POIR 613: Computational Social Science

Pablo Barberá

University of Southern California
`pablobarbera.com`

Course website:

pablobarbera.com/POIR613/

Today

1. Zoom discussion sessions (for now):
   - November 10th
2. Project
   - All project ideas should be in; feedback sent
   - Next milestone: 5-page summary that includes some data analysis by November 5th
3. Word embeddings
   - Overview
   - Applications
   - Bias
   - Demo
4. Ideological scaling using text
5. Solutions to challenge 7
6. More word embeddings

Overview of QTA (Grimmer and Stewart, 2013)

1. Acquire textual data:
    - Existing corpora; scraped data; digitized text
2. Preprocess the data:
    - Bag-of-words vs word embeddings
3. Apply method appropriate to research goal:
    - Describe and compare documents
        - Readability; similarity; keyness metrics
    - Classify documents into known categories
        - Dictionary methods
        - Supervised machine learning
    - Classify documents into unknown categories
        - Document clustering
        - Topic models
    - Scale documents on latent dimension
        - Known dimension: wordscores
        - Unknown dimensions: wordfish

Word embeddings

Beyond bag-of-words

Most applications of text analysis rely on a bag-of-words representation of documents

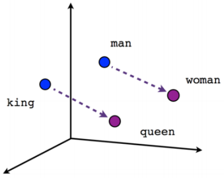
- Only relevant feature: frequency of features
- Ignores context, grammar, word order...
- Wrong but often irrelevant

One alternative: word embeddings

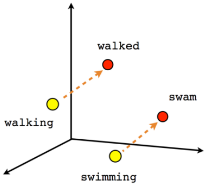
- Represent words as real-valued vector in a multidimensional space (often 100–500 dimensions), common to all words
- Distance in space captures syntactic and semantic regularities, i.e. words that are close in space have similar meaning
  - How? Vectors are learned based on context similarity
  - Distributional hypothesis: words that appear in the same context share semantic meaning
- Operations with vectors are also meaningful

Word embeddings example

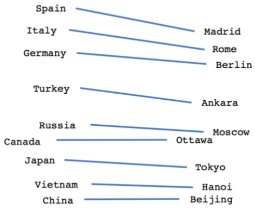
word	D_1	D_2	D_3	...	D_N
man	0.46	0.67	0.05	...	...
woman	0.46	-0.89	-0.08	...	...
king	0.79	0.96	0.02	...	...
queen	0.80	-0.58	-0.14	...	...



Male-Female



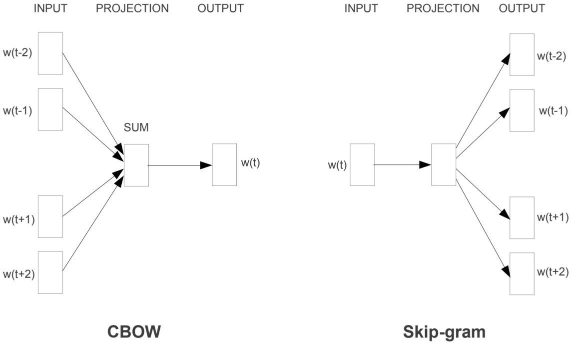
Verb tense



Country-Capital

word2vec (Mikolov 2013)

- Statistical method to efficiently learn word embeddings from a corpus, developed by Google engineer (now at FB)
- Most popular, in part because pre-trained vectors are available
- Two models to learn word embeddings:



word2vec (Mikolov 2013)

How does the model learn about embeddings?

- Consider the following sentences:
  - I study Math at school
  - I study Geography at school
  - You study Biology at school
- The model will learn that the words Math, Geography, and Biology must have a similar meaning because they appear in similar contexts
- i.e. they will be estimated to have similar embeddings

Other embedding methods

- GloVe embeddings (Stanford NLP group)
  - Trained using global co-occurrence
  - Less corpus-specific than word2vec, but differences are minimal (Rodriguez and Spirling, 2021)
- Google's Bidirectional Encoder Representations from Transformer (BERT)
  - Builds on recent advances in deep learning
  - Key difference: words' embeddings depend on context, and are not fixed

Word embeddings

- Overview
- Applications
- Bias
- Demo

Applications

Three main social science applications of word embeddings:

1. Alternative to bag-of-words feature representation in supervised learning tasks:
   - Can improve performance with small labeled sets
   - Takes context into account
2. Support for other automated text analysis tasks:
   - Expand dictionaries
   - Evaluate coherence of topics models
3. Understanding word meaning
   - Analysis of semantic shifts over time
   - Study of word meaning varies by groups

