	Males Count	%	Females Count	%
Alternativet	7	70	3	30
Dansk Folkeparti	5	50	5	50
Det Konservative Folkeparti	6	60	4	40
Enhedslisten	7	70	3	30
Liberal Alliance	6	60	4	40
Radikale Venstre	5	50	5	50
Socialdemokratiet	7	70	3	30
Socialistisk Folkeparti	3	30	7	70
Venstre	6	60	4	40
Total	**52**	**58**	**38**	**42**

Table 1: Gender distribution of dataset per party

second called *against*, declaring opposition to a given topic, and a third, called *neutral*, contains both quotes that are deemed to be neutral towards the topic, as well as quotes for which a specific stance can not be determined. During the initial labelling efforts, it was observed that not all gathered quotes could be categorized along the same axis, and it was therefore decided to divide the dataset into sub-topics.

3.2.1 Defining Sub-topics

A clear division of the dataset was found along whether the quote concerned immigration issues in the context of within the borders of Denmark, or in a more global context. The sub-topic National immigration policy (*national policy* for short) was defined as policy and topics that concern matters within Danish borders, such as the number of asylum seekers the country takes in, how these are housed, and what requirements should be set for them in regards to taking Danish education and employment. An example of a quote within this subtopic can be found below, which concerns the government's initiative to combat communities that they define as ghettos.

> *Det er godt, at der lægges op til højere straffe og en styrket politiindsats i ghettoer. Men regeringen skal passe på ikke at oversælge sit udspil. Det kan ikke løse alle problemer.*
>
> *It is good that harsher penalties and an increased police effort in ghettos is encouraged. But the government should be careful not to oversell its proposal. It can not solve all problems.*
> Martin Henriksen, (Ritzau, 2018b)

Centralized immigration policy (*centralization* for short) is defined as policy and topics that concern immigration on a European or international level, for example distribution of asylum seekers among the member countries of EU, deterring immigrants at EU's boarders or the sending of immigrant from Denmark to refugee camps in other countries. An example of such a quote is:

> *Den danske regering bør i stedet sige til den italienske regering, at Danmark og Italien i fællesskab kan transportere asylansøgerne tilbage til Afrika, så de kan blive sat af på kyste.*
>
> *The Danish government should instead tell the Italian government that Denmark and Italy can transport the asylum seekers back to Africa together, where they can be set ashore on the coast.*
> Martin Henriksen, (Ritzau, 2018f)

Some quotes fit both subtopics. In these cases, a duplicate quote is created, and one is labelled with each subtopic. An example of this is found below, where first half of the quote is concerning the free mobility of labour within EU, and immigration stemming from this, and the second half is concerned with the effect on immigration legislation changes on a Danish level.

> *Det er oplagt at se på, hvordan vi kan understøtte en højere grad af mobilitet i Europa, så danske virksomheder, der har brug for arbejdskraft, kan få den, uden det betyder den indvandring, som vil følge af at sætte beløbsgrænsen ned.*
>
> *It would be natural to look at, how we can support a higher level of mobility in Europe, so Danish companies that need laborers can get them, without it resulting in the immigration, which would result from lowering the threshold.*
> Mette Frederiksen, (Ritzau, 2018d)

3.2.2 Annotation Guidelines

The subtopic *national policy* is defined as tightening the policy within the borders of Denmark on the legislative fields of immigration, integration and asylum. Therefore, a quote would be classified as *for* this topic, if it exhibits one or more of the following traits.

- support for higher restrictions on immigrants or asylum seekers entering the country
- support for lowering public benefits to immigrants or asylum seekers
- a wish to get immigrants or asylum seekers to leave Denmark, after they have entered the country
- making demands specifically of immigrants or asylum seekers, for instance regarding taking language courses or job search

- seeking to make immigrants or asylum seekers change their culture or behaviour
- communicating explicitly or implicitly that immigration is a burden to Danish society
- wishing to implement changes in behaviour though negative incentives such as decreased public benefits

Quotes classified as *against* the *national policy* subtopic, on the other hand, will exhibit one or more of the following traits.

