732A92/TDDE16 Text Mining (2022)

Word embeddings

Marco Kuhlmann
Department of Computer and Information Science

LINKÖPING UNIVERSITY

This session

- Questions and answers
- Word embeddings and stereotypes
- Introduction to the lab

Questions and answers

Overview of word embeddings

1. Introduction to word embeddings
2. Learning word embeddings via matrix factorisation
3. Learning word embeddings with neural networks
4. The skip-gram model
5. Subword models
6. Contextualised word embeddings

Playground

https://projector.tensorflow.org

- Play around with different types of word embeddings and visualisations (PCA, T-SNE, UMAP).
- Upload and visualise your own vectors.

Word embeddings and stereotypes

Project structure

1. Identify your problem — 8 hours (w44–w48)
2. Design your approach — 32 hours (w49–w50)
3. Evaluate your approach — 32 hours (w51–w01)
4. Produce your report — 16 hours (w02)

Embedding bias and occupation participation

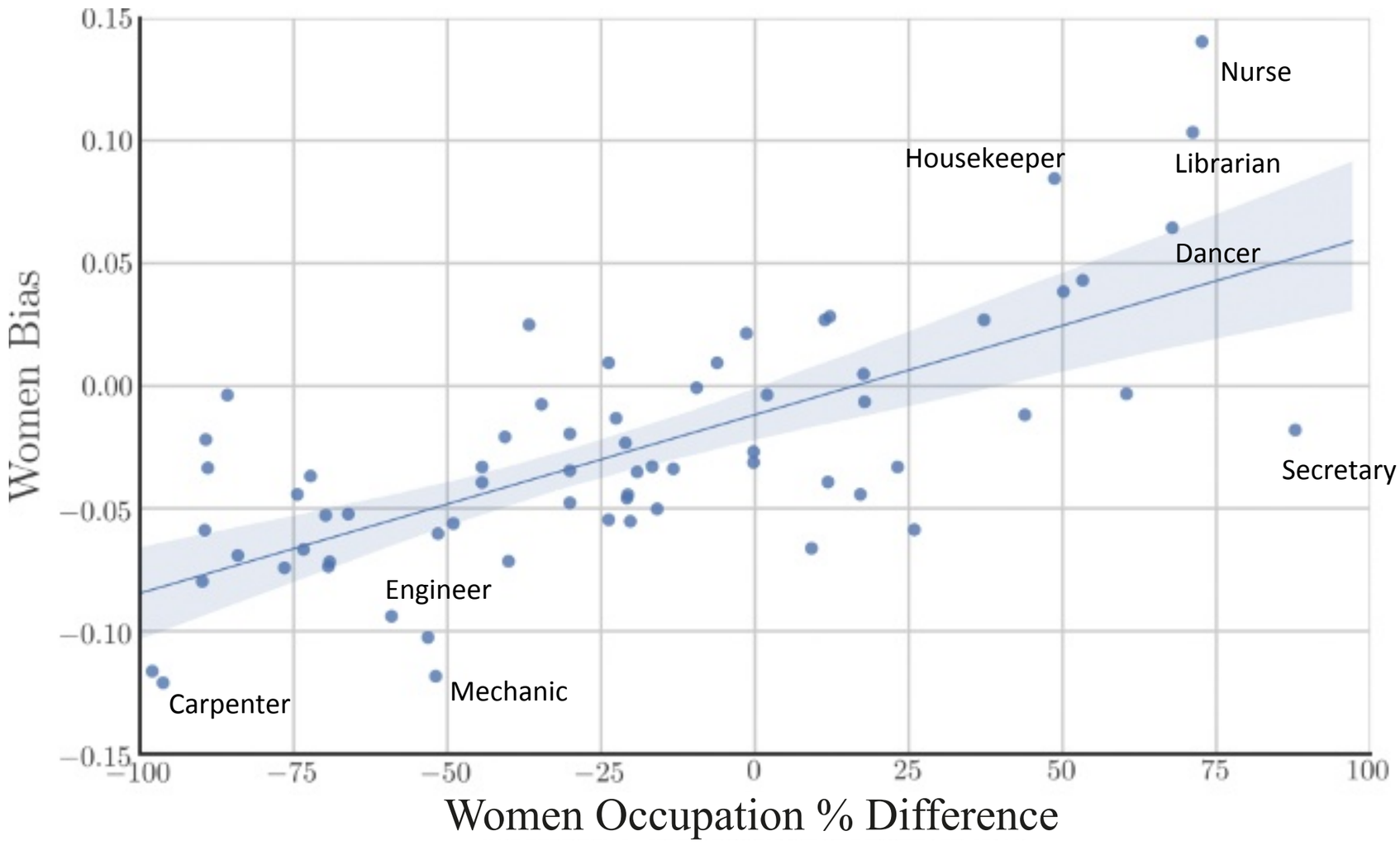

Figure 1 from Garg et al. (2018)

Embedding bias

- Train word embeddings on the data under consideration.
- For each occupation word (e.g. *teacher*, *lawyer*), compute the average embedding distance between that word and reference words that represent women/men (e.g. *she*, *female* vs. *he*, *male*).
- Define the **women bias** of an occupation word as the average distance for men minus the average distance for women.

Project 1

The project aims to explore the strength of gender biases in conservative and liberal media in the US with the help of word embeddings. The work employs the dataset of 80,000 articles from the 6 largest news sources with varying political inclination, and compares semantic distances from male- and female-denoting terms to a set of contextual words. Making use of Python's Gensim, I build word2vec models for every news source, align them and measure Euclidean distance between embedding vectors. A higher difference between the distances for contextual words shows stronger bias. My hypothesis of conservative media sources having stronger biases was not supported by the data. Further examination of the results shows that the structure of biases is complex and requires more data and adjustments in the methodology. Overall, both in conservative and liberal news reports, gender biases are substantial and need further studying.

Project 2

The report investigates the attitude towards immigrants in labor market among Swedish unions, through analyzing documents from Swedish unions' press conferences. A CBOW model configured with hierarchical softmax is employed to train on the dataset over different time span and unions. Relative norm distances between neutral and target word lists for Swedish and immigrants oriented groups are extracted from the trained word embeddings. By analyzing the values of relative norm distances through years and unions, a relatively negative attitude towards immigrants among Swedish unions is detected in 2015 and 2016, and *Fastighetsanställdas förbund* shows a less positive attitude than other unions. Taking limited dataset and resources into account, improvements are possible for further research.

