

Man is to computer programmer as woman is to homemaker? debiasing word embeddings

T Bolukbasi, KW Chang, JY Zou... - *Advances in neural ...*, 2016 - papers.nips.cc

... By retaining the gender component for gender-specific **words**, we maintain analogies such as ... is increasingly reliant on **word embeddings**, in a small way debiased **word embeddings** can hopefully ... applications, one might argue that gender **biases** in the **embedding** (eg computer ...

☆ [🔗](#) Cited by 277 [Related articles](#) [All 9 versions](#) [🔗](#)

Semantics derived automatically from language corpora contain human-like biases

A Caliskan, JJ Bryson, A Narayanan - *Science*, 2017 - science.sciencemag.org

... We developed our first method, the **Word-Embedding Association Test (WEAT)**, a statistical test analogous to the IAT, and applied it to a widely used semantic representation of **words** in AI, termed **word embeddings** ...

☆ [🔗](#) Cited by 323 [Related articles](#) [All 22 versions](#)

Diachronic word embeddings reveal statistical laws of semantic change

WL Hamilton, J Leskovec, D Jurafsky - *arXiv preprint arXiv:1605.09096*, 2016 - arxiv.org

... 8In fact, the SGNS **embeddings** may even be **biased** in the other direction, since higher frequency **words** undergo more SGD updates “away” from this initialization. where $z_{wi} \sim N(0, \sigma_{wi})$ is the random intercept for **word** w_i and $\epsilon(t)_{wi} \in N(0, \sigma)$ is an error term ...

☆ [🔗](#) Cited by 203 [Related articles](#) [All 18 versions](#) [🔗](#)

Word embeddings quantify 100 years of gender and ethnic stereotypes

N Garg, L Schiebinger, D Jurafsky... - *Proceedings of the ...*, 2018 - *National Acad Sciences*

... **Word embeddings** quantify 100 years of gender and ethnic stereotypes. Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. PNAS April 17, 2018 115 (16) E3635-E3644; published ahead of print April 3, 2018 <https://doi.org/10.1073/pnas.1720347115> ...

☆ [🔗](#) Cited by 45 [Related articles](#) [All 10 versions](#)

Improving distributional similarity with lessons learned from word embeddings

O Levy, Y Goldberg, I Dagan - *Transactions of the Association for ...*, 2015 - *MIT Press*

... using windows of 2, 5, and 10 tokens to each side of the focus **word** (the window ... times in the corpus were ig- nored, resulting in vocabularies of 189,533 terms for both **words** and contexts ... Training **Embeddings** We trained a 500- dimensional representation with SVD, SGNS, and ...

☆ [🔗](#) Cited by 684 [Related articles](#) [All 12 versions](#)

Incorporating structural alignment biases into an attentional neural translation model

T Cohn, CDV Hoang, E Vymolova, K Yao... - *arXiv preprint arXiv ...*, 2016 - arxiv.org

... garbage collection prob- lem. Conversely **words** that tend to need high fertil- ity can be learned through the interactions between these features and the **word** and context **embeddings** in Eq 7. Global fertility A second ...

☆ [🔗](#) Cited by 87 [Related articles](#) [All 10 versions](#) [🔗](#)

Quantifying and reducing stereotypes in word embeddings

T Bolukbasi, KW Chang, J Zou, V Saligrama... - *arXiv preprint arXiv ...*, 2016 - arxiv.org

... that a LinkedIn web page containing terms such as “PhD student,” “**embed- dings**,” and “deep ... propose two methods to systematically quantify the gender **bias** in a set of **word embeddings**. First, we quantify how **words**, such as those corresponding to professions, are distributed ...

☆ [🔗](#) Cited by 17 [Related articles](#) [All 5 versions](#) [🔗](#)

Mitigating unwanted biases with adversarial learning

BH Zhang, B Lemoine, M Mitchell - *Proceedings of the 2018 AAAI/ACM ...*, 2018 - dl.acm.org

... let $x = (x_1, x_2, x_3) \in \mathbb{R}^{3d}$ comprise the **word** vectors for the first three **words**, y be ... nor is there projection onto the **bias** subspace and therefore the original **word embeddings** are never ... What is learned is a transform from a **biased embedding** space to a debiased **embedding** space ...

☆ [🔗](#) Cited by 38 [Related articles](#) [All 8 versions](#)

Evaluation methods for unsupervised word embeddings

T Schnabel, I Labutov, D Mimno... - *Proceedings of the 2015 ...*, 2015 - aclweb.org

... We used length-normalized em- beddings, as rare **words** might have shorter vec- tors ... five-fold cross- validation at each threshold frequency in Figure 3. All **word embeddings** do better ... Unlike all other **embeddings**, accuracy for C&W **embed- dings** increases for larger threshold ...

☆ [🔗](#) Cited by 225 [Related articles](#) [All 11 versions](#) [🔗](#)

Biased graph walks for RDF graph embeddings

M Cochez, P Ristoski, SP Ponzetto... - Proceedings of the 7th ..., 2017 - dl.acm.org

... 4 APPROACH In our approach, we adapt neural language models for RDF graph **embeddings** ...
word order in text documents, explicitly modeling the assumption that closer **words** in the ... RDF
graphs, we consider entities and relations between entities instead of **word** sequences ...

☆  Cited by 16 Related articles All 6 versions

Related searches

evaluation of word embeddings

origins of bias word embeddings

diachronic word embeddings

unsupervised word embeddings

computer programmer word embeddings

ethnic stereotypes word embeddings

distributional similarity word embeddings

practical tasks word embeddings

statistical laws word embeddings

semantic change word embeddings

lexical decision word embeddings

reaction times word embeddings

cognitive processes word embeddings