

Introduction to Python Project Proposal: Measuring the French Government Position on Promoting Electric Vehicles

Xiaoyan HU

x.hu@mpp.hertie-school.org

1. Introduction¹

The French government has proposed an ambitious plan in promoting the electric vehicle industry. It aims at multiplying the sales of electric vehicles (EV) by five times between 2018 and 2022. By the end of this five-year-plan, there should be one million electric or hybrid cars circulating in France, which shall contribute to an ultimate goal of the French government: a prohibition on the sale of fossil fuel cars by 2040. However, the reality falls short of the expectation of the French government. To boost the diffusion of electric vehicles, the French government employed various policy tools to boost the EV industry. Those policies range from direct public subvention to maintaining car factories, RD subsidies, to consumer incentives, and parking fee exemptions. State leaders such as Emmanuel Macron and Édouard Philippe visited BEV show rooms in international conferences and production sites, sending supporting signals to the EV market.

At first glance, one may argue that the close relationship between carbon emission reduction and electric vehicle promotion should infer that the French government push on this industry is mainly an environmental concern. After all, the first government plan on “véhicules décarbonés” (decarbonized vehicles) in the 21st century was raised by the French president of that time, Nicolas Sarkozy, as a response to the Grenelle Agreement. Full deployment of electric vehicles yields promising a picture for harnessing the speed of global warming. However, text analysis of more recent politician’s speech and government policy papers confirm that the intention of financial instruments also includes enforcing the competitiveness of the French automobile industry, such as the commission letter from Édouard Philippe on a government consulting report. In the engagement letter of a government commissioned report, French premier minister Edouard Philippe wrote: “it is clear that our industry is suffering a degree of erosion, and that the degradation of our trade balance in this sector has accelerated. There are, however, many opportunities, with ongoing

revolutions in the field of clean mobility and driverless and shared vehicles.” (Edouard Phillippe, 2018) [1]

In 2013, Bradley W. Lane and others [2] 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. In this framework, jurisdictions suite of policies falls in one of three categories: primary risk management, primary industrial policy, or a substantial blend of the two. By reading government documents, Lane concluded that France was an intermediate case with a substantial blend of industrial policy and risk management.

2. Motivation

Lane’s framework of analysis provided useful guidelines, but it has been seven years and two government changes since their research. It is time to apply the theory on more recent policy texts to update our understanding of the intention of this French government on the French EV industry. Instead of repeating the conventional way of generating policy positions, this team project wishes to draw inspiration from computer sciences and explore a more efficient way of collecting and classifying text. By doing so, not only can we collect and process considerably sized corpus with fewer costs but also replace the intuitive understanding of the text to something more measurable, putting subjectivity in a more objective criterion.

Therefore, this team project firstly wishes to **design a more efficient way of collecting text data from the open web by using Python web scrapping technique. Then we will classify the scrapped text with domain-specific knowledge and decide the general policy incentive of the French government.**

The value of this exercise extends beyond the very topic of this research. From a personal perspective, it is a revision of some traditional methods acquired last semester using the R language. While reviewing the knowledge from last year, this project will adapt learned models with a new language Python. Secondly, due to the limited data size that this project planned to collect, we will not be able to use much deep learning technique and will turn to rule-based

¹GitHub page link:<https://github.com/Xiaoyan-Adele/Political-Motivation-behind-French-EV-Promotion>

systems. While machine learning usually gives better results with much less effort involved when we have a considerable corpus size, it does not phase out the value of rule-based systems, especially when we try to simulate experts' reading of a much smaller sized text in specific domains without an already trained model. This much less sophisticated natural language processing poses an alternative to the more advanced algorithm which will be explored with another team project.

3. Evaluation

Successful completion of this team project consists of two parts. The first part is the definitive collection of text from the open web specified in the resources section. It shows that instead of searching keywords on google and inspect each link individually, this project manages to automate the process and generate a ready-to-use corpus for further analysis.

The second part involved topic classification and results representation. This project wishes to reproduce the word cloud technic learned last semester in R with the newly acquainted language Python. This project also aims at detecting document-level covariance and topical prevalence based on previous knowledge of the electric vehicle industry. Once the topic model is generated, we can decide whether the current French government support to the electric vehicle industry is out of environmental concern or industrial revival.

