# Unsupervised Post-processing of Word Vectors via Conceptor Negation


**Tianlin Liu*, Lyle Ungar♯, João Sedoc♯**

*Department of Computer Science and Electrical Engineering,
Jacobs University Bremen, Bremen, Germany.
♯Department of Computer and Information Science,
University of Pennsylvania, Philadelphia, PA, USA.

t.liu@jacobs-university.de, {ungar,joao}@cis.upenn.edu


JACOBS UNIVERSITY

Penn UNIVERSITY of PENNSYLVANIA

## Abstract


To enrich the semantic captured by word vectors, we introduce a novel word vector post-processing technique based on *matrix conceptors*, a family of regularized identity maps. More concretely, we propose to use conceptors to suppress those latent features of word vectors having high variances. The proposed method is purely unsupervised: it does not rely on any corpus or external linguistic database.


## Unsupervised word vector post-processing

Goal: Enhance semantic regularities word vectors' in a lightweight fashion.
Method: Use spectral-decomposition methods akin to SVD and PCA.
Relevant work: *all-but-the-top* (ABTT) method [1].

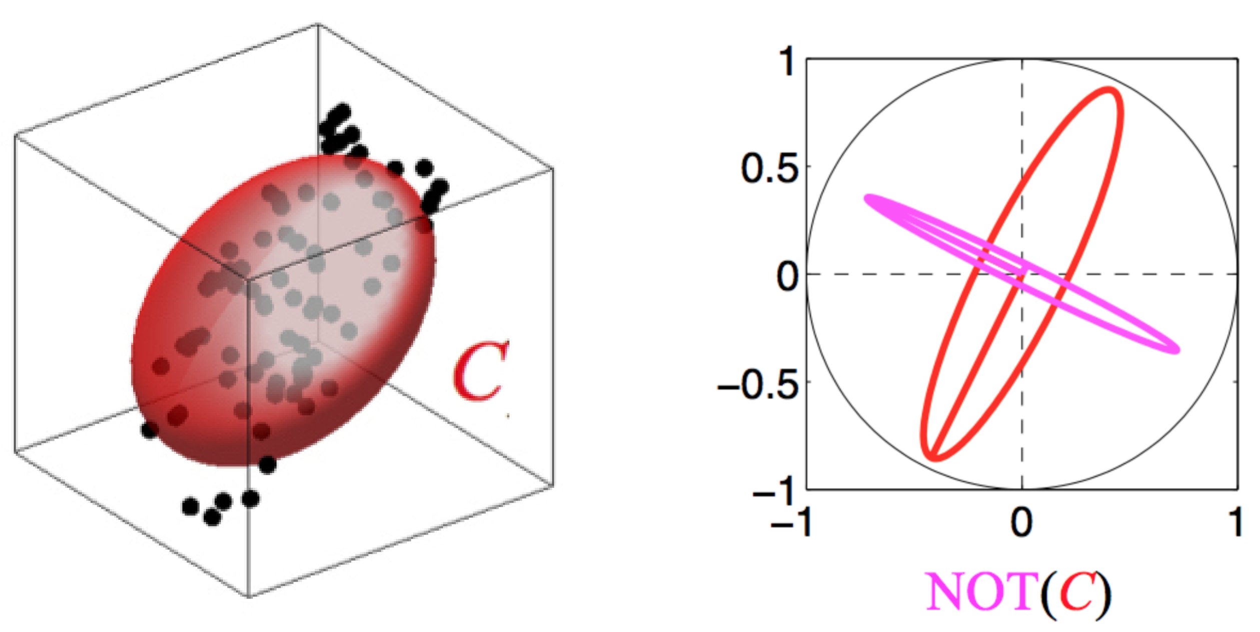
**Input:** (i) $\{v_w \in \mathbb{R}^n : w \in V\}$: word vectors with a vocabulary $V$;
(ii) $d$: the number of PCs to be removed.
**Step 1:** Center the word vectors: Let $\bar{v}_w := v_w - \mu$ for all $w \in V$, where $\mu$ is the mean of the input word vectors.
**Step 2:** Compute the first $d$ PCs $\{u_i \in \mathbb{R}^n\}_{i \in [d]}$ of the column-wisely stacked centered word vectors $[\bar{v}_w]_{w \in V} \in \mathbb{R}^{n \times |V|}$ via a PCA.
**Step 3:** Process the word vectors: $\tilde{v}_w^{\text{ABTT}} := \bar{v}_w - \sum_{i=1}^{d} u_i^\top u_i \bar{v}_w, \forall w \in V$.
**Output:** $\{\tilde{v}_w^{\text{ABTT}}, w \in V\}$.

