Exploring potential gender stereotypes in the distributional semantics of child-directed speech

Benjamin E. deMayo (bdemayo@princeton.edu)
Princeton University

Abstract

Abstract: In three analyses, I explore whether gender stereotypes might be present in the distributional semantics of the CHILDES corpus (MacWhinney 2014), a large compendium of transcribed conversations between caregivers and their children, by training 2 commonly-used word embedding models on the corpus. In the first analysis, I show that word vector representations generated by both models capture some information about gender that is correlated with human judgements of the genderedness of individual words. In the second analysis, I show that this relationship is consistent in speech directed at both boys and girls and across the developmental period spanned by the children in CHILDES. In the third analysis, I examine whether specific stereotypical associations with gender are detectable in the vector space representation of the words.

Keywords: Add your choice of indexing terms or keywords; kindly use a semi-colon; between each term.

Introduction

Gender is a highly salient social category that develops within the first few years of life and maintains its importance across the lifespan (Ruble, Martin, & Berenbaum, 2006). Gender stereotypes, which can be thought of as characteristics that are believed to be true of a gender category as a whole, have their origins in toddlerhood, but become more rigid in the preschool years into middle childhood (Halim & Ruble, 2010). How children form concepts of gender, as well as the stereotypes that are linked to those concepts, has long been a subject of research, with some researchers proposing that language input to children could have a consequential impact on children's conceptualizations of gender categories.

Broadly speaking, two theoretical approaches have attempted to explain how language input to children might shape their gender concepts. One approach has emphasized the communication of knowledge from adults to children in a "top-down" fashion, in which children hear statements that explicitly communicate information about groups, such as generic statements (e.g., "girls are good at reading"; Gelman, Ware, & Kleinberg, 2010). Another approach, which I focus on here, emphasizes how children could pick up on subtle cues about gender concepts and stereotypes from the statistics of their language input in a "bottom-up" fashion. In other words, children could learn that words corresponding to particular activities, traits, occupations, and other characteristics are themselves gendered by virtue of the other words with which they co-occur. This latter approach therefore shares an intimate link with the computational linguistic subfield of distributional semantics, which seeks to characterize how the meaning of linguistic items is related to how those items are distributed in large bodies of text.

Several studies have leveraged the tools of distributional semantics to examine whether gender stereotypes are appreciable natural language corpora; some of these studies focus specifically on language that would likely be heard by children. The general strategy used by these studies has involved taking large bodies of text (usually those with several million tokens, though this has not always been the case Lewis, Borkenhagen, Converse, Lupyan, & Seidenberg, 2020) and using them to train a word embedding models, which generate representations of individual word types in a high-dimensional vector space based on each word type's co-occurrence with other types. The key assumption in such a strategy is that words that frequently co-occur will have similar meanings. Once vector representations of words are obtained, cosine distances between individual lexical items in the vector space are calculated as a proxy of semantic similarity, allowing researchers to examine whether words' vector representations show patterns of similarity to other words that might be expected given prevalent societal stereotypes (e.g., that the word "doll" is closer to the word "girl" than it is to the word "boy"). This general analytic framework has been used to argue that gender stereotypes are present in the distributional structure of large bodies of naturalistic text, including web-based corpora, children's books, and transcripts of films and television shows (Caliskan, Bryson, & Narayanan, 2017; Charlesworth, Yang, Mann, Kurdi, & Banaji, 2021; Lewis & Lupyan, 2020).

In this work, I extend prior findings by examining the human-like gender-stereotypical biases that might emerge in the vector representations obtained from training word embedding models on a body of child-directed speech. Specifically, I use transcripts between caregivers and children between the ages of 1 and 3 years old from the North American English corpora in the Child Language Data Exchange System [CHILDES; MacWhinney (2009)] and extract vector-space semantic representations using 2 commonly-used word embedding models, word2vec (Mikolov, Chen, Corrado, & Dean, 2013) and GloVe (Pennington, Socher, & Manning, 2014).

Method

Data preprocessing

Child-directed language was sourced from all of the North American English transcripts in the CHILDES corpus (MacWhinney, 2009). Transcripts were obtained using the `childes-db` API, which allows researchers to access transcript utterances in a tabular format that includes metadata about each utterance, including its speaker's role (parent, grandparent, child, etc.), the gender of the child in the conversation, and the lemmatized "stem" of the utterance (Sanchez et al., 2019). From this tabular data, the stems of utterances from mothers, fathers, grandparents and adults were extracted and concatenated to create the training data, which contained X conversations from Y dyads including Z children, and was comprised of A word tokens and B word types.

Word embedding model training

Two common word embedding models were used to obtain vector-space representations for words in CHILDES. The first was Word2Vec (Mikolov et al., 2013), which uses a 2-layer neural network to predict a given word in a sentence given its surrounding words (continuous bag of words approach, CBOW) or vice versa (skip-gram approach) and derives vector-space representations of each word based on the neural network weights between the input layer and the single hidden layer of the network. The second was GloVe (Pennington et al., 2014), an unsupervised learning algorithm which takes as input a sparse matrix encoding the co-occurrence frequency of each pair of lexical items in a corpus, and which learns vector representations for these items, such that the inner product between two vectors closely approximates a logarithmic transformation of the probability that those two lexical items co-occur in the text. For our purposes here, the two techniques have the same goal of extracting vector representations of words that are semantically meaningful, even though GloVe's learning strategy emphasizes co-occurrence probability between pairs of words more than Word2Vec, which centers more on the semantic contexts that words appear in. In the following analyses, the context window of each word embedding model is set to 5 words in both directions from a target word. Word representations derived from GloVe are vectors in a X-dimensional space and those from Word2Vec are in a Y-dimensional space.