- support for lower restrictions on immigrants or asylum seekers entering the country
- support for higher public benefits to immigrants or asylum seekers
- immigrants or asylum seekers are free to stay, after having entered the country
- seeking to making fewer demands of, and give more freedom to, immigrants or asylum seekers
- not seeking to make immigrants or asylum seekers change their culture or behaviour
- communicating explicitly or implicitly that immigration is an asset to Danish society
- wishing to implement changes in behaviour though positive incentives such as increased public benefits

The subtopic *centralization* is defined as yielding decision power to EU, and/or solving more immigration issues on a European or international level, rather than on a national level, and *for* and *against* labels are thus more clearly defined for this subtopic. A *for* quote would support yielding power, an example of which is found below.

> *Europa har en fælles udfordring med flygtninge og migranter. Vi må have et fælles asylsystem.*
>
> *Europe has a mutual challenge with refugees and immigrants. We must have a common asylum system.*
> Rasmus Nordqvist, (Ritzau, 2018a)

On the other hand, an *against* quote would be opposed to yielding power, an example of which is found below.

> *Der er for mange spørgsmål, som står ubesvaret hen, og derfor mener vi, at man fra dansk side skal suspendere det samarbejde, indtil der er fuldstændig klarhed over, hvad regeringen har forpligtet sig til på Danmarks vegne.*
>
> *There are too many questions left unanswered, and therefore we believe that Denmark should suspend the collaborative efforts, until there is complete clarity regarding what the government has committed itself to on behalf of Denmark.*
> Martin Henriksen, (Ritzau, 2018g)

3.2.3 Resolving Grey Areas

Not all quotes contain explicit communication of a stance or even clear indicators, like the ones just described. To solve this issue, inspiration is taken from Mohammad et al. (2016), and the questions given to annotators. In line with Mohammad et al. (2016), when labelling quotes, stance is inferred from how the quotee refers to things and people aligned with or opposed to the topic. An example of this would be a politician indicating support towards a ban on the use of burkas, which falls within the subtopic of *national policy*. Seeing as the ban on burkas is a restriction of behaviour, the quote would be labelled as *for*, as the stance of the quote can be induced by proxy. Furthermore, when no clear stance is communicated, and no stance can be determined by proxy, the tone of the quote is analyzed, looking at the use of weighted words, for instance describing immigrants as resources, nuisances or in neutral terms.

4 Annotated Dataset

Looking at the quote count for the dataset as presented in Table 2, it is clear that the dataset is significantly skewed towards the *for* label, containing 57.2% of the quotes, with 23.4% labeled as *against* and 19.3% as neutral when observing the full dataset, and the skewness remains if looking at the two subsets in isolation. Such skewness has shown to be an issue for stance detection models in earlier research, an example of this being the SemEval-2017 competition Task 8, where the dataset contained a majority label with 66% of the data points in the train set and 74% of data points in the test set (Derczynski et al., 2017).

Another potential issue for the stance detection task is the size of the dataset, as a size of 898 instances might not be sufficient to learn the language patterns within the quotes.

4.1 Assessing Representativity in the Dataset

Out of the 90 politicians chosen to be included in the dataset, relevant quotes were only found in the

Party	Topic	# Quotes			
		F	A	N	Total
Alternativet	NP		7	2	9
	C	2			2
	Total	2	7	2	11
Dansk Folkeparti	NP	187	5	25	217
	C	5	18	7	30
	Total	192	23	32	247
Det Konservative Folkeparti	NP	18	1	5	24
	C	2			2
	Total	20	1	5	26
Enhedslisten	NP	3	26	5	34
	C		4	1	5
	Total	3	30	6	39
Liberal Alliance	NP	6	6	6	18
	C				0
	Total	6	6	6	18
Radikale Venstre	NP	7	68	18	93
	C	6		1	7
	Total	13	68	19	100
Socialdemokratiet	NP	92	20	42	154
	C	7	1	1	9
	Total	99	21	43	163
Socialistisk Folkeparti	NP	5	26	2	33
	C	2			2
	Total	7	26	2	35
Venstre	NP	144	14	54	212
	C	38	1	8	47
	Total	182	15	62	259
All parties	NP	462	173	159	**794**
	C	62	24	18	**104**
	Total	**524**	**197**	**177**	**898**

Table 2: Quote count overview for dataset, NP denoting *national policy*, C denoting *centralization*, F denoting For, A denoting Against and N denoting Neutral.

