Visual Exploration of Word Vector Embeddings

Florian Heimerl[*]
University of Wisconsin – Madison

Michael Gleicher[†]
University of Wisconsin – Madison

Abstract

The use of word vector embeddings as the basis for many upstream tasks in text processing has lead to large improvements in accuracy. However, the exact reasons for this success largely remain unclear, as the properties and relations that these embeddings encode are often not well understood. Our goal in this ongoing project is to design effective interactive visualizations that help practitioners and researchers understand and compare such spaces better. The initial steps we have taken is to review relevant literature to identify properties and relations of word vectors that are important for various applications. From these, we derive basic tasks to inform the design of adequate and effective interactive visualizations that help users gain deeper insights into the structure of vector spaces. In addition, we present three initial designs to support these tasks.

Index Terms: text data, task-based design, vector embeddings

1 Introduction

Vector embeddings are a collection of statistical techniques that place complex objects into a vector space. Depending on the desired use case for the embedding, these spaces are optimized to encode different properties and relations between the objects. Although being used for some time, recent research interest has been sparked by positive results that such embeddings have yielded in fields such as image retrieval, biology, medicine, and natural language processing [4, 12].

Despite their popularity, structures captured by vector spaces are often not very well understood [7]. In addition, the non-deterministic nature of many embedding algorithms, and their dependence on critical input parameters (such as the dimensionality of the vector space), can lead to embeddings with differing properties even for the same input data set. There are few tools to help understand and compare such spaces. Most of them rely on projections into 2D space to convey an impression of vector similarities rather than addressing concrete and practical tasks. Such tools are not helpful for understanding and comparing word vector spaces.

In this ongoing work, we review relevant natural language processing (NLP) literature and collect meaningful properties of vector spaces that are used to evaluate or test them. Based on these, we derive six basic tasks that facilitate understanding of these properties. We then propose three design prototypes that help to gain insights into various word vector embeddings.

2 Background

Recent developments in embedding methods have largely been inspired by progress with neural networks [9]. These structures naturally lead to the creation of vector representations for input objects in intermediary levels of the network. There are numerous visual approaches that aim to help understand the structure of such networks, e.g., [8], but it still remains an important problem. However, gaining insights into neural networks does not necessarily help with understanding important characteristics of vector embeddings. In addition, some efficient state-of-the-art embedding methods are not based on a neural networks, e.g., [11]. Word embeddings can be evaluated by measuring performance of downstream NLP tasks [10]. A popular alternative is to directly evaluate encoded linguistic relations between words through the use of specialized evaluation datasets. Some authors, e.g., [12] even let human users rate results. Analyzing and comparing statistical models of language and other types of data is an important problem that has been addressed by visualization research before. Alexander and Gleicher [1], for example, create visual designs in a task-based manner that allow users to gain insights into and compare topic modeling results.

3 Properties and Tasks

We have reviewed NLP literature concerned with word vector embeddings and their evaluation. While some of the evaluations were based on downstream methods, we collected those that directly evaluate vector relations based on linguistic tasks. Table 1 lists three embedding properties and six tasks that can be facilitated by interactive visualization. We derive tasks that deal with single embeddings (abbreviated "Emb." in Table 1), and for comparing properties between multiple embeddings. Table 2 lists linguistic goals, and shows how they can be broken down into the three basic properties.

Property	**Single Emb. Task**	**Multiple Emb. Task**
Nearest neighbors (*nn*)	Understand structure of neighborhood	Compare neighborhoods
Combination (*comb*)	Understand vector combinations	Compare vector combinations
Axis alignment (*axis*)	Understand concept relations	Compare concept relations

