Modeling **word** meaning in context with substitute **vectors**

O Melamud, I Dagan, J Goldberger - … of the 2015 Conference of the …, 2015 - aclweb.org

… **word** slot, as reflected in the respective context substitute vector representa- tion, denoted Scond,1000. Finally, we rank the candidates using the scores in our in-context paraphrase **vectors** from Section 6.2. However, this time we check the effect of injecting a stronger **bias** …

☆ Cited by 23 Related articles All 8 versions »

Not all contexts are created equal: Better **word** representations with variable attention

W Ling, Y Tsvetkov, S Amir, R Fermandez… - Proceedings of the …, 2015 - aclweb.org

… Computing the attention of all words in the input requires 2b operations, as it simply requires re- trieving one value from the lookup matrix K for each **word** and one value from the **bias** s for each **word** in the window … 3 Experiments 3.1 **Word Vectors** …

☆ Cited by 67 Related articles All 13 versions »

Large-scale learning of **word** relatedness with constraints

G Halawi, G Dror, E Gabrilovich, Y Koren - Proceedings of the 18th ACM …, 2012 - dl.acm.org

… vice" can occur with many other words (eg, "vice president"), hence its **bias** is much … shares resemblance with LDA in terms of modeling the words through latent factor **vectors**. However, CLEAR dif- fers in modeling short **word** sequences, or sentences, whereas LDA requires …

☆ Cited by 115 Related articles All 9 versions

Text classification and minimal-**bias** training **vectors**

K Ahmad, TA Bale, D Burford - IJCNN'99. International Joint …, 1999 - ieeexplore.ieee.org

… question in neural network research (Amari, [2]). In order to minimise training **bias** we have … of texts under examination, and is thus a good candidate for a domain feature **word** … Binary training **vectors** representing the news items were then created according to the presence or …

☆ Cited by 5 Related articles All 2 versions

Effective use of **word** order for text categorization with convolutional neural networks

R Johnson, T Zhang - arXiv preprint arXiv:1412.1058, 2014 - arxiv.org

… Each computation unit (oval) computes a non-linear function σ(W·rl(x)+b) of a small region rl(x) of input image x, where weight matrix W and **bias** vector b are shared by all the units in the … Kim (2014) studied fine-tuning of pre-trained **word vectors** to produce input to parallel CNN …

☆ Cited by 413 Related articles All 6 versions »

Gated **word**-character recurrent language model

Y Miyamoto, K Cho - arXiv preprint arXiv:1606.01700, 2016 - arxiv.org

… Rd is the vector representation of the **word** wt using a character input. The generated **vectors** xword wt and xchar wt are mixed by a gate gwt as gwt = σ (vg xword wt + bg) xwt = (1 − gwt)xword wt + gwt xchar wt , (3) where vg ∈ Rd is a weight vector, bg ∈ R is a **bias** scalar, σ …

☆ Cited by 46 Related articles All 5 versions »

Bilbowa: Fast bilingual distributed representations without **word** alignments

S Gouws, Y Bengio, G Corrado - 2015 - jmlr.org

… in §2, the primary challenges with existing bilingual embedding models are their computational com- plexity (due to an expensive softmax or an expensive reg- ularization term, or both), but most importantly, the strong domain **bias** that is … The **word**-count **vectors** are then …

☆ Cited by 202 Related articles All 16 versions »

Sequential short-text classification with recurrent and convolutional neural networks

JY Lee, F Dernoncourt - arXiv preprint arXiv:1603.03827, 2016 - arxiv.org

… At each gradient descent step, weight matrices, **bias vectors**, and **word vectors** are updated. For regular- ization, dropout is applied after the pooling layer, and early stopping is used on the validation set with a patience of 10 epochs. 4 Results and Discussion …

☆ Cited by 137 Related articles All 10 versions »

Predicting polarities of tweets by composing **word** embeddings with long short-term memory

X Wang, Y Liu, SUN Chengjie, B Wang… - Proceedings of the 53rd …, 2015 - aclweb.org

… 2011). These works re- veal that social network relations of opinion hold- ers could bring an influential **bias** to the textual models … model. Such **word vectors** ease the syntactic and semantic sparsity of bag-of-words representations …

☆ Cited by 99 Related articles All 4 versions »

Neural machine translation by jointly learning to align and translate

D Bahdanau, K Cho, Y Bengio - arXiv preprint arXiv:1409.0473, 2014 - arxiv.org

… The model then predicts a target **word** based on the context **vectors** associated with these source positions and all the previous generated target words. *CIFAR Senior Fellow 1 arXiv:1409.0473 v7 [cs.CL] 19 May 2016 Page 2. Published as a conference paper at ICLR 2015 …

  Cited by 6614 Related articles All 20 versions