Ritzau database for 63 politicians. This might constitute an issue in terms of representativity, if a certain gender, party or political orientation is more likely to be quoted by news outlets.

Dividing parties based on their placement on the political axis, defining Alternativet, Enhedslisten, Radikale Venstre, Socialdemokratiet and Socialistisk Folkeparti as left-wing parties and Dansk Folkeparti, Det Konservative Folkeparti, Liberal Alliance and Venstre as right-wing parties, a skewness towards the right-wing parties within the dataset can be observed, as seen in Table 3. We observe an over-representation of right-wing parties with 61% of the quotes.

	Quote #		
	Left-wing	Right-wing	Total
For	400	124	479
Neutral	105	72	177
Against	45	152	197
Total	**550**	**348**	**898**

Table 3: Quote count divided by political axis

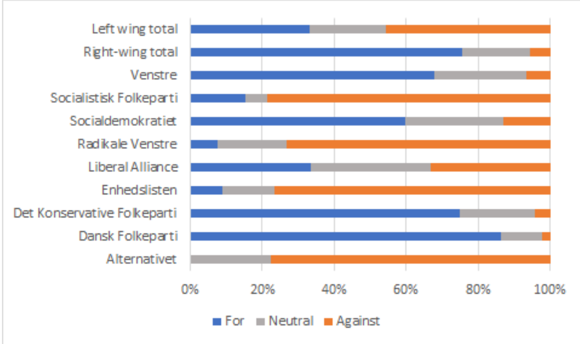

Figure 1: Quote distribution for the subtopic *national policy* between the labels for, against and neutral, for each party, in percentage, totals calculated as sums of quotes

Similarly, a skewness towards the male gender can be observed in the data, shown in Table 4. This is, however, likely to be a reflection of the skewness in the number of male and female politicians included in the party, observed in Table 1.

	Gender	
	Male	Female
For	316	208
Against	124	72
Neutral	98	80
Total	**538**	**360**
%	60	40

Table 4: Quote count divided by gender

The skewness of data, as presented within this section, is likely to constitute a weakness in any classifier built on the dataset, as the classifier will likely be better at recognizing quotes from right-wing than from left-wing parties, and from males than from females.

4.2 Quote Distribution within Parties

Figure 1 shows the distribution of policy quotes on one topic, over parties. Alternativet is the only party univocally against implementing tighter immigration policy. Enhedslisten, Radikale Venstre and Socialistisk Folkeparti are largely *against*, with approximately 80% of quotes within this class. It is worth noting that the quote distributions of both Socialdemokratiet and Liberal Alliance differ significantly from the rest of the parties within their half of the political spectrum, and to a higher degree resembles that of their political opponents. With a *for* distribution of 60%, Socialdemokratiet resides more closely to the right-wing total of 75% than the 32% of the left-wing total, and with a *for* distribution of 32% Liberal Alliance matches that of the left-wing total. However, Liberal Alliance has a lower *against* quote distribution than the left-wing total, and Socialdemokratiet has a larger *against* distribution than the right-wing total. Venstre, Det Konservative Folkeparti and Dansk Folkeparti all have a very low *against* distribution, with Dansk Folkeparti holding the smallest at just a few %.

5 Method

Pretrained fastText word embeddings of size 300 are used as word representations. These are supplemented by what will be denoted as context-based features. These consists of two sets of one-hot embeddings, representing the politicians present in the dataset and the nine parties presently in the Danish parliament respectively. For each quote, a context-based feature vector is generated with a flag raised at the index of the politician behind the quote, and the party affiliation of this politician.

5.1 LSTM Implementation

The initial approach in classifying the stances within the quote dataset was based on a recurrent LSTM-based architecture (Hochreiter and Schmidhuber, 1997), applying forget gate (f_t), input gate (i_t), cell state (C_t), output gate (o_t) and the output vector (h_t) as:

$$H = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}$$
$$f_t = \sigma(W_f \times H + b_f)$$
$$i_t = \sigma(W_i \times H + b_i)$$
$$C_t = f_t \times C_{t-1} + i_t \times tanh(W_c \times H + b_c)$$
$$o_t = \sigma(W_o \times H + b_o)$$
$$h_t = o_t \times tanh(C_t)$$

x_t denotes the input vector at time step t, and h_{t-1} denotes the output of the model at time step t-1. W denotes trainable weight matrices and b denotes trainable biases. By using an LSTM, it was sought to preserve knowledge of long-range dependencies between words, while circumventing the vanishing and exploding gradient problem (Pascanu et al., 2013).