Analyses

Analysis 1: Broad comparison between Word2Vec and GloVe

The first set of analyses is meant a coarse indication of whether Word2Vec and GloVe are capturing roughly similar semantic information for words in the CHILDES corpus, particularly as it concerns individual words' genderedness.

Analysis 1A: Association between genderedness of word representations in Word2Vec and GloVe In Analysis 1A, I examine whether the genderedness of a word's representation in Word2Vec is associated with the genderedness of a the same word's GloVe vector representation. A word's *genderedness* is operationalized here as the average cosine distance between the word's vector representation and the vector representations of each of a set of "anchor words" which correspond to the concept of "boy" or "girl." More precisely, for a given word w and sets of anchor words G and B for the concepts *girl* and *boy*, respectively:

$$G = \{\text{girl, woman, sister, she, her, daughter}\}$$
$$B = \{\text{boy, man, brother, he, him, son}\}$$
$$s(w,G) = \text{mean}_{g\in G}(\cos(\vec{w},\vec{g}))$$
$$s(w,B) = \text{mean}_{b\in B}(\cos(\vec{w},\vec{b}))$$

First-Level Headings

First level headings should be in 12 point , initial caps, bold and centered. Leave one line space above the heading and 1/4˜line space below the heading.

Second-Level Headings

Second level headings should be 11 point , initial caps, bold, and flush left. Leave one line space above the heading and 1/4˜ line space below the heading.

Third-Level Headings Third-level headings should be 10 point , initial caps, bold, and flush left. Leave one line space above the heading, but no space after the heading.

Formalities, Footnotes, and Floats

Use standard APA citation format. Citations within the text should include the author's last name and year. If the authors' names are included in the sentence, place only the year in parentheses, as in (1972), but otherwise place the entire reference in parentheses with the authors and year separated by a comma (Newell & Simon, 1972). List multiple references alphabetically and separate them by semicolons (Chalnick & Billman, 1988; Newell & Simon, 1972). Use the et. al. construction only after listing all the authors to a publication in an earlier reference and for citations with four or more authors.

For more information on citations in RMarkdown, see **here.**

Footnotes

Indicate footnotes with a number[1] in the text. Place the footnotes in 9 point type at the bottom of the page on which they appear. Precede the footnote with a horizontal rule.[2] You can also use markdown formatting to include footnotes using this syntax.[3]

Figures

All artwork must be very dark for purposes of reproduction and should not be hand drawn. Number figures sequentially,

[1] Sample of the first footnote.
[2] Sample of the second footnote.
[3] Sample of a markdown footnote.

placing the figure number and caption, in 10 point, after the figure with one line space above the caption and one line space below it. If necessary, leave extra white space at the bottom of the page to avoid splitting the figure and figure caption. You may float figures to the top or bottom of a column, or set wide figures across both columns.

Two-column images

You can read local images using png package for example and plot it like a regular plot using grid.raster from the grid package. With this method you have full control of the size of your image. **Note: Image must be in .png file format for the readPNG function to work.**

You might want to display a wide figure across both columns. To do this, you change the `fig.env` chunk option to `figure*`. To align the image in the center of the page, set `fig.align` option to `center`. To format the width of your caption text, you set the `num.cols.cap` option to `2`.

One-column images

Single column is the default option, but if you want set it explicitly, set `fig.env` to `figure`. Notice that the `num.cols` option for the caption width is set to `1`.

Figure 2: One column image.

R Plots

You can use R chunks directly to plot graphs. And you can use latex floats in the fig.pos chunk option to have more control over the location of your plot on the page. For more information on latex placement specifiers see **here**

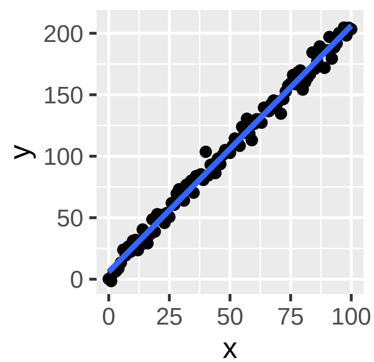

Figure 3: R plot

Tables

Number tables consecutively; place the table number and title (in 10 point) above the table with one line space above the caption and one line space below it, as in Table 1. You may float tables to the top or bottom of a column, set wide tables across both columns.

You can use the xtable function in the xtable package.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.09	0.11	0.8	0.43
x	2.06	0.10	20.4	0.00

Table 1: This table prints across one column.

Acknowledgements

Place acknowledgments (including funding information) in a section at the end of the paper.

References

10 Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. In *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510–516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Charlesworth, T. E., Yang, V., Mann, T. C., Kurdi, B., & Banaji, M. R. (2021). Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. *Psychological Science*, *32*(2), 218–240.

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, *61*(3), 273–301.

Halim, M. L., & Ruble, D. (2010). Gender identity and stereotyping in early and middle childhood. In *Handbook of gender research in psychology* (pp. 495–525). Springer.

Lewis, M., Borkenhagen, M. C., Converse, E., Lupyan, G., & Seidenberg, M. S. (2020). What might books be teaching young children about gender?

Figure 1: This image spans both columns. And the caption text is limited to 0.8 of the width of the document.

Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour*, *4*(10), 1021–1028.

MacWhinney, B. (2009). The CHILDES project. *Tools for Analyzing Talk–Electronic Edition*, *2*.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv Preprint arXiv:1301.3781*.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Retrieved from `http://www.aclweb.org/anthology/D14-1162`

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development.

Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2019). Childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior Research Methods*, *51*(4), 1928–1941.