Predicting political parties based on the language of their socal media activities

Marc Weitz, Manuela Weigold
Eberhard Karls Universität Tübingen

In this article we investigated if political parties can be distinguished by the language of their social media activities. We used a simple two-layer neuronal network to predict political parties by their social media posts. The neuronal network was trained on the facebook posts of eight German parties with a simple Naive Discriminative Learning rule. A cross-validation analysis revealed good accuracy of the predictions for all political parties and additionally showed that the accuracies for the different parties are not homogeneous. A post-hoc analysis of the most activating cues revealed that the best predictors for each party are the names of famous politicians. Furthermore, the post-hoc analysis of the vector semantics of the model showed different patterns of the activation of the cues for two parties.

Keywords: two-layer network, Naive Discriminative Learning, Rescorla-Wagner

Introduction

Uniqueness and discriminability are important in the competition of political parties in their attempt to attract more votes. But people tend to think that political parties become more and more indistinguishable over the last decades.
It is assumable that language plays an important role in the discriminability of political parties. In the last decades, several models of language processing have been developed to analyse language. Recently, a simple model based on the Naive Discriminative Learning rule was presented (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011).
Naive Discriminative Learning is based on single event learning using the Resorla-Wagner equation (Rescorla & Wagner, 1972):

$$V_{CS(n+1)} = V_{CS(n)} + \Delta V_{CS(n)} \tag{1}$$

Where $V_{CS(n)}$ denotes the association strength of the conditioned stimulus in trial n and $\Delta V_{CS(n)}$ denotes the change in association strength from trial to trial:

$$\Delta V_{CS(n)} = \alpha_{CS} \cdot \beta_{US(n)} \cdot (\lambda_{US(n)} - V_{all(n)}) \tag{2}$$

The product of α_{CS}, which represents the salience of the conditioned stimulus (CS), and $\beta_{US(n)}$, which represents the association value of the unconditioned stimulus (US) in trial n, represents the learning rate of the association. $\lambda_{US(n)}$ is the maximal association strength of the US in trial n and $V_{all(n)}$ is the sum of association strengths of all presented stimuli.
The Naive Discriminative Learning approach additionally takes into account that the calculation of ΔV depends on the different cases of combinations of the cues C and the outcomes O (Baayen et al., 2011):

$$\Delta V_i = \begin{cases} 0 & \text{if } C_i \text{ wasn't present} \\ \alpha \cdot \beta \cdot (\lambda - \sum_{\text{Present}(C_j,t)} V_j) & \text{if } O \text{ was present} \\ \alpha \cdot \beta \cdot (0 - \sum_{\text{Present}(C_j,t)} V_j) & \text{if } O \text{ was not present} \end{cases} \tag{3}$$

We used this approach for language processing to examine data we retrieved from public social media accounts of different political parties and to test our hypothesis whether the language used for social media activities differs for different parties.

Methods

The material was retrieved from the public facebook accounts of eight German parties (AfD, Bündnis 90/Die Grünen, CDU, CSU, die Linke, FDP, NPD and SPD) using FacePager (Keyling & Jünger, 2013), a tool for retrieving publicly available information from several JSON-based APIs. The data was retrieved on 9th October 2016.
We preprocessed the data to disjunct events based on a facebook post each using R (R Core Team, 2016). Each event consists of a set of cues, a set of outcomes and the frequency of occurrence. In our analysis we used the normalized words (punctuation and capitalisation was removed using Python (Python Software Foundation, 2016)) of the facebook post as cues to predict the authorship as our outcome. Based on practical reasons we assumed that each post was only published once for each party, therefore we set the frequency for each event to one.
To avoid position effects we randomized the order of events. The training of the model was simulated using the R package ndl2 (Arppe et al., 2015). The analysis of the model was done with R (R Core Team, 2016).

Table 1
Frequencies of the predicted outcomes (rows) compared to the correct outcome (columns)

	Outcomes							
Predicted Outcome	afd	cdu	csu	fdp	gruene	linke	npd	spd
afd	2794	32	70	35	74	90	316	138
cdu	129	5301	193	96	137	116	196	310
csu	108	120	4606	93	227	190	288	309
fdp	45	36	63	3536	83	40	59	119
gruene	134	101	217	94	3695	251	236	292
linke	86	36	64	38	62	2881	174	102
npd	262	83	166	93	129	148	6412	163
spd	414	461	709	531	835	712	655	12113

Results

The model was analysed using a Leave-one-out cross-validation (LOOCV) with a chunksize of 1000 events. Therefore we have split our dataset leaving 1000 events out each. On each of these subsets we trained a model and let it predict the outcome of the missing 1000 events by calculating the activations for each event. The highest activation yields the most likely party based on the knowledge of the model. Table 1 shows the frequencies of the the predicted outcome depending on the true outcome.
To analyse the accuracy of the model we pooled the data by whether the prediction was correct ("hit") or false ("miss") for all parties. Based on these frequencies we calculated the accuracy of predicting each party correctly by dividing the frequency of getting predicted correctly by the overall frequency the party was predicted. The accuracies for all parties are shown in Figure 1.
A χ^2-test revealed a significant inhomogeneity of prediction correctness between the different parties ($\chi^2(7) = 901.11, p < .001$).