5.1.1 Conditional Encoding

The first model implemented, denoted Conditional LSTM, applies conditionality, inspired by Augenstein et al. (2016), by initializing the LSTM layer

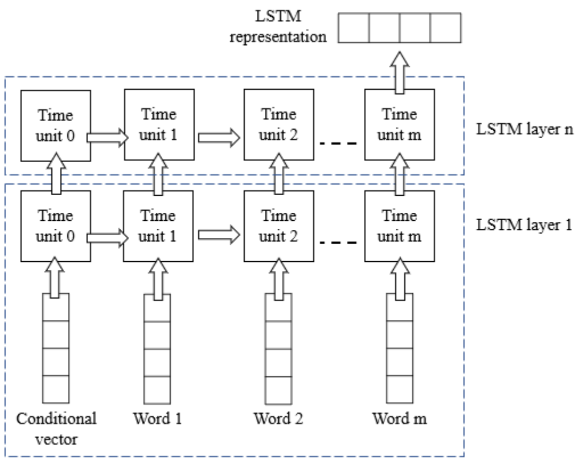

Figure 2: Diagram of Conditional LSTM layer(s)

at time step t_0 on the one-hot embedding representing the quoted politician, and the party of said politician. Thus, the model learns politician and party-dependent quote representations.

The model takes a quote as input, generated as a matrix of word embeddings of the size $E \times L$, E denoting the word embedding size, 300 for the FastText embeddings used within this paper, L denoting the length of the quote. For any value x_i in the quote embedding matrix, it is true that $x_i \in R| - 1 \leq x_i \leq 1$. At each time step, the model takes a single word embedding as input. This LSTM layer type is depicted in Figure 2.

5.2 Multi-Layered Perceptron

The second model is a simple multi-layered perceptron (Rosenblatt, 1961), denoted MLP, which applies average quote embeddings, generated as vectors where the value of the vector is the average of all word embeddings in the quote. The vector will be of length 300, when using the FastText word embeddings, and for a quote of length N, average quote embeddings are calculated as:

$$x_i = \frac{\sum_{i=0}^{N} x_{word}}{N}$$

Quote embeddings are concatenated with the one-hot representation of the quoted politician and the politician's party affiliation. See Figure 3.

5.3 Full Model Architecture

The number of deep learning layers are variable, and used as a parameter in hyperparameter search. The output of the deep learning layers are passed

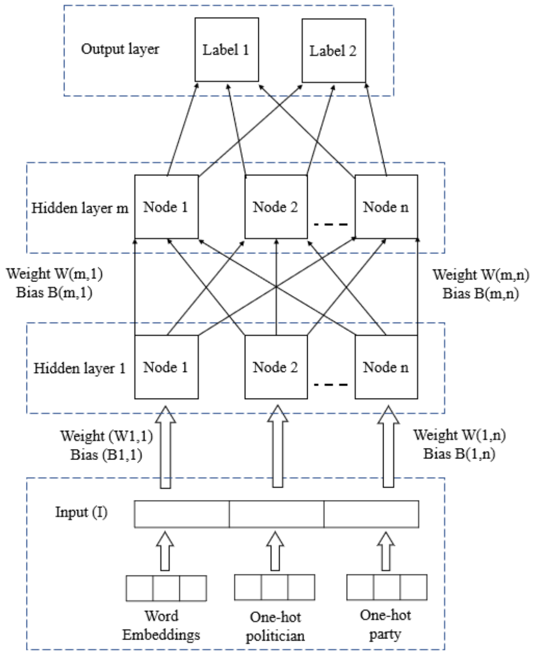

Figure 3: Diagram of Multi-layered perceptron layer(s)

to a number of linear layers containing ReLU activation functions (Richard H. R. Hahnloser and Seung, 2000), the number of which are likewise used as a parameter in hyperparameter search, followed by a softmax layer allowing for classification and optimization using categoric cross entropy loss. Both models had the number of deep learning layers and units, number of ReLU layers and units and L2 optimized, using a grid-wise search of the hyperparameter space. A learning rate of 0.001 and dropout of 0.5 is applied to both models.