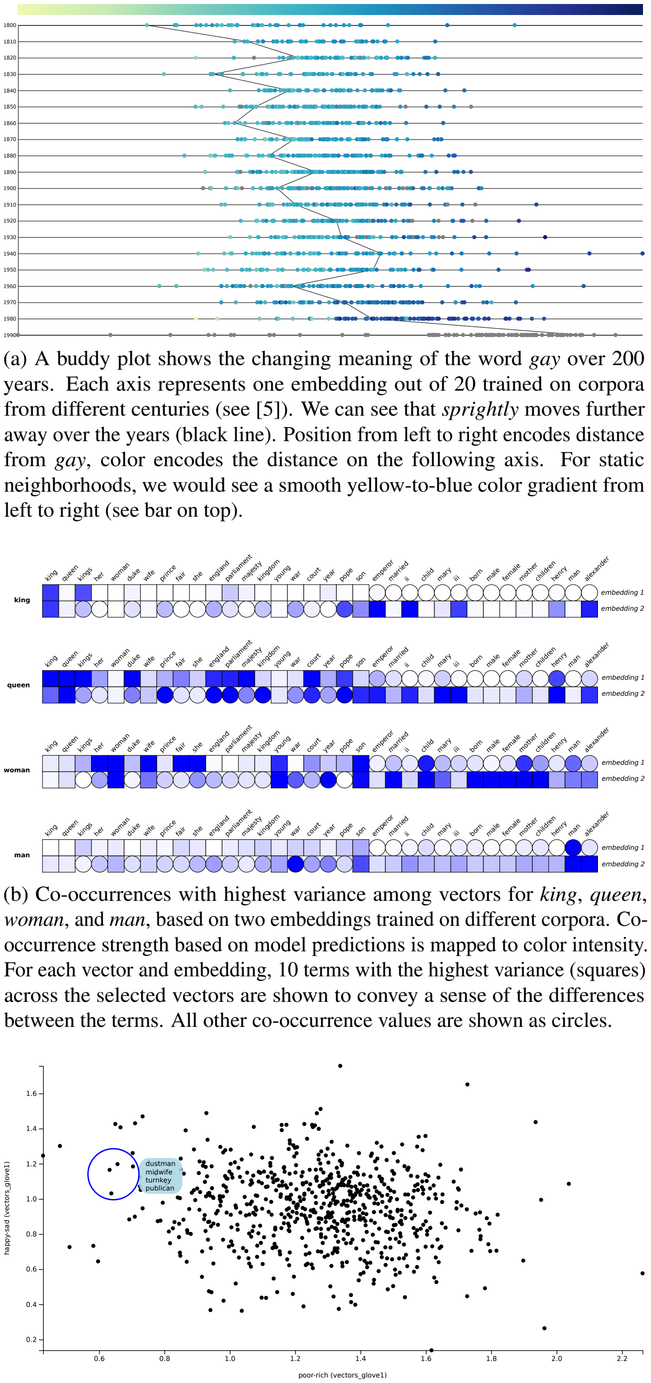
Table 1: Six basic tasks that can help users understand and compare word vector embeddings for different linguistic goals.

Linguistic goal	**Properties**	**Examples**
Word analogy	*comb*, *nn*	[2, 3, 7, 9, 12]
Word fields	*comb*, *nn*	[4]
Semantic relatedness	*nn* (ranked)	[2, 12]
Synonymy detection	*nn* (candidates)	[2]
Concept categorization	*nn* (candidates)	[2, 12]
Selectional preferences	*comb*, *nn*	[2, 12]
Concept axis	*axis*	[3]
Changes in meaning	*nn* (over time)	[5, 6]

Table 2: Linguistic tasks from NLP literature and word embedding properties relevant for them. *nn* is listed with three variations, *ranks*, *candidates*, and *over time*. These target neighborhood relations in a ranked fashion, based on candidate expressions, and comparing neighborhoods over time, respectively. All of them are supported by our designs.

[*] e-mail: heimerl@cs.wisc.edu

[†] e-mail: gleicher@cs.wisc.edu



(a) A buddy plot shows the changing meaning of the word *gay* over 200 years. Each axis represents one embedding out of 20 trained on corpora from different centuries (see [5]). We can see that *sprightly* moves further away over the years (black line). Position from left to right encodes distance from *gay*, color encodes the distance on the following axis. For static neighborhoods, we would see a smooth yellow-to-blue color gradient from left to right (see bar on top).

