

Voorspellen van partij affiliatie in de Handelingen van de
Tweede Kamer
Onderzoeksvoorstel

Jasper van der Heide
10732721

5 april 2018

1 Inleiding

Teksten van politieke partijen kunnen bruikbaar zijn voor het bepalen van ideologische positie van andere teksten, aangezien zij zowel een tekst leveren als ook een bekende ideologie. Deze informatie kan vervolgens toegepast worden bij andere teksten die wellicht ideologisch van aard zijn. Bijvoorbeeld, op basis van deze informatie kan men teksten uit kranten classificeren op basis van ideologie. Diermeier et al. (2007) deed onderzoek naar het classificeren op basis van ideologische positie in de Amerikaanse Senaat (101e tot en met 108e Congres). Dit onderzoek wist de ideologie van de senatoren te voorspellen met een 92 procent nauwkeurigheid.

Als een vervolg op dit onderzoek deed Graeme Hirst et al. (2010) een vergelijkbaar onderzoek naar zowel het Canadese als het Europese Parlement. In dit onderzoek maken zij gebruik van support-vector machines. In tegenstelling tot het onderzoek van Diermeier et al. (2007), vinden zij minder dat de woorden van de sprekers een uiting zijn van ideologie. Daarentegen vinden zij wel een grotere invloed van oppositie tegenover regering in de woorden van de sprekers.

2 Beschrijving van het onderwerp

In dit onderzoek wordt getracht de lessen van bovengenoemde onderzoeken toe te passen op de Nederlandse politiek en te kijken in hoeverre er een voorspelbaarheid is in partij affiliatie van Tweede Kamerleden aan de hand van hun speeches. Het voornaamste doel is hierbij een zo accuraat mogelijke voorspelling te maken van het partijlidmaatschap van een spreker. Maar daarnaast van grote relevantie is kijken naar welke factoren zeer voorspelbaar zijn en wat dit kan betekenen.

3 Onderzoeksvraag

In hoeverre is partij affiliatie van een spreker in de Tweede Kamer te voorspellen op basis van de speeches in de Tweede Kamer?

4 Onderzoeksmethode

Het onderzoek zal zich in eerste instantie richten op het repliceren van het onderzoek van Graeme Hirst et al. (2010), maar dan met data van het Nederlandse parlement, de Tweede Kamer. Vervolgens zal gekeken worden naar het verbeteren van de classificatie door variaties op het onderzoek van Graeme Hirst et al. (2010) toe te passen. Voor de hand liggende variaties zijn:

- Verschillende wegingen voor woord (tf, tf-idf, etc.)
- Verschillende woorden buiten beschouwing laten bij classificatie (stopwoorden, 500 meest voorkomend, etc.)

- Verschillende classificatie methoden (Support Vector Machine, Naive Bayes Classifier, etc.)

Daarnaast wordt in het onderzoek van Graeme Hirst et al. (2010) al opgemerkt dat er sprake kan zijn van confoundings. In hun geval valt voornamelijk de tegenstelling van oppositie en regering op. In dit onderzoek zal getracht worden meerdere van deze confoundings te vinden of uit te sluiten. Waar mogelijk kan dit dan meegenomen worden in het vinden van een zo optimaal mogelijke classificatie.

4.1 De data

De data die gebruikt zullen worden voor het onderzoek zijn de Handelingen van de Tweede Kamer en bijbehorende metadata zoals die te vinden zijn op de website van de Nederlandse overheid, specifiek de website officielebekendmakingen.nl. Deze Handelingen zijn beschikbaar vanaf het vergaderjaar 1994/1995 tot en met 2018. Welke en hoeveel Handelingen precies gebruikt worden, hangt af van praktische haalbaarheid, maar het ligt voor de hand om over meerdere kabinetsperioden te nemen, aangezien dit mogelijk maakt om de invloed van regering tegenover oppositie te onderzoeken en hiermee rekening te houden in de voorspelbaarheid.

Eventueel kan ook gebruik gemaakt worden van de reeds verzamelde Handelingen door PoliticalMapshup, die ruim 200 jaar terug gaan.

Deze data zullen dan verwerkt worden en mogelijk bewerkt worden. Hierbij kan gedacht worden aan Part-Of-Speech, lemmatisation en/of stemming.

Voorgaande onderzoeken maakten gebruik van Support Vector Machines met bag-of-words features. Dit is daarom ook de meeste logische machine learning methode om mee te beginnen, maar afhankelijk van dit resultaat en de tijd, is het ook mogelijk en wellicht interessant om te kijken naar andere methoden.

5 Planning

Week 14	Voltooien Inleiding en onderzoeksvoorstel
Week 15	Voltooien reproductie onderzoek Graeme Hirst et al. (2010)
Week 16	Voltooien stuk over classificatiemethoden
Week 17	Evaluatie met begeleider, inleveren inleiding, verkennen confounding biases
Week 18	Voltooien onderzoek naar confounding biases
Week 19	Inleveren methode & theoretische achtergrond
Week 20	Uitwerken resultaten
Week 21	Voltooien resultaten
Week 22	Discussie en Conclusie voltooien
Week 23	Eerste versie inleveren
Week 24	Feedback verwerken
Week 25	Feedback verwerken & inleveren definitieve versie
Week 26	Vorbereiden presentatie
Week 27	Presentatie scriptie

In principe elke week afspreken met begeleider kwartier afspreken als dat nodig is.

6 Naam en akkoord van begeleider

Maarten Marx

7 Literatuur

Diermeier, D. J-F. Godbout, B. Yu and S. Kaufmann. 2007. Language and ideology in Congress. Annual Meeting of the Midwest Political Science Association.

Hirst, Graeme Riabinin, Yaroslav Graham, Jory Boizot-Roche, Magali Morris, Colin. (2014). Text to Ideology or Text to Party Status?. 93-116. 10.1075/dapsac.55.05hir.