Assessment criteria – Method

Is the data used in the project suitable for the stated problem?
Are technical concepts, models and algorithms applied correctly?
Are the experimental results validated with appropriate evaluation methods?

- F – The problem should have been approached differently.
  The choice of the data, models, algorithms or evaluation methods is not appropriate, or there is too little information in the report to assess whether the choice was appropriate.

- E – The data used in the project is suitable for the stated problem.
  Technical concepts, models and algorithms are applied correctly.
  The experimental results are validated with appropriate evaluation methods.

- A – The data is created specifically for the project.
  The project involves non-trivial modifications or combinations of models and algorithms.
  The experimental results are validated using several complementary evaluation methods.

Group discussions

Choose one of the two projects.

- Choose one aspect of the project that you found particularly interesting. Motivate your choice.
- Suppose that you would want to replicate the experimental results of the chosen project. What information do you need?
- Which parts of the evaluation method do you not understand or have concerns about (based on the abstract)?

Suggested structure (1)

- **Introduction**

  What problem did you address in the project? Why is this problem interesting? What can we learn by solving the problem?

- **Theory**

  Present relevant theoretical background, and in particular those concepts and methods that were not covered in the course.

Suggested structure (2)

- **Data**

  What data did you use in your project? How was this data created? What preprocessing did you do (if any), and why?

- **Method**

  Explain how you approached the stated problem. Aim to be detailed enough for others to reproduce your results.

- **Results**

  Present your results in an objective way. Use tables and charts, but do not forget to also include a summary in text form.

Suggested structure (3)

- **Discussion**

  Analyse your results. Discuss the limitations of your work. Compare your study to related work, such as internet materials or scientific articles.

- **Conclusion**

  Summarise your analysis. To what extent did you solve your stated problem? What else do you take away from your project?

Introduction to the lab