

# NLP For Economists

## NLP for Economics: An Overview

Sowmya Vajjala

Munich Graduate School of Economics - LMU Munich

Guest Course, October 2020

# Goal for this session

- ▶ Introduce some usecases for NLP in Economics research

Based on: Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. Journal of Economic Literature, 57(3), 535-74.

# Why bother about NLP if I am an Economist?

- ▶ Language is an important means of communication for all human beings.
- ▶ Lot of research and news is still published in the form of text.
- ▶ Economists too have to read and understand a lot of text to do their job.
- ▶ Text as data also serves as a rich complement to the kind of numeric/tabular data you usually see in economics research.

# What is different about text data?

- ▶ Inherently high-dimensional (unlike other kinds of data economists encounter normally).
- ▶ There is a need to figure out how to represent raw text as a fixed dimensional array.
- ▶ Then, you have to understand how to use this array to predict, describe or do causal analysis etc. (although a more common use case is prediction)

## Example usecases

- ▶ In financial economics, "text from financial news, social media, and company filings is used to predict asset price movements and study the causal impact of new information."

## Example usecases

- ▶ In financial economics, "text from financial news, social media, and company filings is used to predict asset price movements and study the causal impact of new information."
- ▶ "In macroeconomics, text is used to forecast variation in inflation and unemployment and estimate the effects of policy uncertainty."

## Example usecases

- ▶ In financial economics, "text from financial news, social media, and company filings is used to predict asset price movements and study the causal impact of new information."
- ▶ "In macroeconomics, text is used to forecast variation in inflation and unemployment and estimate the effects of policy uncertainty."
- ▶ "In media economics, text from news and social media is used to study the drivers and effects of political slant."

## Example usecases

- ▶ In financial economics, "text from financial news, social media, and company filings is used to predict asset price movements and study the causal impact of new information."
- ▶ "In macroeconomics, text is used to forecast variation in inflation and unemployment and estimate the effects of policy uncertainty."
- ▶ "In media economics, text from news and social media is used to study the drivers and effects of political slant."
- ▶ "In industrial organization and marketing, text from advertisements and product reviews is used to study the drivers of consumer decision making."



## Example usecases

- ▶ In financial economics, "text from financial news, social media, and company filings is used to predict asset price movements and study the causal impact of new information."
- ▶ "In macroeconomics, text is used to forecast variation in inflation and unemployment and estimate the effects of policy uncertainty."
- ▶ "In media economics, text from news and social media is used to study the drivers and effects of political slant."
- ▶ "In industrial organization and marketing, text from advertisements and product reviews is used to study the drivers of consumer decision making."
- ▶ "In political economy, text from politicians' speeches is used to study the dynamics of political agendas and debate."

# Example usecases with causal inference goals

- ▶ using "Google search data to estimate local areas' racial animus, then studies the causal effect of racial animus on votes for Barack Obama in the 2008 election."
- ▶ using "congressional and news text to estimate each news outlet's political slant, then study the supply and demand forces that determine slant in equilibrium."
- ▶ "measure local news coverage of earnings announcements, then use the relationship between coverage and trading by local investors to separate the causal effect of news from other sources of correlation between news and stock prices."

# Examples from a methods perspective

- ▶ authorship attribution approaches from NLP "to answer an authorship question of more direct interest to economists: who invented instrumental variables?"
- ▶ word frequencies are used for "a regression approach to construct an index of news-implied market volatility based on text from the Wall Street Journal from 1890–2009."
- ▶ topic models on diaries of CEOs to identify the behavioral patterns of leaders/managers.
- ▶ sentiment analysis to study financial stability

.... and so on.

# Various forms of data

- ▶ news articles
- ▶ social media posts
- ▶ CEO diaries
- ▶ research articles
- ▶ quarterly earnings call transcripts
- ▶ web searches
- ▶ company disclosures
- ▶ patents

.. ...

# Summary

- ▶ Clearly, there are many use cases for NLP methods in various branches of economics.
- ▶ All sorts of texts from tweets to bank documents are used in these analyses.
- ▶ A range of NLP approaches, coupled with machine learning methods are commonly used for predictive analyses.
- ▶ Although not as common, using these methods to understand causal inferences is also not uncommon.

.. ...

# ToDo for you

- ▶ I strongly recommend reading Gentzkow et.al.'s article!
- ▶ Check the below article for asomewhat different approach, and examples in German! *Siegel, M. (2018). Text Mining in Economics. In Semantic Applications (pp. 63-73). Springer Vieweg, Berlin, Heidelberg.*
- ▶ Check the Assignment 0 on course website (<https://econnlpcourse.github.io/lectures>)
- ▶ Think about what kind of problems in your research area can benefit from knowing about NLP methods.