**Algorithm 1:** *The all-but-the-top (ABTT) algorithm for word vector post-processing.*

Downside of ABTT: ABTT either **completely removes** the a latent feature (taking form as a PC of the word vectors), or **keeps it intact**.

Our improvement over ABTT: **softly** gate away variances explained by the leading PCs of word vectors using conceptor matrices [2].

## Conceptors




(a) (b)

**Figure 1:** *(a): The red conceptor characterizes linear subspace occupied by point clouds. (b): NOT operation on conceptors.*

The conceptor, $C$, is a soft projection matrix on the linear subspace where the samples of $x$ lie. From a set of $n$-dimensional points $\{x_i\}_{i \in I}$, $C$ is defined as:

$$\underset{C}{\text{argmin}} \frac{1}{|I|} \sum_{i \in I} ||x_i - Cx_i||^2 + \alpha^{-2} ||C||_{\text{F}}^2 \quad (1)$$

where $\alpha$ is a hyperparameter and $||\cdot||_{\text{F}}$ is the Frobenius norm, $|I|$ is the cardinality of $I$. This optimization problem has a closed-form solution

$$C = R(R + \alpha^{-2}\mathbf{I})^{-1} \quad \text{where} \quad R = \frac{1}{|I|} XX^\top \quad (2)$$

where $X$ is a matrix with $x_i$ as columns, and $\mathbf{I}$ is the $n \times n$ identity matrix. Relationship between singular values of $R$ and $C$:

$$R = U \Sigma U^\top \qquad C = USU^\top$$
$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \qquad S = \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{bmatrix} \quad (3)$$
$$s_i = \sigma_i / (\sigma_i + \alpha^{-2}) \in [0, 1)$$

Boolean logic on conceptor $C$:

$$\neg C := I - C.$$

The negated conceptor, $\neg C$, softly projects the data onto a linear subspace that can be roughly understood as the orthogonal complement of the subspace characterized by $C$.

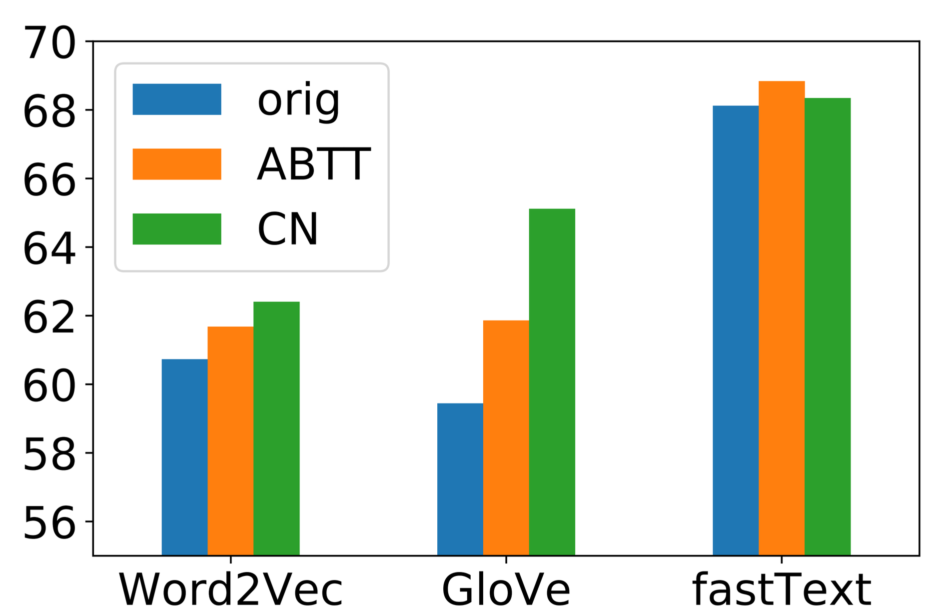
## Post-processing word vectors with Conceptor Negation

