

Research Report

İsa Ali Demir

CSSM 502

Introduction

What do Kurdish leaders understand from democracy? Which aspects of democracy are highlighted more in the public speeches of Kurdish party leaders? Are regularly held free and fair national elections discussed more than freedom of speech as a central prerequisite for democracy? Or is being respectful and protective towards minority rights put forward more than protecting property rights as a key component to be considered as a democracy? In my research for the course, I deal with these questions. I examine what Kurdish leaders understand from democracy by employing an unsupervised machine learning text analysis model, namely Latent Dirichlet Allocation (LDA), to weekly group speeches of Kurdish parties between the years of 2010 and 2020. Python packages of *NLTK*, *stop_words*, and *gensim* are used in the process of tokenization, stemming, constructing the document-term matrix, and finally building the model.

Parties that have a group in the Grand National Assembly of Turkey normally hold weekly group meetings during their legislative periods. Party leaders make speeches at these meetings, deputies, party members and various guests attend these speeches, and speeches can be broadcast live on television. In these speeches, the leaders touch on the issues on the political agenda and give messages to their own party members, other parties, and citizens. Focusing on these weekly group speeches in order to understand the Kurdish political party leaders' perceptions of democracy is advantageous in several respects. Since parliamentary group talks are held every week and are not tied to a specific political process (eg elections),

leaders are quite free about their content. For this reason, determining whether leaders prefer to talk about democracy, and if they do, which aspect of democracy they emphasize and what political phenomena they associate with (eg elections, freedoms, the rule of law, etc.) will give important clues about their perception of democracy. On the other hand, these speeches are not only for party members or deputies, but they are also valuable resources for shaping and understanding the prevailing understanding of democracy in the country, as they are messages to other political actors and citizens.

Literature Review

LDA is an unsupervised machine learning technique that can scan a range of documents, detect word and phrase patterns within, and automatically cluster the phrases and similar expressions that best characterize a group of documents. This technique has been termed an “unsupervised” technique because it does not require a predefined label or list of training data classified by humans. An LDA model detects word frequency and distance between words, collecting similar texts and the words and phrases that appear most often together. These words and phrases that appear together repeatedly in a particular set of documents are called "topics". With this information, it is possible to determine the content of each text group. For example, political texts may come from different political parties, and LDA makes it possible to aggregate documents for different political parties by subject (ie, phrases that often go together). LDA and other topic modeling methods have been used in the political science literature on how legislators in the US Senate present senate speeches (Quinn et al., 2010) and press releases (Grimmer, 2010), open-ended survey responses (Roberts et al., 2014), and how international newspapers present news (Roberts et. al., 2016) and Twitter messages (Lucas et al., 2015). Also, Elçi (2019) has shown that it is possible to convert the texts of these speeches

into data by compiling 569 parliamentary group speeches made by political party leaders between 2011 and 2019.

Methodology

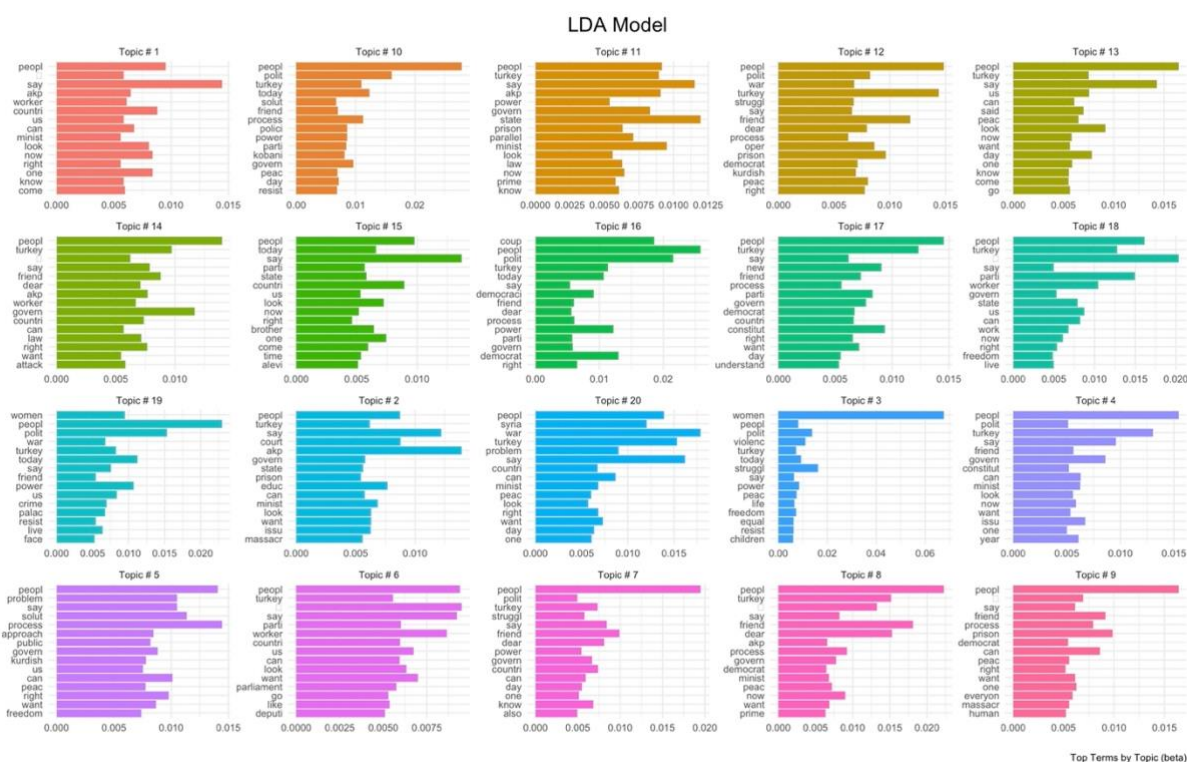
The Latent Dirichlet Allocation (LDA) model was applied to a corpus of speeches made by Kurdish party leaders in Turkey. The speeches were collected from various sources and pre-processed to remove irrelevant information and prepare the text for analysis. The pre-processing steps included tokenization, lemmatization, and removal of stop words. The LDA model was implemented using the Gensim library in Python. The number of topics was set to 20 based on the size of the corpus and the desired level of granularity in the analysis. The model was trained for a total of 100 iterations, with a random seed set for reproducibility.

