Deep Neural Models of Semantic Shift

Alex Rosenfeld & Katrin Erk

Aminata Ndiaye & Edimah Songo
M2 MASH

Main question

How do words and their meanings change over time ?

Figure: "Cell" - 1670s

Figure: "Cell" - 1980s

Figure: "Cell" - 1990s

With a specific Neural Network architecture, we can DETECT & QUANTIFY these over time changes !

Modelling the change of words over time : then and now

Past approach :

- Separate data in time bins (eras, centuries, decades, etc.)
- Train different model on each bin
- Output : word vector

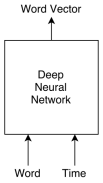


Figure: DiffTime model's approach

New approach :

- Time as a continuous variable
- One model
- Output : function that outputs word vector for a given time

Figure: Semantic trajectories of words

Outline

Outline

1 State of the art models

2 What is the DiffTime Model ?

3 What is the best model?

4 How to measure the speed of word change ?

5 Conclusion

State of the art models

No distributional models

- Topic modeling
- Sentiment analysis

Distributional models

Diachronic models

- Period of time separated in several bins
- Synchronic model for each time bin trained independently of the others

The bins approach

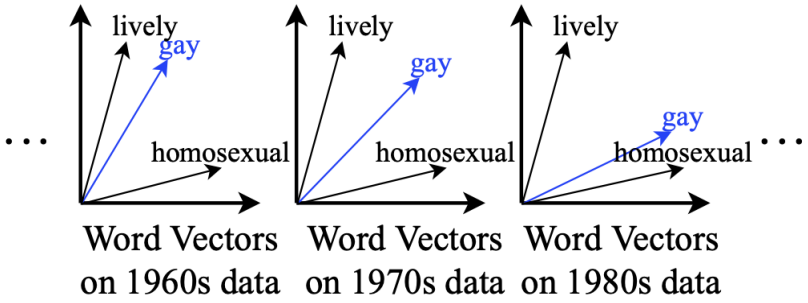


Figure: Evolution of the representation vectors of the word 'gay'

Different bins size

Large bins

- **LargeBin**
- Bins of 5 years
- Hamilton et al. (2016)

Small bins

- **SmallBinPreInit**
- Bins of 1 year
- Kim et al. (2014)
- preinitializing each bin with the vectors of the previous bin training

Prior and Transition Probabilities Model

- Bamler and Mandt (2017)
- transform the SGNS loss function into a probability distribution over the target and context vectors.
- discourage a vector variable from deviating from the previous bin's vectors.
- **SmallBinReg**

Outline

The DiffTime model

Goal

To obtain a word vector for each timepoint t.

- Input :
  - t : continuous variable
  - w : target word
- Output : $use_w(w, t)$ a word vector

Same goes for word vectors for context words c : output is $use_c(c, t)$.

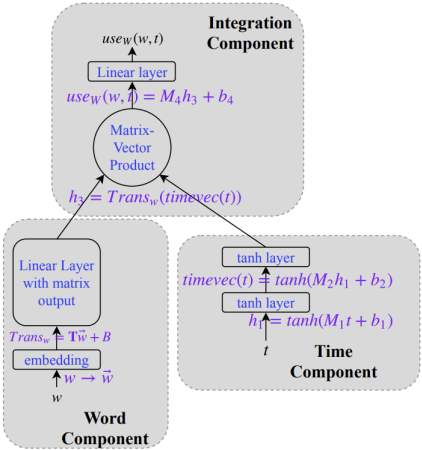


Figure: Diagram of DiffTime

Outline

Synthetic tasks

Figure: Banana

Figure: Lobster

- Sigmoidal path
- banana + lobster $\rightarrow$ banana$\circ$lobster

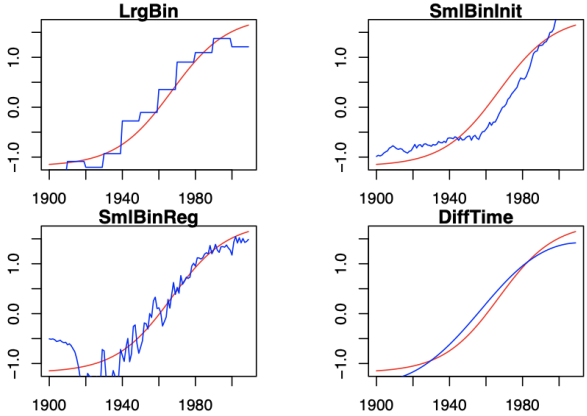


Figure: Graph Comparisons between synthetic path (red) and predicted paths (blue) for the synthetic word pistol∘elm.

The Results

Method	AMSE 1900–2009	AMSE 1950–2009
LargeBin	62.52	51.71
SmallBinPreInit	171.43	49.88
SmallBinReg	106.79	42.67
DiffTime	**25.67**	**11.48**

Figure: Model performance under the synthetic evaluation. The values are the mean sum of squares error (MSSE) for each method. Lower value is better. The first column is MSSE using all times. The second column is MSSE using years 1950 to 2009.

Bias of the evaluation technique

Strength

- Easy to understand and interpret
- Provide useful information on the quality of a diachronic distributional model

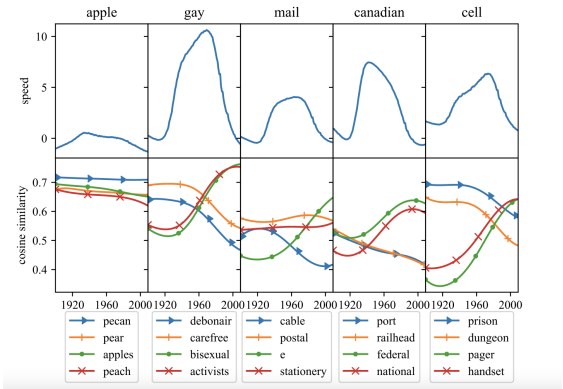
Weakness

- operates on synthetic words
- only generate words that shift from one sense to another
- we may have privileged continuous models that incorporate a sigmoidal function in their architecture.

Outline

From the differentiability of the model :

$$\text{speed} = ||\nabla_t(\frac{use_W(w,t)}{||use_W(w,\cdot)||_2})||$$



Outline

Conclusion

Article's takeaways

- Modelling semantic shift with time as a **continuous** variable.
- New way to evaluate the performance of the model
- Measure the speed of change in word meaning

Drawbacks

- Sometimes blurry explanations & graphs
- Evaluation method untested on real data

Wide range of applications :

- Humanities research (linguistics, history, etc.)
- Marketing & politics (emerging trends online, etc.)
- NLP (improve text classification, sentiment analysis, etc.)