

Escaping the Filter Bubble

Emily Ng





Getting Stuck In a Filter Bubble

Traditional information recommender systems promote filter bubbles.

Dangers:

- Decreases empathy
- Promotes ignorance
- Hinders progress



GOAL:

Create a system that recommends articles of the
opposite polarity.

Pipeline

Data



NLP



Clustering



Sentiment



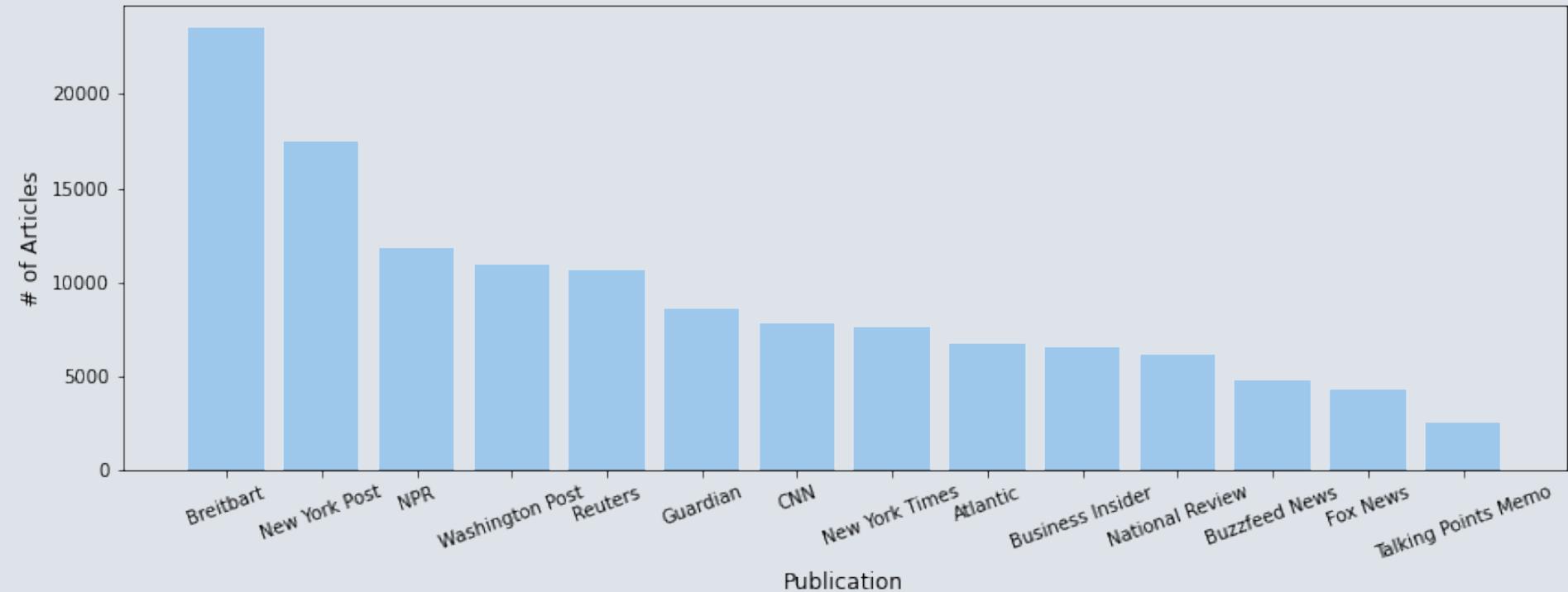
Recommender

143,000 articles

15 American Publications



Articles by Publication

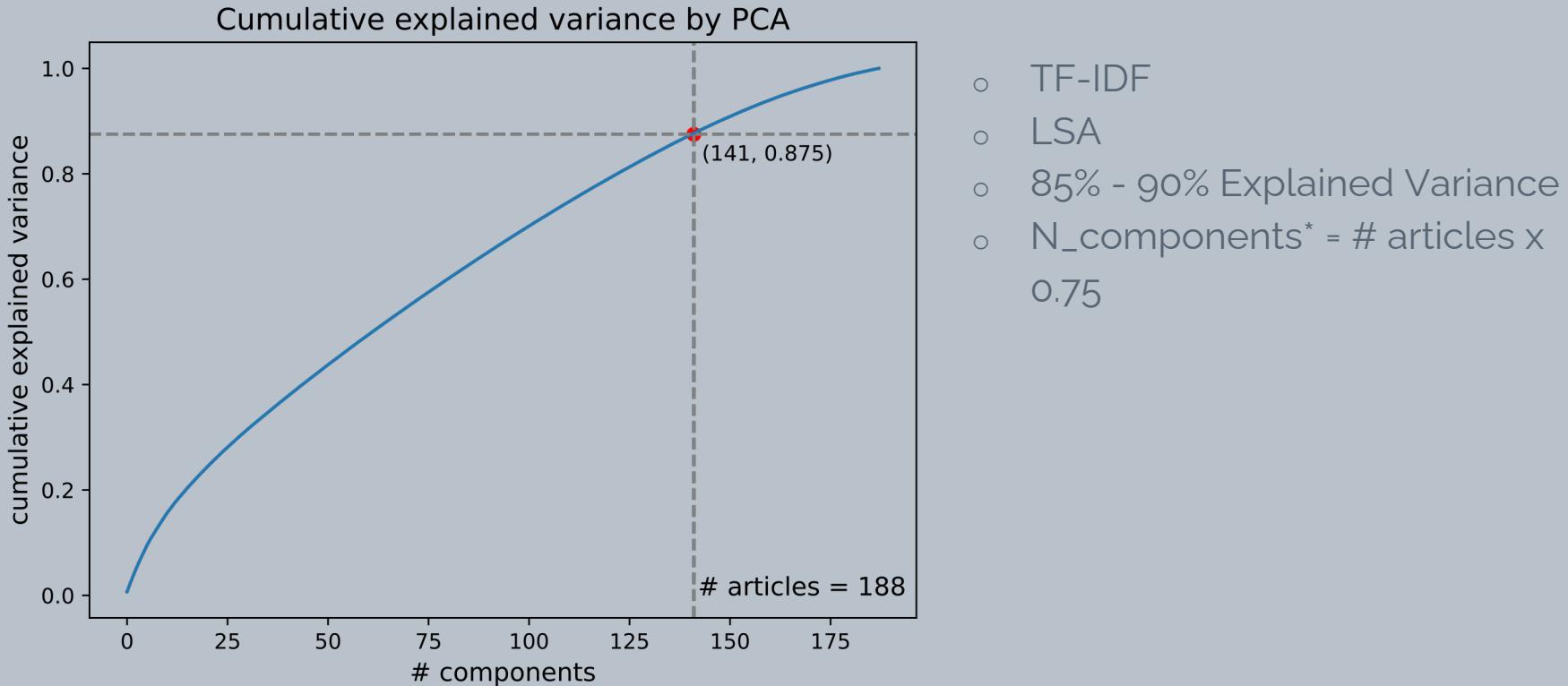


Model

WORLD BUSINESS | B3

countries have
effect and will bring material
benefits. That's
why we are
optimistic.

NLP Implementation



Recursive LSA vs Hierarchical Clustering

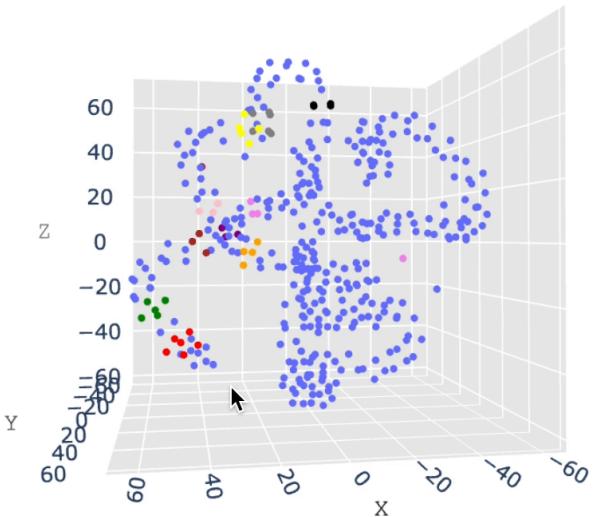
LSA

- Classify article as topic with highest coefficient
- Required multiple passes to get desired resolution of clusters
- More likely to get inaccurate clustering of articles (false positives)

Hierarchical Clustering (Agglomerative)

- $N_{clusters} = N_{components}$
- Higher precision than LSA
- More likely to split a large group of articles into 2 smaller ones despite them being able to be classified as one.

TSNE Visualization