Dictionary expansion

Using word embeddings to expand dictionaries (e.g. incivility)

```
> distance(file_name = "FBvec.bin",
+          search_word = "libtard",
+          num = 10)
Entered word or sentence: libtard

Word: libtard  Position in vocabulary: 5753
          word              dist
1          lib 0.798957586288452
2        lefty 0.771853387355804
3       libturd 0.762575328350067
4     teabagger 0.744283258914948
5      teabilly 0.715277075767517
6       liberal 0.709996342658997
7        retard 0.690707504749298
8       dumbass 0.690422177314758
9          rwnj 0.684058785438538
10 republitard 0.678197801113129
```

```
> distance(file_name = "FBvec.bin",
+          search_word = "idiot",
+          num = 10)
Entered word or sentence: idiot

Word: idiot  Position in vocabulary: 646
        word              dist
1   imbecile 0.867565214633942
2    asshole 0.848560094833374
3      moron 0.781079053878784
4     asshat 0.772150039672852
5     a-hole 0.765781462192535
6      ahole 0.760824918746948
7    asswipe 0.742586553096771
8  ignoramus 0.735219776630402
9   arsehole 0.732272684574127
10     idoit 0.720151424407959
```

Source: Timm and Barberá, 2019

The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings

American Sociological Review
1–45

DOI: 10.1177/0003122419877135
journals.sagepub.com/home/asr
SAGE

Austin C. Kozlowski,[a] Matt Taddy,[b] and James A. Evans[a,c]

Abstract

We argue word embedding models are a useful tool for the study of culture using a historical analysis of shared understandings of social class as an empirical case. Word embeddings represent semantic relations between words as relationships between vectors in a high-dimensional space, specifying a relational model of meaning consistent with contemporary theories of culture. Dimensions induced by word differences (*rich* – *poor*) in these spaces correspond to dimensions of cultural meaning, and the projection of words onto these dimensions reflects widely shared associations, which we validate with surveys. Analyzing text from millions of books published over 100 years, we show that the markers of class continuously shifted amidst the economic transformations of the twentieth century, yet the basic cultural dimensions of class remained remarkably stable. The notable exception is education, which became tightly linked to affluence independent of its association with cultivated taste.

Keywords

word embeddings, *word2vec*, culture, computational sociology, methodology, text analysis, content analysis

Source: Kozlowski et al, ASR 2019

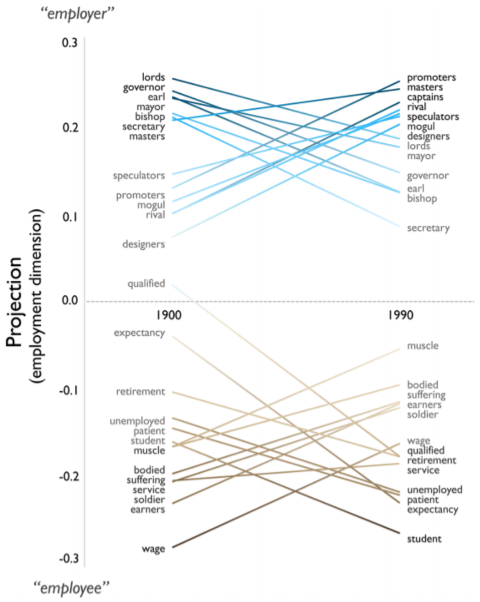


Figure 10. Words That Project High and Low on the Employment Dimension of Word Embedding Models Trained on Texts Published at the Beginning and End of the Twentieth Century; 1900–1919 and 1980–1999 Google Ngrams Corpus