4. Resources

The policy context of this team project is closely linked to the master thesis research topic of the author. Therefore, prior to adopting computational methods, the author already has several government policy papers in French explaining the motivation behind fiscal support from both industrial and environmental perspectives. Those documents can help us to generate the classification keywords and conditions which are necessary for designing the web scrapping code

For web scrapping, there are ample code book and learning materials online. This project will look for those training materials on the open web and after understanding the mechanism, write our code for detecting relevant articles. To understand the French government positions on BEV market, this project will scrap materials from will be from the French president website (<https://www.elysee.fr/>) and the French prime minister website (<https://www.gouvernement.fr/discours-et-rapports>). Besides official release on government websites, we wish to explore twitter accounts of both state leaders since their inception to the office so as to complete our collection of the corpus.

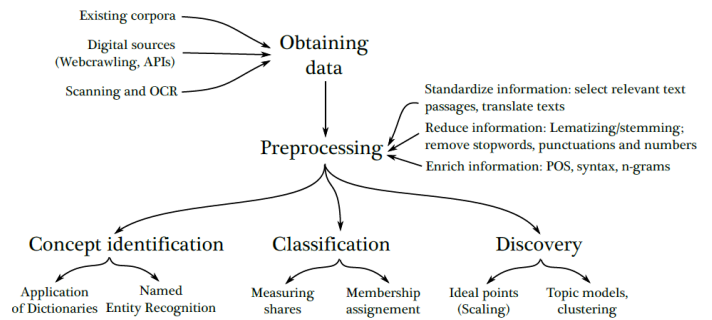


Figure 1. Overview of text-as-data applications by Gilardi and Wüest (2018)

During the methodology identification period, this paper consulted a recent publication by Gilardi and Wüest[3] to understand the location of our puzzle in the research world. The following mind map is their summary of text-as-data applications. Following the classification and literature review of Gilardi and Wüest, this project finds itself in the domain of topic modelling. Past studies has already explored potential toolkit for topic classifications. Starting from rule-based practice in the early 2000s to the popular machine learning models, existing literature provide rich methodologies and it is up for our project to verify which one is the most feasible for our project.

Roberts et al (2013) developed the structural topic model (STM), which allow us to estimate the quantities of interest in applied problems directly. While being an inspiring example, STM model is not available in Python. However, closely related to this R computational method, Latent Dirichlet Allocation (LDA), was already used in detecting causal inference mechanism from political science. Babera et al. (2014)[4] analysis tweets with LDA to generate the topics from both constituents and the members of Congress. This project will continue to explore the potential of employing LDA model in generating topics from unstructured text and combine it with other statistical methods to calculate and compare the covariance between the two sets of relations what we wish to examine in this project.

5. Contributions

As a one-person team project, instead of describing division of work int this section, the author will present the planned steps to meet the requirement of the course. The motivation of this project is to explore a more efficient way of summarising text using the Python knowledge learned in this course and applying it to a conventional text analysis study. Thus, the author will design the web scrapping code and use it to look for input data from the websites mentioned in Resources Section. After obtaining the raw dataset, the author shall pre-process the data including special character

cleaning, downcase the corpus, remove punctuation signs, equalize the predicting power of those processive pronouns, stemming/lemmatizing the corpus, and remove stop words from the nltk package. Upon successful cleaning of the text, the author shall be responsible for initial exploratory data analysis and build a topic classifier manually and apply the scrapped text onto the designed system.

In order to ensure that the rule-based systems reflect the semantic relationships in the corpus accurately, the author will request external help on the French language. This help is restricted to confirm the linguistic accuracy of the classification methods and thus will not introduce contribution to methodology design and result interpretation.

References

- [1] Xavier Mosquet, Patrick Pélata (2019). *Reinforcing the attractiveness and competitiveness of France in tomorrow's automotive industry and mobility*. Conseil National Industrie.
- [2] Lane, B. W.; Messer-Betts, N.; Hartmann, D.; Carley, S. (2013). *Government Promotion of the Electric Car: Risk Management or Industrial Policy?*. European Journal of Risk Regulation, 4(2), 227-246.
- [3] Fabrizio Gilardi and Bruno Wüest (2018). *Text-as-Data Methods for Comparative Policy Analysis*. <https://www.fabriziogilardi.org/resources/papers/Gilardi-Wueest-TextAsData-Policy-Analysis.pdf>
- [4] P. Barbera, R. Bonneau, P. Egan, J. T. Jost, J. Nagler, and J. Tucker.(2014). *Leaders or followers? Measuring political responsiveness in the U.S. congress using social media data..* Working Paper
- [5] Ryan Wesslen (2018). *Computer-Assisted Text Analysis for Social Science: Topic Models and Beyond*. arXiv:1803.11045v2 [cs.CL]