- All Articles
- Dodd-Frank Executive Order
- Kellyanne Conway-Bowling Green
- Snapchat IPO
- Trump-Australia Call
- Super Bowl-Tom Brady
- Iran Sanctions
- Job Growth
- Yemen Raid
- Incohesive Grouping
- North Korea Tension

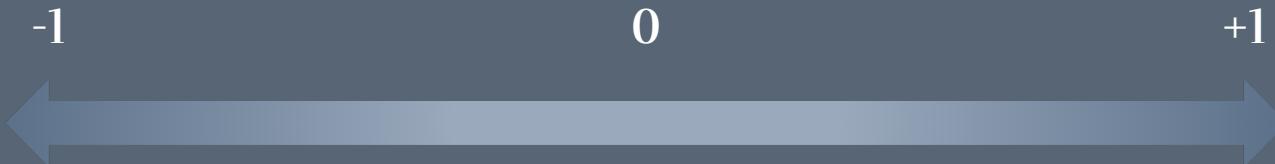


Recommender

Sentiment Assignment

- NLTK VADER
 - -1: Negative Polarity
 - 0: Neutral
 - +1: Positive Polarity
- Based on intro and conclusion segments

Polarity



Article

1 2 3 4 5



News Article Recommender



Enter the date of publication, publication source, and article content and we will recommend an article for you.

Date (YYYY-MM-DD)

2017-02-03

Publication

CNN

Current Limitations

- Current model is limited to only articles written on the same day.
- Disproportionate publication representation may skew polarity scale.
- Many of the articles are unique and can not be assigned to a cluster.
- Limited by the number of articles and publications represented in dataset.

Future Work

- Develop a metric to evaluate accuracy of cluster assignments
- Scale with dataset provided by [Components](#)
 - 2.7M articles, 27 publications
- Improve model efficiency



THANKS!

Any questions?