## A latent variable model approach to pmi-based **word** embeddings

S Arora, Y Li, Y Liang, T Ma, A Risteski - Transactions of the Association …, 2016 - MIT Press

… as a random walk on words, which are assumed to be embedded as **vectors** in a … interpretation
and has proven useful in subsequent work, eg understand- ing structure of **word** embeddings
for … Also our work clarifies some weighting and **bias** terms in the training objec- tives of …

☆  ⁇  Cited by 64    Related articles    All 6 versions

## Adam: A method for stochastic optimization

DP Kingma, J Ba - arXiv preprint arXiv:1412.6980, 2014 - arxiv.org

… Good default settings for the tested machine learning problems are α = 0.002, β1 = 0.9 and β2 =
0.999. With βt 1 we denote β1 to the power t. Here, (α/(1 − βt 1)) is the learning rate with the
**bias**-correction term for the first moment. All operations on **vectors** are element-wise …

☆  ⁇  Cited by 20744    Related articles    All 17 versions    »

## Distributional **vectors** encode referential attributes

A Gupta, G Boleda, M Baroni, S Padó - Proceedings of the 2015 …, 2015 - aclweb.org

… is present in text, these representations are able to capture many aspects of **word** meaning … We
show that a simple supervised algo- rithm applied to **vectors** can retrieve them so … of the conceptual
nature of the knowledge encoded in dis- tributional data, that **bias** their predictions …

☆  ⁇  Cited by 25    Related articles    All 11 versions    »

## A unified learning framework of skip-grams and global **vectors**

J Suzuki, M Nagata - Proceedings of the 53rd Annual Meeting of the …, 2015 - aclweb.org

… A comparison with the results of the second row, finds no mean- ingful benefit to using the **bias**
terms. In contrast, obviously, the elapsed time for model learning is consistently shorter since
we can discard the **bias** term update … Glove: Global **vectors** for **word** representation …

☆  ⁇  Cited by 10    Related articles    All 5 versions    »

## Distributed **word** representation learning for cross-lingual dependency parsing

M Xiao, Y Guo - Proceedings of the Eighteenth Conference on …, 2014 - aclweb.org

… J(D) = 1 N N ∑ i=1 max(0,1 − s(xi) + s(xi)) (5) We perform a random initialization over the
look-up table and weight model parameters, and set all the **bias** model parameters to zeros …
training effort. The dimension k of the embedding **word vectors** in R is set as 200 …

☆  ⁇  Cited by 48    Related articles    All 12 versions    »

## Deep visual-semantic alignments for generating image descriptions

A Karpathy, L Fei-Fei - Proceedings of the IEEE conference on …, 2015 - cv-foundation.org

… We condition the RNN's predictions on the image informa- tion (bv) via **bias** interactions on the
first step … Since we use dif- ferent **word vectors**, dropout for regularization and different
cross-validation ranges and larger embedding sizes, we re- implemented their loss for a fair …

☆  ⁇  Cited by 2308    Related articles    All 27 versions    »

## Adaptive decision directed speech recognition **bias** equalization method and apparatus

BH Juang, D Mansour, JG Wilpon - US Patent 5,812,972, 1998 - Google Patents

… sequence or most likely HMM for the sequence of feature **vectors**. Each most likely state sequence
preferably represents a **word** model. The recognizer 220 employs a novel segmentation technique
that includes adaptive equalization to compensate for signal **bias** caused by …

☆  ⁇  Cited by 47    Related articles    All 2 versions    »

## Sparse overcomplete **word** vector representations

M Faruqui, Y Tsvetkov, D Yogatama, C Dyer… - arXiv preprint arXiv …, 2015 - arxiv.org

… Our work builds on recent explorations of spar- sity as a useful form of inductive **bias** in NLP and
machine learning more broadly … di- mensions has been shown to improve dimension interpretability
(Murphy et al., 2012; Fyshe et al., 2014) and usability of **word vectors** as features …

☆  ⁇  Cited by 74    Related articles    All 9 versions    »

## Ask, attend and answer: Exploring question-guided spatial attention for visual question answering

H Xu, K Saenko - European Conference on Computer Vision, 2016 - Springer

… 30]. Specifically, we compute $$\begin{aligned} Q = W\_Q \cdot V + b\_Q \end{aligned}$$.
(4). where \(W\_Q \in \mathbb {R}^T\) represents the BOW weights for **word vectors**
V, and \(b\_Q \in \mathbb {R}^{N}\) is the **bias** term. The …

☆  ⁇  Cited by 319    Related articles    All 5 versions

## Evaluating **word** embeddings using a representative suite of practical tasks

N Nayak, G Angeli, CD Manning - Proceedings of the 1st Workshop on …, 2016 - aclweb.org

… 7.2 Avoiding **bias** Since this method of evaluation involves training a number of neural network models, there is a sig- nificant danger of overfitting to the embeddings used to find the hyperparameters … 2014. Community evaluation and exchange of **word vectors** at **word**- **vectors** …

☆ 〃 Cited by 30 Related articles All 10 versions ≫

## Evaluating **word** embeddings using a representative suite of practical tasks

N Nayak, G Angeli, CD Manning - Proceedings of the 1st Workshop on …, 2016 - aclweb.org

… 7.2 Avoiding **bias** Since this method of evaluation involves training a number of neural network models, there is a sig- nificant danger of overfitting to the embeddings used to find the hyperparameters … 2014. Community evaluation and exchange of **word vectors** at **word**- **vectors** …

☆ 〃 Cited by 30 Related articles All 10 versions ≫