# Neural Embedding Allocation:
# Distributed Representations of Words, Topics, and Documents

**Kamrun Naher Keya**
IS Department, UMBC
kkeya1@umbc.edu

**Yannis Papanikolaou**
Aristotle Univeristy of Thessaloniki
ypapanik@csd.auth.gr

**James R. Foulds**
IS Department, UMBC
jfoulds@umbc.edu

Word embedding models such as the skip-gram improve the performance of Natural Language Processing (NLP) methods by revealing the latent structural relationship between words (Mikolov et al., 2013). The skip-gram learns a vector-space representation of words by using a log-bilinear classifier to parameterize the distributions over the words in the context of a given word. The similarity relationships between the semantic meanings of words are reflected in the similarity of the vectors. On the other hand, topic models such as latent Dirichlet allocation (LDA) (Blei et al., 2003) construct latent representations of topical themes and of documents, and these can be used to subsequently derive representations for words. For instance, Griffiths et al. (2007) used these semantic representations of words extracted from LDA to predict human responses to a word association task. However, LDA does not directly capture nuanced relationships between words using vector-space embeddings.

In this work, we develop a model which jointly recovers vector-space embeddings of words, documents, and topics, in the same semantic space. Our approach is to deconstruct topic models such as LDA by reparameterizing them using vector-space embeddings. This leads to an embedding version of LDA, which we call *neural embedding allocation* (NEA). We train NEA by minimizing the KL-divergence to the data distribution of the corresponding LDA model, using a stream of simulated data from the model. We can view our proposed method as *learning to mimic a topic model with a skip-gram style embedding model* to reveal underlying semantic representations. Although the NEA model generally loses some model capacity relative to the topic models, it provides vector representation which encode valuable similarity information between words, e.g. strong is more similar to powerful than to Paris.

**Latent Dirichlet allocation**

- For each document $d$
  - For each word in the document $w_{di}$
    - Draw $z_{di} \sim$ Discrete$(\theta^{(d)})$
    - Draw $w_{di} \sim$ Discrete$(\phi^{(z_{di})})$

**Neural embedding allocation**

- For each document $d$
  - For each word in the document $w_{di}$
    - Draw $z_{di} \propto \exp(\bar{v}_{z_{di}}^T v_d)$
    - Draw $w_{di} \propto \exp({v'_{w_{di}}}^T \bar{v}_{z_{di}})$

Figure 1: Our neural embedding allocation model reparameterizes topic models with embeddings.

In Figure 1, we show how our proposed NEA model reparameterizes LDA. The NEA model (bottom) is parameterized by document vectors $v_d$, topic vectors $\bar{v}_k$, and "output" word vectors $v'_w$ which mimic the discrete distributions of LDA (top) such as document distributions over topics, $\theta^{(d)}$ and topic distribution over words, $\phi^{(k)}$ by re-encoding them using log-bilinear models. The topic vectors $\bar{v}_k$ are used as both the "output" vectors for $\theta^{(d)}$, and the input vectors for $\phi^{(k)}$. Therefore, our developed model will be able to encode similarity relationships between documents, topics, and words within a shared latent space. We have also considered a "skip-gram" variant of NEA in which a word's meaning is assumed to be determined by the contexts in which it is found, and where topics generate the context of a word rather than the word itself (not shown for space).

Finally, we are aiming to make our model scalable to massive datasets by speeding up the NEA training. In addition to using noise contrastive estimation (NCE) and/or negative sampling to speed up estimation (Mnih and Kavukcuoglu, 2013), we plan to use the alias method to accelerate the simulation from the topic model when training NEA to mimic it (Li et al., 2014).

## References


David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Thomas L Griffiths, Mark Steyvers, and Joshua B Tenenbaum. 2007. Topics in semantic representation. *Psychological Review*, 114(2):211.

Aaron Q Li, Amr Ahmed, Sujith Ravi, and Alexander J Smola. 2014. Reducing the sampling complexity of topic models. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 891–900. ACM.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Andriy Mnih and Koray Kavukcuoglu. 2013. Learning word embeddings efficiently with noise-contrastive estimation. In *Advances in Neural Information Processing Systems*, pages 2265–2273.