6 Results and Analysis

The overall evaluation of the three models was performed with the full dataset, as well as with the *national policy* subset, using the optimal hyperparameters. No experiments were made using only the *centralization* dataset, as this was deemed too small at a quote count of just 104. The models were compared on both $F1_{micro}$ and $F1_{macro}$. However, due to the skewed label distribution within the dataset, as pointed out in Section 4, $F1_{macro}$ is the primary metric for model evaluation.

Seeing as the dataset was generated specifically for this piece of research, there exists no prior benchmarks with which to compare the models. For this reason, two benchmark models are built, namely a Gaussian Naive Bayes classifier and Random Forest classifier, both out-of-the-box implementations from the scikit-learn Python library (Pedregosa et al., 2011).

Full Dataset

	GNB	RF	Cond.	MLP
$F1_{macro}$	0.266	0.387	0.375	**0.575**
$F1_{micro}$	0.306	0.461	0.400	**0.717**
F_{acc}	0.442	0.267	0.580	**0.826**
A_{acc}	**0.6**	0.2	0.244	0.120
N_{acc}	0.029	**0.797**	0.440	**0.797**

National Policy dataset

	GNB	RF	Cond.	MLP
$F1_{macro}$	0.254	0.435	0.358	**0.585**
$F1_{micro}$	0.283	0.560	0.372	**0.774**
F_{acc}	0.337	0.525	0.256	**0.963**
A_{acc}	**0.696**	0.435	0.480	0.043
N_{acc}	0.036	0.821	0.478	**0.804**

Table 5: Performance comparison of all models, including benchmark models, using optimized hyperparameters, GNB referring to Gaussian Naïve Bayes, RF referring to Random Forest

There are three majority-based baselines. The first majority baseline-based model uses the overall majority class of the full dataset to classify quotes. The second applies the majority class for each politician to classify quotes from that politician, and the third model does the same, instead using the majority class for each party.

6.1 Model Comparison

From Table 5 it can be observed that the MLP outperforms all four other models in terms of $F1_{macro}$ on both the full and *national policy* datasets. The MLP also performs best in regards to $F1_{micro}$ on both the full dataset and the *national policy* dataset.

Table 6 show that the MLP model out-performs both majority baseline models in terms of $F1_{macro}$. However, the politician-level baseline outperforms the MLP in terms of $F1_{micro}$.

	$\mathbf{F1_{macro}}$	$\mathbf{F1_{micro}}$
MLP	**0.575**	0.717
Majority	0.253	0.611
Majority_{pol}	0.299	**0.835**
Majority_{party}	0.270	0.696

Table 6: Comparison of majority baseline performance to MLP performance

Policy dataset

	F	A	N
F	77	0	3
A	8	15	2
N	10	57	2

Full dataset

	F	A	N
F	22	52	1
A	5	16	2
N	2	52	2

Table 7: Confusion matrices for Multi-layered perceptron, using optimized hyperparameters

6.2 Misclassification Analysis

Table 7 shows the confusion matrix for the MLP, run with optimal hyperparameters on both the full and *national policy* dataset, can be found. For both datasets, the strength of the MLP is its ability to correctly classify *for* and *neutral* quotes. As a function of this, the model more or less ignores the *against* class, in pursuit of correctly classifying the two larger classes instead. A tendency can be observed towards classifying *against*-quotes as *for*, and to some extend also misclassifying some *neutral* quotes as *for*. This is not surprising, *for* being the majority class. Both the *against*-quotes and *neutral*-quotes classified as *for* are generally found to contain a large number of negative words targeting some other topic than immigration. Generally, the quotes within the *for* label apply a large number of negative words, suggesting that the classifiers mis-interpret the target of the negative words. An example of a *neutral*-quote labeled as *for* is:

> *Det er ingen hemmelighed, at vi i Dansk Folkeparti opfatter Dansk Industri som meget manipulerende og utroværdig i diskussionen om udenlandsk arbejdskraft.*
>
> *It is no secret that we in Dansk Folkeparti perceive Dansk Industri as being very manipulative and untrustworthy in the debate regarding foreign labor.*
> Martin Henriksen, (Ritzau, 2018c)

And an example of an *against* quote:

> *Det er bestemt ikke problemfrit at integrere flygtninge. Men løsningen er da ikke at eksportere problemerne til for eksempel Nordafrika, hvor man i forvejen står med en kæmpemæssig opgave.*
>
> *Integrating refugees is definitely not without its challenges. But the solution is not to export the problem to, for instance, North Africa, where the region is already faced with a huge task.*
> Johanne Schmidt Nielsen, (Ritzau, 2018e)

6.3 Post-hoc Exploratory Experiments

Additional experiments were performed using the MLP trained on the full dataset, to gain additional insight into the model's performance.

6.3.1 The Effect of Context-based Features

Comparing Table 8 and Table 5, it can be observed that removal of either party or politician from the context-based features significantly reduces the MLP's results. A model applying a feature vector composed only of the two context-based features out-performs models applying combinations of the text-based features and one of the two context-based features, but is in turn out-performed by the model applying all three features. This shows that the inclusion of context-based features significantly improves the model's performance, but that the model still relies on text-based features for optimal performance.

Feature	**$F1_{macro}$**	**$F1_{micro}$**
FastText	0.138	0.261
FastText, $Vector_{pol}$	0.405	0.522
FastText, $Vector_{party}$	0.441	0.594
$Vector_{pol}$, $Vector_{party}$	0.439	0.583
FastText, $Vector_{pol}$, $Vector_{party}$	**0.575**	**0.717**

Table 8: Results of experiments on MLP with reduced contextual features

6.3.2 Size of the Dataset

It is assumed that the small size of the quote dataset is a significant factor in preventing the models from achieving better performance, seeing as a smaller dataset size makes generalization to unobserved data points more difficult. To test this hypothesis, experiments were performed on the MLP using the optimal model hyperparameters, but a reduced training set sizes, in the range of 10 - 100% of the total quote dataset, the results of which can be found in Table 9. From this table, it is clear that decreasing the training set size reduces the performance of the model. It is assumed that the opposite is also true, and that a dataset of larger size would thus increase performance of the generated models.

6.3.3 Choice of Optimizer

The models were implemented using a simple stochastic gradient descent optimizer from PyTorch (Paszke et al., 2017). This decision was made early in the development process, prior to the search of hyperparameter spaces for models. Thus, little testing was performed for the alternative, more advanced, optimizers. To gain insight into whether the use of alternative optimizers would have improved performance, a comparative experiment was performed, the results of which are presented in Table 10.

	Quotes	Optimal epoch	F1$_{micro}$	F1$_{macro}$
10%	72	Any	0.383	0.185
20%	144	Any	0.383	0.185
30%	216	Any	0.383	0.185
40%	288	200	0.5	0.33
50%	360	300	0.478	0.33
60%	432	200	0.567	0.425
70%	504	200	0.583	0.428
80%	576	200	0.656	0.488
90%	648	300	0.727	0.52
100%	720	300	**0.717**	**0.575**

Table 9: Dataset size impact on MLP performance

	Adagrad		Adadelta		Adam	
rate ϵ	0.001	0.01	0.001	0.01	0.001	0.01
Epoch	Any	200	300	30	30	Any
F1$_{micro}$	0.383	0.633	**0.722**	0.622	0.661	0.383
F1$_{macro}$	0.185	0.490	0.518	0.536	**0.547**	0.185

Table 10: Performance of the Adagrad, Adadelta and Adam optimizers in the MLP

Learning Rate	Epoch	F1$_{macro}$	F1$_{micro}$
0.001	30	0.185	0.383
	50	0.185	0.383
	70	0.185	0.383
	100	0.348	0.506
	200	0.525	**0.733**
	300	**0.575**	0.717
0.01	30	**0.526**	**0.733**
	50	0.410	0.606
	70	0.396	0.594
	100	0.454	0.578
	200	0.478	0.633
	300	0.410	0.500
0.1	30	0.423	0.589
	50	**0.504**	**0.706**
	70	0.442	0.617
	100	0.437	0.639
	200	0.397	0.572
	300	0.497	0.667