The results of the LDA model were analyzed to identify the key themes and keywords in the speeches. The model generated a set of topics, each represented by a set of keywords, and the speeches were assigned to one or more topics based on the keywords they contained. The most frequent words were also identified. It's worth noting that LDA is an unsupervised model, and the labels of the topics are not assigned by the model but by the user, so it's important to check if the labels assigned to the topics are coherent with the content of the speeches.

In order to increase the reliability of the results, it would be beneficial to perform a manual validation of the topics generated by the LDA model. This can be done by reviewing a sample of the speeches assigned to each topic and comparing them to the keywords associated with that topic. Additionally, it would be useful to apply other natural language processing techniques, such as sentiment analysis, to gain further insights into the speeches.

It's also important to keep in mind that LDA is a generative probabilistic model, and its results may depend on the initial conditions, such as the random seed, the number of iterations, the number of topics, etc. Therefore, it's important to run the model several times with different parameters to check if the results are consistent.

Results / Discussion



The Latent Dirichlet Allocation (LDA) model applied to speeches of Kurdish party leaders in Turkey revealed several key themes that are prevalent in the speeches. These themes include issues of peace, democracy, human rights, and self-governance for the Kurdish community.

Topic 12, for example, seems to focus on issues of peace and democracy, with keywords such as "peace," "democrat," "Kurdish," and "prison" being prominent. This suggests that the speakers may have been discussing the importance of peace and democracy for the Kurdish community in Turkey, and possibly addressing the issue of Kurdish individuals being held in

prison. The Kurdish community in Turkey has long been marginalized and denied basic rights and freedoms, and it seems that the party leaders are advocating for peace and democracy as a means to rectify this situation.

Topic 9 also appears to focus on the theme of peace, with keywords such as "process," "massacre," "human," and "peace" being prominent. This suggests that the speakers may have been discussing ongoing peace processes and the importance of human rights in achieving peace. The Kurdish community in Turkey has been subject to violence and human rights abuses for decades, and it seems that the party leaders are calling for peace processes that prioritize the protection of human rights.

Topic 3 centers on issues of women's rights, with keywords such as "women," "violence," "freedom," and "equal" being prominent. This suggests that the speakers may have been discussing issues of violence against women and the importance of equal rights for women. The Kurdish community in Turkey has a long history of patriarchal attitudes and discrimination against women, and it seems that the party leaders are advocating for the rights and equality of Kurdish women.

Topic 20 seems to focus on issues related to Syria and Turkey's relationship to the war, with keywords such as "Syria," "war," "problem," "minister," and "Turkey" being prominent. This suggests that the speakers may have been discussing the impact of the Syrian war on Turkey and the role of the Turkish government in addressing the problem. The Kurdish community in Turkey has been deeply affected by the Syrian war, and it seems that the party leaders are calling for the government to take action to address the problems caused by the war.

Topic 5 centers on issues of governance and the rights of the Kurdish population, with keywords such as "process," "govern," "Kurdish," and "right" being prominent. This suggests that the speakers may have been discussing ongoing political processes and the importance of ensuring that the Kurdish population has the right to self-governance. The Kurdish community in Turkey has been denied the right to self-governance for decades, and it seems that the party leaders are advocating for political processes that would allow the Kurdish population to govern themselves.

In general, the results of the LDA model suggest that the Kurdish party leaders in Turkey are primarily focused on issues of peace, democracy, human rights, and self-governance for the Kurdish community. The most frequently used words such as "people", "right", "Turkey", "Polit", "Peace", "Democrat", "Democracy", "Process" indicate that the speeches have a strong emphasis on the rights of the people, the importance of democracy, and the need for peaceful and political process to solve the problems. The Kurdish community in Turkey has long been denied basic rights and freedoms, and it seems that the party leaders are advocating for peace, democracy, human rights, and self-governance as a means to rectify this situation.

References

Barberá, P., Boydston, A. E., Linn, S., McMahon, R., & Nagler, J. (2020). Automated Text Classification of News Articles: A Practical Guide. *Political Analysis*, 29(1), 19–42.
<https://doi.org/10.1017/pan.2020.8>

Elçi, E. (2019). The Rise of Populism in Turkey: A Content Analysis. *Southeast European and Black Sea Studies*, 19(3), 387–408. <https://doi.org/10.1080/14683857.2019.1656875>

Grimmer, J. (2010). A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis*, 18(1), 1–35.
<https://doi.org/10.1093/pan/mpp034>

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2015). Computer-Assisted Text Analysis for Comparative Politics. *Political Analysis*, 23(2), 254–277. <https://doi.org/10.1093/pan/mpu019>

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespín, M. H., & Radev, D. R. (2010). How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science*, 54(1), 209–228. <https://doi.org/10.1111/j.1540-5907.2009.00427.x>