Bi-Weekly Colloquium

10 December 2021

Özge Erten, ESR8

Dynamic Word Embeddings for Evolving Semantic Discovery

Authors:Zijun Yao, Yifan Sun, Weicong Ding, Nikhil Rao, Hui Xiong

Year: 2018

Dynamic Word Embeddings for Evolving Semantic Discovery

- ➔ Problem: Adequately capturing the evolving language structure and vocabulary.
- ➔ Proposed Model: A dynamic statistical model to learn time-aware word vector representation.

For instance;

real estate → television → republican

Methodology

T = time

D = text corpus and Dt,t =1,...T, is the corpus of all documents in the t -th time slice.

V = words present in the corpus at any point in time and V = {w1, . . . , wV } overall vocabulary

L = window size, for exp. This is a sample sentence for explaining window size.

uw = The static embedding for word w

U(t) = The embedding matrix of all words

Let's assume we have a time period from January to April and we have collections about magazine. D1 denotes text corpus of magazines in January and DT denotes text corpus of magazines in April. V contains all vocabulary from January to April.

Time-Agnostic Embeddings

➔ Pointwise Mutual Information (PMI) matrix

- Observation: “Semantically similar words often have similar neighboring words in a corpus”.

$$\mathrm{PMI}(\mathcal{D}, L)_{w,c} = \log\left(\frac{\#(w,c)\cdot|\mathcal{D}|}{\#(w)\cdot\#(c)}\right)$$

- PMI matrix calculates for w and c-th entry to find embedding vectors uw and uc.

The Alignment Problem

- ➔ Alignment for different time periods is an issue in general if embeddings are learned independently for each time slice.
- ➔ Solution: In the paper, the word embeddings learn across time jointly
  - ◆ Thus obviating the need to solve a separate alignment problem.
- ➔ Proposed method:

Temporal embeddings in all time slices concurrently + Regularization terms

Temporal Word Embeddings

➔ An extension: Using matrix factorization technique on each time slice Dt separately.

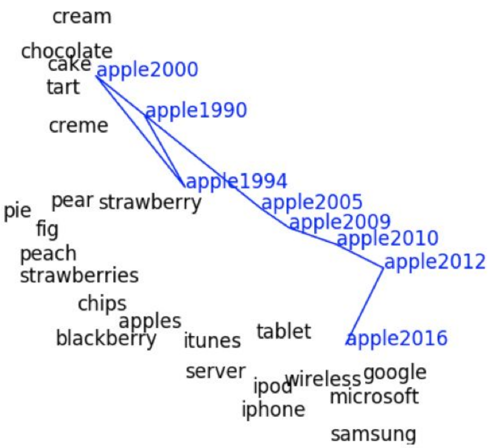
$$\mathrm{PPMI}(t, L)_{w,c} = \max\{\mathrm{PMI}(\mathcal{D}_t, L)_{w,c}, 0\}. := Y(t).$$

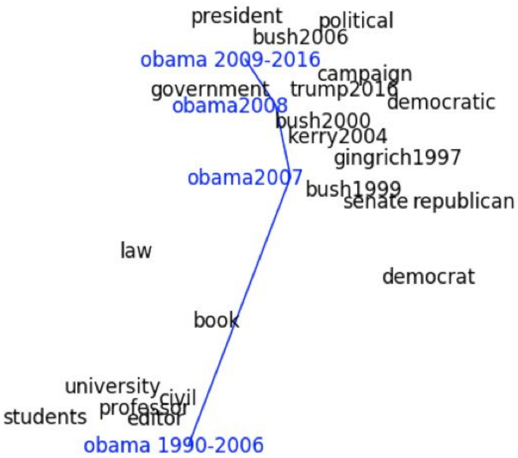
➔ The model;

$$\min_{U(1),\ldots,U(T)} \frac{1}{2}\sum_{t=1}^{T} \|Y(t) - U(t)U(t)^T\|_F^2 + \frac{\lambda}{2}\sum_{t=1}^{T}\|U(t)\|_F^2 + \frac{\tau}{2}\sum_{t=2}^{T}\|U(t-1) - U(t)\|_F^2,$$

Dataset and Experiments

- New York Times, published a total of 99,872 articles
  - Between January 1990 and July 2016.
  - D1 = articles in January 1990 and DT = articles in July 2016.
  - Time: They use yearly time slices, dividing the corpus into T = 27 partitions.
  - V: 20, 936 unique words.
- Trajectory visualization:

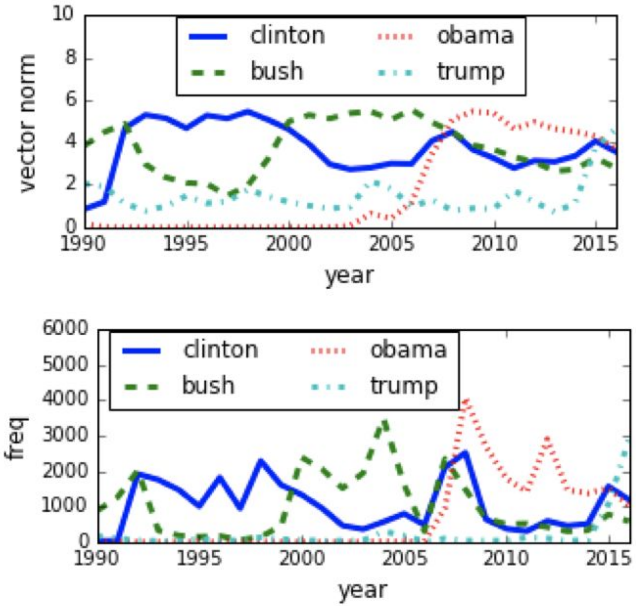




➔ Equivalence searching

<table>
<tr><th>Question</th><th>US president</th><th>NYC mayor</th></tr>
<tr><td>Query</td><td>obama, 2016</td><td>blasio, 2015</td></tr>
<tr><td>90-92</td><td>bush</td><td rowspan="2">dinkins</td></tr>
<tr><td>93</td><td rowspan="2">clinton</td></tr>
<tr><td>94-00</td><td rowspan="2">giuliani</td></tr>
<tr><td>01</td><td rowspan="5">bush</td></tr>
<tr><td>02-05</td><td>bloomberg</td></tr>
<tr><td>06</td><td>n/a*</td></tr>
<tr><td>07</td><td rowspan="3">bloomberg</td></tr>
<tr><td>08</td></tr>
<tr><td>09-10</td><td rowspan="4">obama</td></tr>
<tr><td>11</td><td>cuomo*</td></tr>
<tr><td>12</td><td>bloomberg</td></tr>
<tr><td>13-16</td><td>blasio</td></tr>
</table>

→ Popularity determination



Conclusion

- → Higher interpretability for embeddings,
- → Better quality with less data,
- → More reliable alignment for across-time querying.

Reference

- Yao, Z., Sun, Y., Ding, W., Rao, N., & Xiong, H. (2018, February). Dynamic word embeddings for evolving semantic discovery. In Proceedings of the eleventh acm international conference on web search and data mining (pp. 673-681).

Thank you!