Discussion

It is common belief that political parties have become more and more indistinguishable. But can empirical data confirm this theory?
If the common opinion is correct, a model based on any kind of discriminative learning would be forced to fail. In this case the prediction accuracy would be chance level because the model had no possibility to learn anything. That is the case because its learning is based on the discriminability of the outcome, here the political parties.
However, our model yields good prediction accuracy for all political parties far above chance level as shown in Figure 1. Even the prediction accuracy is not homogeneous for all parties. These findings lead us to the assumption that political parties are distinguishable at least for the aspect of their language.

On the other hand it must be questioned which words cause the distinction. Therefore the most strongly predicting words for all eight parties are shown in Table 2 and Table 3. Taking a look at these tables reveals that a large amount of words in the Top 30 predicting words are names of famous

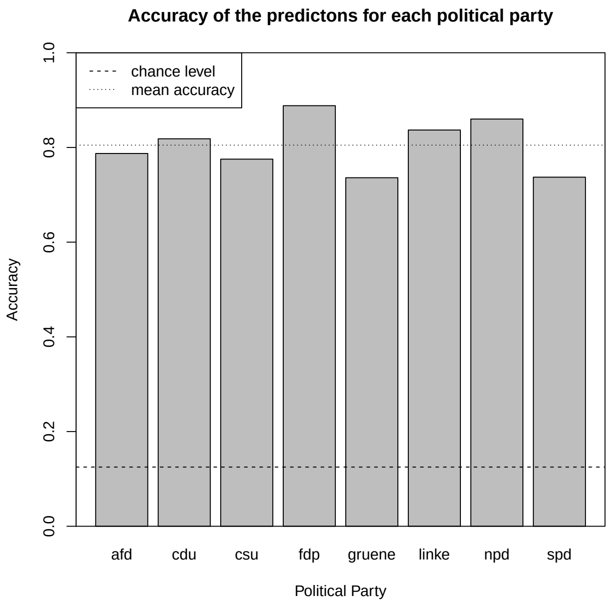
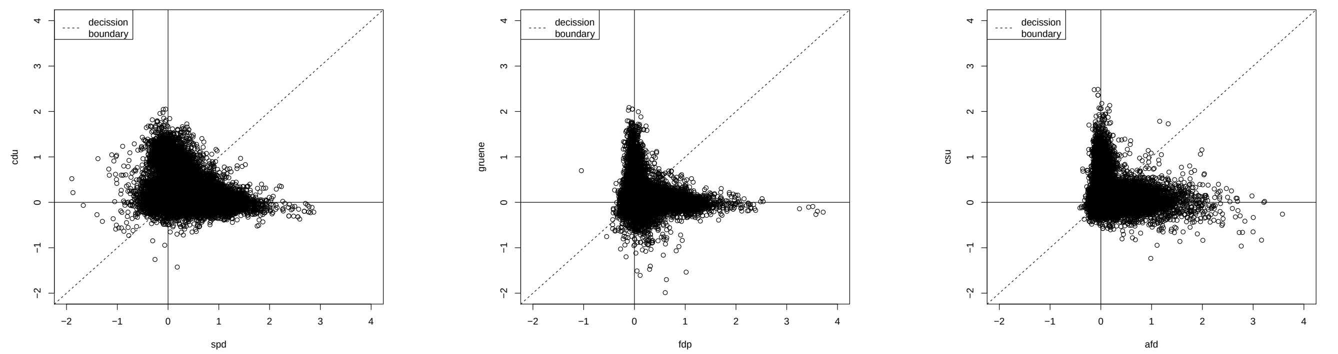


Figure 1. Accuracy of the predictions for each party with average accuracy and chance-level



(a) SPD vs CDU (b) FDP vs Buendnis 90/Die Gruene (c) AFD vs CSU

Figure 2. Scatteplots for the activations for two parties each.

politicians of each party. It would be interesting to run the model again without these cues and have a look at how much the prediction accuracy will drop.

Another interesting aspect to take a look at are the semantic vectors of the model. From these vectors we can gain insights into more specific features like how the activation of cues between two outcomes is. Three examples are shown in Figure 2. Assuming the parties are indistinguishable all points were positioned on a line with an intercept of zero and a slope of 1 because each cue would activate both outcomes identically. As it can be seen in Figure 2 there is a large number of cues with an activation around zero. These cues are assumably common words giving nearly no information of the authorship. On the opposite, cues can provide a large amount of information for one outcome but not for any of the others, indicated by a straight line in horizontal or vertical direction. For example Figure 2b shows how a bulk of cues activate Buendnis 90/Die Gruene but not the FDP (vertical line) and vice versa (horizontal line). However, Figure 2a does not show a clear pattern as found in Figure 2b which indicates that some pairs are better distinguishable than others. A slightly different pattern is shown in Figure 2c. In this plot is a bulk of cues presented which indicate AFD more strongly than the CSU. On the other hand, the cues indicating the CSU are activating the AFD homogeneously (they appear to activate the other party independently of how much they indicate the CSU).