Table 11: Results of Learning Rate experiments on Quote LSTM using optimal hyperparameters

From this table it can be seen, that the Adam optimizer reaches an $F1_{macro}$ score of 0.547, comparable to the best score of the basic SGD optimizer which was 0.575, despite hyperparameters being trained using the basic SGD optimizer. It is worth noting that this result is achieved after only 30 epochs, whereas the basic SGD optimizer required 300 epochs. This indicates that using an adaptive optimizer would not necessarily lead to higher performance than stochastic gradient descent, for this task, but can be a more efficient choice of optimizer.

6.3.4 Alternative Learning Rates

As can be seen in Table 11, a higher learning rate decreases the convergence time on high F1-scores significantly, however reducing the performance of models for higher numbers of epochs. The fact that the models applying a higher learning rate can not achieve as strong a performance as that using a learning rate of 0.001 is likely to be due to the models skipping some maxima. One solution would be a variable learning rate, reducing the learning rate once the model shows per-epoch diminishing loss reduction, thus achieving both quick convergence and precision.

7 Conclusion

This work created both a dataset and approach for political stance detection in Danish. A dataset of quotes from Danish politicians, including the quoted politician and the quoted politician's party, annotated for use in stance detection was generated, and annotation guidelines for this dataset were defined. Two deep learning-based classifiers were designed, implemented and optimized for the task. The simple MLP model that took an averaged quote embedding as input far outperformed the more advanced LSTM model, which took a single word at each time step. The generated dataset is applicable for use in future research within the field of stance detection in Danish, and the created models can be used as benchmarks when testing stance detection classifiers on this dataset.

Labeled quote data and code for this project is available on GitHub **(link)**.

Acknowledgements

We thank the reviewers, whose suggestions proved highly useful, and Victor Petrén Bach Hansen and Jakob Bang Helvind, whose sparring and feedback was invaluable during both the research behind this paper, and the writing of the paper itself. This research was conducted under the NeIC/NordForsk sponsored Nordic Language Processing Laboratory (NLPL) project.

References

Ahmet Aker, Leon Derczynski, and Kalina Bontcheva. 2017. Simple open stance classification for rumour analysis. In *Proc. RANLP*.

Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance Detection with Bidirectional Conditional Encoding. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885.

Emily M. Bender and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604.

Michael Collins and Nigel Duffy. 2001. Convolutional Kernels for Natural Language. In *Proceedings of the 14th International Conference on Neural Information Processing Systems*, pages 625–632, Vancouver, Canada.

Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours. In *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*, page 69–76, Vancouver, Canada.

Omar Enayet and Samhaa R. El-Beltagy. 2017. Determining Rumour and Veracity Support for Rumours on Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*, pages 470–474, Vancouver, Canada.

Kenneth C. Enevoldsen and Lasse Hansen. 2017. Analysing political biases in danish newspapers using sentiment analysis. *Journal of Language Works*, 2(2).

Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. 2019. SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 845–854, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political Ideology Detection Using Recursive Neural Networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 1113–1122, Baltimore, Maryland, USA.

Kristen Johnson and Dan Goldwasser. 2016. "All I know about politics is what I read in Twitter": Weakly Supervised Models for Extracting Politicians' Stances From Twitter. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2966–2977, Osaka, Japan.

Andreas Kirkedal, Barbara Plank, Leon Derczynski, and Natalie Schluter. 2019. The Lacunae of Danish Natural Language Processing. In *Proceedings of the Nordic Conference on Computational Linguistics (NODALIDA)*.

Kristoffer Kvalvik. 2017. Køn, indkomst og geografi: Her er vælgernes vigtigste dagsordener. https://www.altinget.dk/artikel/163051-koen-indkomst-og-geografi-her-er-vaelgernes-vigtigste-dagsordener.

Mirko Lai, Delia Irazu Hernandez Farias, Vivian Patti, and Paolo Rosso. 2016. Friends and Enemies of Clinton and Trump: Using Context for Detecting Stance in Political Tweets. In *Advances in Computational Intelligence: 15th Mexican International Conference on Artificial Intelligence, MICAI 2016*, pages 155–168, Cancun, Mexico.