**Input:** (i) $\{v_w \in \mathbb{R}^n : w \in V\}$: word vectors of a vocabulary $V$;
(ii) $\alpha \in \mathbb{R}$: a hyper-parameter.
**Step 1**. Compute the conceptor $C$ from word vectors: $C = R(R + \alpha^{-2}\mathbf{I})^{-1}$, where $R$ is estimated by $\frac{1}{|V|}\sum_w v_w v_w^\top$
**Step 2**. Compute $\neg C := \mathbf{I} - C$
**Step 3**. Process the word vectors: $\tilde{v}_w^{\text{CN}} := \neg C v_w, \forall w \in V$
**Output:** $\{\tilde{v}_w^{\text{CN}} : w \in V\}$.

**Algorithm 2:** *The conceptor negation (CN) algorithm for word vector post-processing.*

## Experiments

### Word Similarity




*Post-processing results (Spearman's rank correlation coefficient $\times$ 100) averaged across seven word similarity benchmarks in [1].*

### Semantic Textual Similarity

| | WORD2VEC | | | GLOVE | | |
|---|---|---|---|---|---|---|
| | orig. | ABTT | CN | orig. | ABTT | CN |
| STS 2012 | 57.22 | **57.67** | 54.31 | 48.27 | 54.06 | **54.38** |
| STS 2013 | 56.81 | 57.98 | **59.17** | 44.83 | 51.71 | **55.51** |
| STS 2014 | 62.89. | 63.30 | **66.22** | 51.11 | 59.23 | **62.66** |
| STS 2015 | 62.74 | 63.35 | **67.15** | 47.23 | 57.29 | **63.74** |
| SICK | 70.10 | 70.20 | **72.71** | 65.14 | **67.85** | 66.42 |

Post-processing results ($\times$100) on the semantic textual similarity tasks.

### Concept Categorization

| | WORD2VEC | | | GLOVE | | |
|---|---|---|---|---|---|---|
| | orig. | ABTT | CN. | orig. | ABTT | CN |
| ESSLLI | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| AP | 87.28 | 88.3 | **89.31** | 86.43 | 87.19 | **90.95** |
| BM | 58.15 | 59.24 | **60.19** | 65.34 | 67.35 | **67.63** |

Purity ($\times$ 100) of the clusters in concept categorization task with fixed centroids.

### Neural Belief Tracker

A deep neural network based dialogue state tracking system [3].

| | WORD2VEC | | | GLOVE | | |
|---|---|---|---|---|---|---|
| | orig. | ABTT | CN. | orig. | ABTT | CN |
| Food | 48.6 | **84.7** | 78.5 | 86.4 | 83.7 | **88.8** |
| Price range | 90.2 | 88.1 | **92.2** | 91.0 | 93.9 | **94.7** |
| Area | 83.1 | 82.4 | **86.1** | 93.5 | **94.9** | 93.7 |
| Average | 74.0 | 85.1 | **85.6** | 90.3 | 90.8 | **92.4** |

The goal accuracy of food, price range, and area.

## References


[1] Mu, J. and Viswanath, P. All-but-the-top: Simple and effective postprocessing for word representations. In *International Conference on Learning Representations*, 2018.

[2] Jaeger, H. 2014. Controlling recurrent neural networks by conceptors. Technical report, Jacobs University Bremen.

[3] Mrkšić, N and Vulić, I. Fully statistical neural belief tracking. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 108113. Association for Computational Linguistics.


Full paper Codes


The 33rd AAAI Conference on Artificial Intelligence (AAAI), Honolulu, Hawaii, USA, 2019