Besides these insights our results demonstrate how simple computational techniques like two-layer neuronal networks in combination with the Naive Discriminative Learning approach can solve even complex questions like predicting a subject by its behaviour. Therefore it should be noticed that the model we used only used the information provided within the posts and was trained on very sparse data. Additional information like syntax or word frequency were not used, but will likely provide additional information for discriminating the authorship.

However, it is very interesting how far we can go with these parsimonious models in comparison to more expensive models like deep neuronal networks, which are very common today. An interesting question remaining is how accurate this model's predictions are when the entropy is high, e.g. there are many possible outcomes to discriminate. Therefore it could be interesting to train the model on a large number of facebook authors from different areas. Furthermore, this attempt could yield sets of domain specific words which are good predictors for a specific domain.

References

Arppe, A., Bitschau, S., Schilling, N., Hendrix, P., Milin, P., Baayen, R. H., & Shaoul, C. (2015). *Ndl2: naive discriminative learning*. R package version 0.1.0.9002.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological review*, *118*(3), 438.

Keyling, T. & Jünger, J. (2013). Facepager (version, f.e. 3.3). an application for generic data retrieval through apis. Source code available from https://github.com/strohne/Facepager.

Python Software Foundation. (2016). *Python 3 (version 3.4.2)*. Retrieved from https://www.python.org/

R Core Team. (2016). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rescorla, R. A. & Wagner, A. R. (1972). A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning: current research and theory*.

Table 2
Best predicting words for the parties AFD, CSU, Buendnis 90/Die Gruene and NPD

	Political Parties			
	afd	csu	gruene	npd
1	lucke	bayern	grüne	info
2	afd	dobrindt	gruene	npd
3	hr	csu	grün	franz
4	sprecher	seehofer	tage	asylanten
5	programmhinweis	bayerns	grünen	volk
6	kollegen	bär	cem	udo
7	gauland	scheuer	greenprimary	überfremdung
8	veränderung	bayerische	özdemir	asyl
9	petry	schönes	claudia	frank
10	frauke	city	teilt	asylbetrüger
11	alternative	wünschen	atomkraft	voigt
12	freitagsgedanken	andreas	roth	heimat
13	link	betreuungsgeld	rebecca	souveränität
14	henkel	guttenberg	bdk	jn
15	bernd	aschermittwoch	harms	nationale
16	freunde	bayerischen	trittin	landesverband
17	demo	söder	eckardt	deutsche
18	nun	the	göring	islamisierung
19	wahlen	edmund	atomausstieg	mir
20	pretzell	dorothee	vs	gefällt
21	landesparteitag	bayernkurier	jürgen	teilen
22	positionen	türkei	katrin	abschieben
23	griechenland	aigner	wach	stoppen
24	bezahlen	stoiber	bild	asylbetrug
25	these	weber	atom	com
26	möglichkeit	br	lemke	muß
27	zahlen	ilse	demonstrieren	stein
28	deren	lounge	länderrat	ausländer
29	erfurt	herrmann	steffi	ds
30	prof	manfred	bundesdelegiertenkonferenz	klick

Table 3
Best predicting words for the parties CDU, FDP, Die Linke and SPD

	Political Parties			
	cdu	fdp	linke	spd
1	cdu	lindner	linke	spd
2	tauber	fdp	linken	stimmefuervernunft
3	gröhe	lambsdorff	gysi	©
4	flugblatt	graf	gregor	finale
5	medianight	kubicki	höhn	butzmann
6	angela	nicola	riexinger	dominik
7	bundesfinanzminister	beer	kipping	martin
8	leyen	wissing	golze	sigmar
9	bundesumweltminister	christian	diana	peer
10	kauder	brüderle	matthias	schulz
11	wahlfakten	suding	drin	walter
12	röttgen	rainer	bernd	faktencheck
13	merkel	demokraten	ernst	plambeck
14	ursula	liberale	gabi	anschauen
15	schavan	aktionswoche	zimmer	schluss
16	schäuble	wolfgang	kandidatur	willy
17	arbeitsmarkt	alexander	bartsch	scholz
18	maizière	volker	lafontaine	motiv
19	david	buchstabe	liebknecht	digitalen
20	bundesinnenminister	lexikon	lötzsch	mindestlohn
21	interview	katja	ramelow	manuela
22	annette	freien	liste	schmidt
23	bundesarbeitsministerin	julis	katja	andrea
24	mcallister	uwe	ceta	hendricks
25	adenauer	lencke	gesine	nahles
26	artikel	steiner	syriza	stephan
27	peter	dürr	bodo	stream
28	magazin	solms	katharina	yasmin
29	wachstum	barth	parteivorstand	glückwunsch
30	leistungen	diener	klaus	schwarzgelblog