

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual labeling assisted by associated documents

Topic labeling parameters

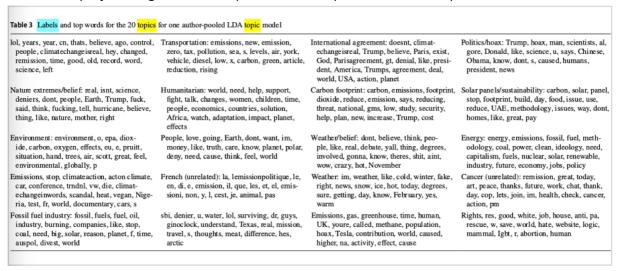
Nr of inspected words: 50

Label generation

The labels were chosen by looking at the top 50 words per topics with their probabilities and examining the tweets that are most likely per topic.

Some cells have the same topic; tweets that fall into either topic should be grouped together in the analysis. Cells with incomprehensible topics are not labeled.

Table 3 displays one good author-pooled LDA topic model with topic labels.



The relevant topics that could be labeled are transportation, international agreement, politicshoax, nature extremes (hurricane) belief, humanitarian, carbon footprint, solar panelssustainability, environment, weather belief, energy, weather, and fossil fuel industry.

Motivation

One sign of a good topic model is when the topics can be given a label that corresponds to the top words in that topic.

And

Many of the topics in Table 3 are not related to climate change. This shows that topic modeling (with labeling) can be done to filter out irrelevant documents of tweets that may have been detected by the keyword search but have nothing to do with climate change.

Topic modeling

LDA

Topic modeling parameters

Nr of topics (k): {5, 20, 80}

a: 1/k

η: 1/k

Nr. of topics

20 (Many different topic models with random seeds, each generating 20 topics)

Label

Manually assigned single or multi word label

Label selection

\

Label quality evaluation

\

Assessors

١

Domain

Paper: Climate change

Dataset: Social media (Twitter)

Problem statement

In this study, a large dataset of geotagged tweets containing certain keywords relating to climate change is analyzed using volume analysis and text mining techniques such as topic modeling and sentiment analysis.

Latent Dirichlet allocation was applied for topic modeling to infer the different topics of discussion, and Valence Aware Dictionary and sEntiment Reasoner was applied for sentiment analysis to determine the overall feelings and attitudes found in the dataset. These techniques are used to compare and contrast the nature of climate change discussion between different countries and over time. Sentiment analysis shows that the

overall discussion is negative, especially when users are reacting to political or extreme weather events. Topic modeling shows that the different topics of discussion on climate change are diverse, but some topics are more prevalent than others. In particular, the discussion of climate change in the USA is less focused on policy-related topics than other countries.

Corpus

Origin: Twitter

Nr. of documents: 390,016 (366,244 after pre-processing)

Details:

tweets from July 1, 2016, to February 28, 2018

Keywords: climate change, carbon dioxide, fossil fuel, carbon footprint, emissions

Document

Tweet message, geographic data, id, userid, postdate

Postdate	Latitude	Longitude	Message	Country	Region	City
2017-01-30 02:19:21	-41.2529983521	174.754333496	What is the future for #NZ forests? @ sciblogsnz: "Forest health in a chang- ing climate" https://t.co/X8oE9kjMPi #climatechange #environment	NZ	Wellington	Kelburn
2017-10-29 11:35:54	35.30904769900000	4 – 98.7169952393	@DaveLeeC3 @JacobAWohl @real- DonaldTrump So are we giving Obama credit for the hurricanes too? What about Global Warming?	USA	Oklahoma	Weatherfor
2017-12-05 08:33:26	60.18334198	24.6751346588	.@miapetrakumpula: Climate change is real and we need European action and commitment to change the future. #FutureofEurope	FI	Uusimaa	Espoo

Pre-processing

- manually examining a small sample of the dataset to see which kinds of false positives
 exist and then removing many false positives of that type from the entire dataset
 programmatically.
- Reverse geocoding: process of obtaining a readable place name or address from a latitude/longitude point. This adds:
 - country name, region name (if the city belonged to some political region), and city name.
- Removal of tweets from bot or spam users
- hyperlinks and non-alphabetic characters such as punctuation and numbers are removed

- Tokenisation
- stopword removal
- frequent keyword removal (keywords "climate," "change," "climatechange," "global," "warming," and "globalwarming")
- words that rarely occur (less than 200 times) in the corpus are removed

abstract = {Social media websites can be used as a data source for mining public opinion on a variety of subjects including climate change. Twitter, in particular, allows for the evaluation of public opinion across both time and space because geotagged tweets include timestamps and geographic coordinates (latitude/longitude). In this study, a large dataset of geotagged tweets containing certain keywords relating to climate change is analyzed using volume analysis and text mining techniques such as topic modeling and sentiment analysis. Latent Dirichlet allocation was applied for topic modeling to infer the different topics of discussion, and Valence Aware Dictionary and sEntiment Reasoner was applied for sentiment analysis to determine the overall feelings and attitudes found in the dataset. These techniques are used to compare and contrast the nature of climate change discussion between different countries and over time. Sentiment analysis shows that the overall discussion is negative, especially when users are reacting to political or extreme weather events. Topic modeling shows that the different topics of discussion on climate change are diverse, but some topics are more prevalent than others. In particular, the discussion of climate change in the USA is less focused on policy-related topics than other countries.},

```
author = {Dahal, Biraj and Kumar, Sathish A. P. and Li, Zhenlong},
date-added = {2023-04-11 11:33:34 +0200},
date-modified = {2023-04-11 11:33:34 +0200},
day = {10},
doi = {10.1007/s13278-019-0568-8},
issn = {1869-5469},
journal = {Social Network Analysis and Mining},
month = {Jun},
number = {1},
pages = {24},
```

```
title = {Topic modeling and sentiment analysis of global climate change
tweets},
url = {https://link.springer.com/content/pdf/10.1007/s13278-019-0568-8.pdf},
volume = {9},
year = {2019},
bdsk-url-1 = {https://link.springer.com/content/pdf/10.1007/
s13278-019-0568-8.pdf},
bdsk-url-2 = {https://doi.org/10.1007/s13278-019-0568-8}}
```

#Thesis/Papers/BS