Semantics- and Syntax-Related Subvectors in the Skip-Gram Embeddings (Student Abstract)

Maxat Tezekbayev, Zhenisbek Assylbekov, Rustem Takhanov
Department of Mathematics, School of Sciences and Humanities, Nazarbayev University, Nur-Sultan, Kazakhstan

Abstract
We show that the skip-gram embedding of any word can be decomposed into two subvectors which roughly correspond to semantic and syntactic roles of the word.

Introduction
Assuming that words have already been converted into indices, let \( \{1, \ldots, n\} \) be a finite vocabulary of words. Following the setups of the widely used WORD 2VEC (Mikolov et al. 2013) model, we consider two vectors per each word \( i \):
- \( w_i \) is an embedding of the word \( i \) when \( i \) is a center word,
- \( c_i \) is an embedding of the word \( i \) when \( i \) is a context word.

We follow the assumptions of Assylbekov and Takhanov (2019) on the nature of word vectors, context vectors, and text generation, i.e.
1. A priori word vectors \( w_1, \ldots, w_n \in \mathbb{R}^d \) are i.i.d. draws from isotropic multivariate Gaussian distribution: \( w_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \frac{1}{d} I) \), where \( I \) is the \( d \times d \) identity matrix.
2. Context vectors \( c_1, \ldots, c_n \) are related to word vectors according to \( c_i = Qw_i, i = 1, \ldots, n \), for some orthogonal matrix \( Q \in \mathbb{R}^{d \times d} \).
3. Given a word \( j \), the probability of any word \( i \) being in its context is given by
   \[
   p(i \mid j) \propto p_i \cdot e^{w_i^\top c_i}.
   \]  

Hypothesis. Under the assumptions 1–3 above, Assylbekov and Takhanov (2019) showed that each word’s vector \( w_i \) splits into two approximately equally-sized subvectors \( x_i \) and \( y_i \), and the model (1) for generating a word \( i \) in the context of a word \( j \) can be rewritten as
   \[
   p(i \mid j) \approx p_i \cdot e^{x_i^\top x_j - y_i^\top y_j}.
   \]

Interestingly, embeddings of the first type (\( x_i \) and \( x_j \)) are responsible for pulling the word \( i \) into the context of the word \( j \), while embeddings of the second type (\( y_i \) and \( y_j \)) are responsible for pushing the word \( i \) away from the context of the word \( j \). We hypothesize that the \( x \)-embeddings are more related to semantics, whereas the \( y \)-embeddings are more related to syntax. In what follows we provide a motivating example for this hypothesis and then empirically validate it through controlled experiments.

Motivating Example
Consider a phrase
\[ \text{the dog barking at strangers} \]

The word ‘barking’ appears in the context of the word ‘dog’ but the word vector \( w_{\text{barking}} \) is not the closest to the word vector \( w_{\text{dog}} \) (see Table 2). Instead, these vectors are split in such way that the quantity \( x_{\text{dog}}^\top x_{\text{barking}} - y_{\text{dog}}^\top y_{\text{barking}} \) is large enough. We can interpret this as follows: the word ‘barking’ is semantically close enough to the word ‘dog’ but is not the closest one: e.g. \( w_{\text{puppy}} \) is much closer to \( w_{\text{dog}} \) than \( w_{\text{barking}} \). On the other hand the word ‘barking’ syntactically fits better being next to the word ‘dog’ than ‘puppy’, i.e. \( -y_{\text{dog}}^\top y_{\text{puppy}} < -y_{\text{dog}}^\top y_{\text{barking}} \). This combination of semantic
We evaluate the whole vectors and the subvectors on tagging the Brown corpus with the Universal POS tags. The resulting accuracies are provided in Table 3. We can see that the y-subvectors are more suitable for POS-tagging than the x-subvectors, which means that the y-parts carry more syntactic information than the x-parts.

**Conclusion**

Theoretical analysis of word embeddings gives us better understanding of their properties. Moreover, theory may provide us interesting hypotheses on the nature and structure of word embeddings, and such hypotheses can be verified empirically as is done in this paper.

**Acknowledgements**

This work is supported by the Nazarbayev University Collaborative Research Program 091019CRP2109, and by the Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan, IRN AP05133700.

**References**

