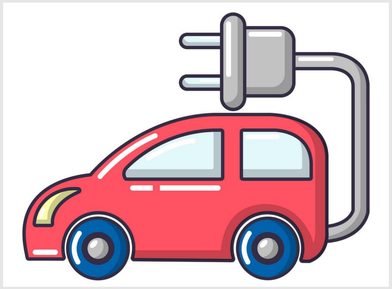


Measuring the Political Motivation behind French Government's Endorsement for Electric Vehicles

Team Project by Hu, Xiaoyan for the *Course: Python Programming for Data Scientists, Hertie School*

Background

In 2013, Bradley W. Lane and others identified two motivations for why governments seek to promote the electric car: risk management in response to ecological or energy concern versus industrial policy seeking an economic upgrade. He believed the French government lied in between these two positions.

Research Motivation

1. Update Lane's 2013 conclusion on French government's political motivation;
2. Explore an alternative position articulation method other than the conventional human text reading approach.

Project Implementation and Results

Web Scraping

for getting the right text

Method:
This project set up a web spider with the Scrapy library to crawl the article content following each url link yield by keyword ("véhicules électrique") search results from the French government website and export it to an csv file

Experiment:
scraped 159 items from the French government website and exported it to a CSV file sized 302141 bytes.

In total, the process took 78.82 seconds.

Analysis:
Although the spider performance was satisfactory, the scraped text were not clean enough for direct preprocessing and therefore, manual filtering was conductedto delete senseless phrase and some irrelavant content.

Topic Modeling

to discover the hidden relations

Method:
Latent Dirichlet Allocation (LDA), an unsupervised, probabilistic modelling method which extracts topics from a collection of corpus. The coding implementation relied mainly on gensim library in python.

Experiment:
1. Library import (some basic ones including numpy, pandas, os and pprint)
2. Data cleaning (rely mainly on spacy, nltk, and re)
3. Create word dictionary
4. Model training (gensim)
5. Model tuning

Analysis:
The initial LDA model had a coherence score of 0.4523. After the fine-tuning, I improved coherence score to 0.4784, which would e a qualified result for further analysis.

Results Visualization

for clear presentation of the findings

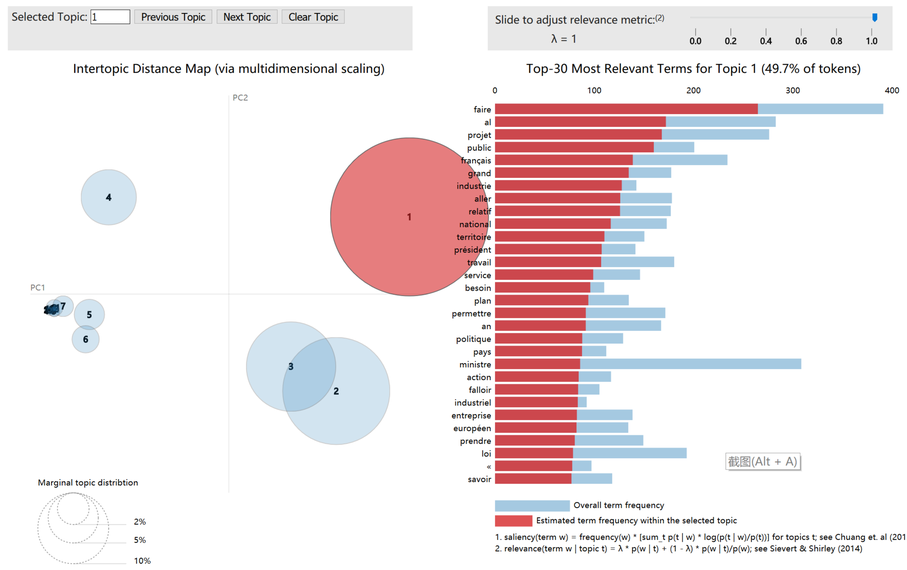
Method:
pyLDAvis library

Experiment:
generated two graphs for both the initial LDA model and the tuned one

Analysis:
The biggest topic cluster accounted for 49.7% of the total tokens. It had terms such as "industrie", "grand", "industriel", "enterprise", "français" etc. This suggests that Topic One could be highly industry/business relevant.

Topic two (22.9% of the tokens) seems to be rather environment related. However, compared to an environmental concern, topic two mentioned "mobilité", "énergétique", "port", or "transport" too often.

Environment related terms were more spread into other high ranking topics than industry related



Topic visualization using pyLDAvis on the tuned results

Analysis and Conclusion

1. This project finds that although the political motivation of the French government behind its electric vehicle policy remains mixed judging from publications of this government, **the eager for industrial transition upgrade outweighs environmental concerns.**

2. **environment issues** is associated with many topics and thus is maybe a **more wide-spread concern** of the French government. However, on concrete working projects, policymakers probably still let national industrial transition occupy much of their mind.

Both Scrapy and LDA are easily scalable and can be adapted to future research of a similar nature with a much bigger corpus. However, it still leaves an open question on how much domain knowledge is still required to select the right text for machine analysis or the "truth" will ultimately reveal itself as long as we feed the algorithm with enough data.