(b) Co-occurrences with highest variance among vectors for *king*, *queen*, *woman*, and *man*, based on two embeddings trained on different corpora. Co-occurrence strength based on model predictions is mapped to color intensity. For each vector and embedding, 10 terms with the highest variance (squares) across the selected vectors are shown to convey a sense of the differences between the terms. All other co-occurrence values are shown as circles.

(c) Historic occupations projected along two axes, *happy - sad* and *poor - rich*, from an embedding trained on historic texts. Users can explore the space with an interactive lens that summarizes information of words under it.

Figure 1: Three designs to support the tasks in Table 1.

4 Designs

In this section, we list three proposed designs to support the tasks from Table 1. Figure 1 shows examples for them. Figure 1a is a buddy plot design [1] that supports the analysis of nearest neighbors (nn). It allows exploration of neighborhood structures for different words from within the same embedding, or for the same word across different embeddings. Buddy plots also support the analysis of vector combinations based on their local neighborhoods. It enables users to find answers to analysis questions such as how the neighborhood of the vector *queen* changes, if they subtract *woman* from it.

In addition, the original context of the words and their variation across combinations are an important property to understand the meaning of vector operations in relation to the base corpus. To provide users with insight into variances across word vectors, we have created a design that conveys them for each vector involved in a combination based on one or multiple embeddings (see Figure 1b).

The final task we support is to explore projections to axes spanned by vectors that represent user-defined concepts. We have chosen to use scatter plots for this (see [6]) as they convey general distributions, and show correlations between both axes. Scatter plots are flexible to support the comparison of terms mapped to identical axes in two different embeddings, as well as different axes in the same embedding. Figure 1c shows an example of the latter.

5 Discussion and Conclusion

Currently, the three designs are implemented individually and lack interaction necessary for practical analysis. In addition to seamless switching between them, users should have access to examples of the base text, e.g., to interpret co-occurrences within their original contexts. Moreover, the scatter plot design does not scale to very large numbers of words. This can either be achieved by restricting it to word fields (see [4]), or by using suitable aggregation methods. So far, we have focused on word embeddings. However, our designs are extendable to other data types, such as medical data. We plan to explore these domains and adapt our designs accordingly.

To summarize, we have identified tasks to better understand word vector embeddings and propose three designs to support them. In the future, we plan to extend our designs and eventually make them available, embedded within a more comprehensive analysis framework.

Acknowledgments

This work was in part supported by NSF award 1162037.

References

[1] E. Alexander and M. Gleicher. Task-driven comparison of topic models. *IEEE TVCG*, 22(1):320–329, January 2016.

[2] M. Baroni, G. Dinu, and G. Kruszewski. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proc. of ACL*, pages 238–247, 2014.

[3] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *NIPS*, pages 4349–4357, 2016.

[4] E. Fast, B. Chen, and M. S. Bernstein. Empath: Understanding topic signals in large-scale text. In *Proc. of CHI*, pages 4647–4657, 2016.

[5] W. L. Hamilton, J. Leskovec, and D. Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In *Proc. of ACL*, pages 1489–1501, 2016.

[6] A. Jatowt and K. Duh. A framework for analyzing semantic change of words across time. In *Proc. of JCDL*, pages 229–238, 2014.

[7] O. Levy and Y. Goldberg. Linguistic regularities in sparse and explicit word representations. In *Proc. of CoNLL*, pages 171–180, 2014.

[8] M. Liu, J. Shi, Z. Li, C. Li, J. Zhu, and S. Liu. Towards better analysis of deep convolutional neural networks. *IEEE TVCG*, 23(1):91–100, January 2017.

[9] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Proc. of NIPS*, pages 3111–3119, 2013.

[10] N. Nayak, G. Angeli, and C. D. Manning. Evaluating word embeddings using a representative suite of practical tasks. In *Proc. of the ACL Workshop on Evaluating Vector-Space Representations for NLP*, pages 31–35, 2016.

[11] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Proc. of EMNLP*, pages 1532–1543, 2014.

[12] T. Schnabel, I. Labutov, D. Mimno, and T. Joachims. Evaluation methods for unsupervised word embeddings. In *Proc. of EMNLP*, pages 298–307, 2015.