Cooperation in the international system

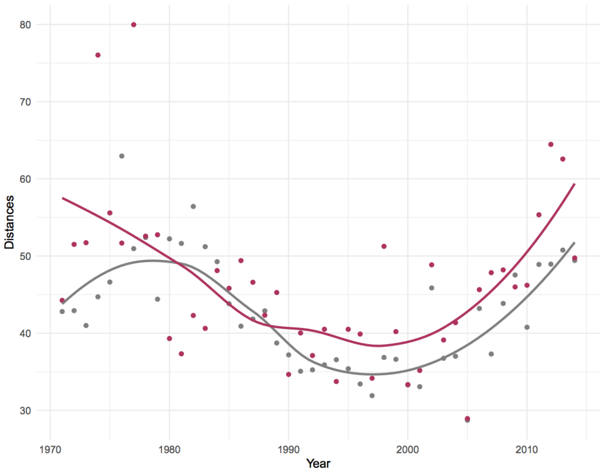
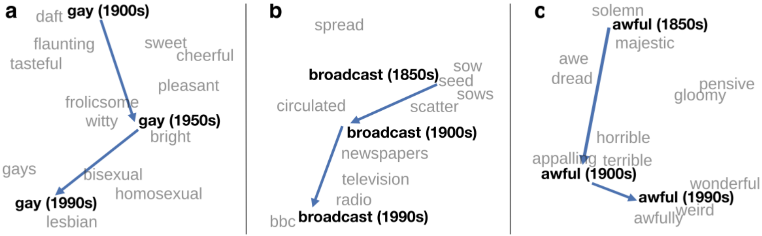


Figure 4: *Distances by core countries*. Plot of Euclidian distances between US and Russia (gray), and US and China (maroon).

Source: Pomeroy et al 2018

Semantic shifts

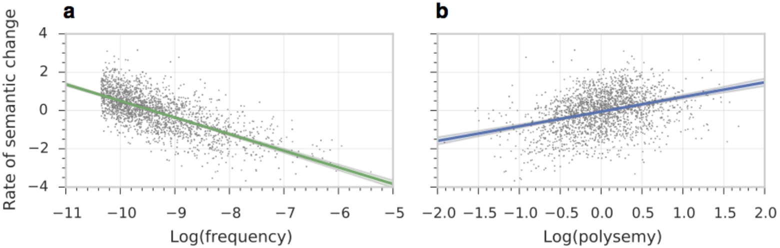
Using word embeddings to visualize changes in word meaning:



Source: Hamilton et al, 2016 ACL.
https://nlp.stanford.edu/projects/histwords/

Application: semantic shifts

1. **Law of conformity**: words that are used more frequently change less and have more stable meanings
2. **Law of innovation**: words that are polysemous (have many meanings) change at faster rates.



Source: Hamilton et al, 2016 ACL.
https://nlp.stanford.edu/projects/histwords/

Word embeddings

- Overview
- Applications
- Bias
- Demo

Bias in word embeddings

Semantic relationships in embeddings space capture stereotypes:

- Neutral example: man – woman ≈ king – queen
- Biased example: man – woman ≈ computer programmer – homemaker

Gender stereotype *she-he* analogies.

sewing-carpentry	register-nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	hairdresser-barber

Gender appropriate *she-he* analogies.

queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

Source: Bolukbasi et al, 2016. arXiv:1607.06520

See also Garg et al, 2018 PNAS and Caliskan et al, 2017 Science.

Word embeddings

- Overview
- Applications
- Bias
- Demo

Ideological scaling using text as data

Wordscores (Laver, Benoit, Garry, 2003, APSR)

- Goal: estimate positions on a latent ideological scale
- Data = document-term matrix $\mathbf{W_R}$ for set of "reference" texts, each with known A_{rd}, a policy position on dimension d.
- Compute $\mathbf{F}$, where F_{rm} is relative frequency of word m over the total number of words in document r.
- Scores for individual words:
  - $P_{rm} = \frac{F_{rm}}{\sum_r F_{rm}} \rightarrow$ (Prob. we are reading r if we observe m)
  - Wordscore $S_{md} = \sum_r (P_{rm} \times A_{rd})$
- Scores for "virgin" texts:
  - $S_{vd} = \sum_w (F_{vm} \times S_{md}) \rightarrow$ (weighted average of scored words)
  - $S^*_{vd} = (S_{vd} - \overline{S_{vd}}) \left(\frac{SD_{rd}}{SD_{vd}}\right) + \overline{S_{vd}} \rightarrow$ Rescaled scores.

Wordfish (Slapin and Proksch, 2008, AJPS)

- Goal: unsupervised scaling of ideological positions
- Ideology of politician *i*, θ_i is a position in a latent scale.
- Word usage is drawn from a Poisson-IRT model:

$$W_{im} \sim \mathrm{Poisson}(\lambda_{im})$$
$$\lambda_{im} = exp(\alpha_i + \psi_m + \beta_m \times \theta_i)$$

- where:

  α_i is “loquaciousness” of politician *i*
  ψ_m is frequency of word *m*
  β_m is discrimination parameter of word *m*

- Estimation using EM algorithm.
- Identification:
  - Unit variance restriction for θ_i
  - Choose *a* and *b* such that $\theta_a > \theta_b$