Anders Edelbo Lillie, Emil Refsgaard Middelboe, and Leon Derczynski. 2019. Joint rumour stance and veracity. In *Proceedings of the Nordic Conference on Computational Linguistics*.

Saif M. Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiodan Zhu, and Colin Cherry. 2016. A Dataset for Detecting Stance in Tweets. In *LREC*, Portoroz, Slovenia.

Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT) - Special Issue on Argumentation in Social Media and Regular Papers*.

Finn Årup Nielsen. 2011. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, pages 93–98.

Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 2013. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on International Conference on Machine Learning*, pages 1310–1318.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. In *NIPS-W*.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. 2011. Rumor has it: identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599, Edinburgh, United Kingdom.

Misha A. Mahowald Rodney J. Douglas Richard H. R. Hahnloser, Rahul Sarpeshkar and H. Sebastian Seung. 2000. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405:947–951.

Ritzau. 2018a. Alternativet vil i EU med fyret græsk finansminister. *Ritzau*.

Ritzau. 2018b. DF siger nej til S-plan om jobcentre i Sydeuropa. *Ritzau*.

Ritzau. 2018c. OVERBLIK: Det siger politikerne om ghettoplanen. *Ritzau*.

Ritzau. 2018d. S vil skabe fem danske jobcentre i sydeuropa. *Ritzau*.

Ritzau. 2018e. Socialdemokratisk asyludspil deler rød blok. *Ritzau*.

Ritzau. 2018f. Støjberg vil have migranter fra dansk skib i land i Italien. *Ritzau*.

Ritzau. 2018g. Tyrkiske kulturformænds sympati for rabiat bevægelse alarmerer politikere. *Ritzau*.

Ritzau. 2019. Ejerskab, bestyrelse og årsregnskab. https://ritzau.com/ejerskab-bestyrelse-og-aarsregnskab/.

Frank Rosenblatt. 1961. *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Cornell Aeronautical Laboratory, Buffallo, New York, USA.

Maria Skeppstedt, Vasiliki Simaki, Carita Paradis, and Andreas Kerren. 2017. Detection of stance and sentiment modifiers in political blogs. In *Proceedings of the International Conference on Speech and Computer, SPECOM*, pages 1589–1599, Hatfield, United Kingdom.

Martin Tutek, Ivan Sekulic, Paula Gombar, Ivan Paljak, Filip Culinovic, Filip Boltuzic, Mladen Karan, Domagoj Alagic, and Jan Snajder. 2016. TakeLab at SemEval-2016 Task 6: Stance Classification in Tweets Using a Genetic Algorithm Based Ensemble. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 464–468.

Wan Wei, Xiao Zhang, Xuqin Liu, Wei Chen, and Tengijao Wang. 2016. pkudblab at SemEval-2016 Task 6 : A Specific Convolutional Neural Network System for Effective Stance Detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 384–388.

Guido Zarrella and Amy Marsh. 2016. MITRE at SemEval-2016 Task 6: Transfer Learning for Stance Detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 458–463.

Li Zeng, Kate Starbird, and Emma S. Spiro. 2016. Unconfirmed: Classifying Rumor Stance in Crisis-Related Social Media Messages. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM)*, pages 747–750, Cologne, Germany.

Appendix 1: Data Statement

Curation rationale Quotes from Danish politicians published by the Ritzau news agency.

Language variety BCP-47: da-DK

Speaker demographic

- Danish politicians.
- Age: approx. 25-70.
- Gender: mixed; see Section 3.1.
- Race/ethnicity: mostly white with Scandinavian background.
- Native language: Danish.
- Socioeconomic status: minimum 56494.17 DKK per month ($8470 USD).
- Different speakers represented: 63.
- Presence of disordered speech: Quotes are mostly curated, so not prevalent.

Annotator demographic

- Age: 20-30.
- Gender: male.
- Race/ethnicity: white northern European.
- Native language: Danish.
- Socioeconomic status: higher education student.

Speech situation Quotes given by politicians in parliament during debate or discussion, during verbal interviews or in writing, transcribed and then published in edited newswire.

Text characteristics Danish Newswire.

Provenance Originally taken from Ritzau.