**Dynamic Word Embeddings**

**Robert Bamler 1 Stephan Mandt 1**

**Abstract**

We present a probabilistic language model for time-stamped text data which tracks the semantic evolution of individual words over time. The model represents words and contexts by latent trajectories in an embedding space. At each moment in time, the embedding vectors are inferred from a probabilistic version of word2vec (Mikolov et al., 2013b). These embedding vectors are connected in time through a latent diffusion process. We describe two scalable variational inference algorithms—skip- gram smoothing and skip-gram filtering—that al- low us to train the model jointly over all times; thus learning on all data while simultaneously al- lowing word and context vectors to drift. Experimental results on three different corpora demon- strate that our dynamic model infers word embedding trajectories that are more interpretable and lead to higher predictive likelihoods than competing methods that are based on static models trained separately on time slices.

**摘要**

我们提出了一种针对时间戳文本数据的概率语言模型，该模型可跟踪单词随时间的语义演变。该模型通过嵌入空间中的潜在轨迹来表示单词和上下文。在每个时刻，词向量是从word2vec的概率版本推断出来的（Mikolov等，2013b）。这些词向量通过潜在扩散过程及时连接。我们描述了两种可扩展的变分推理算法 - 跳过式平滑和跳过式克式过滤 - 这使得我们可以在任何时候同时训练模型从而学习所有数据，同时允许其中的字和上下文向量移动。通过三个不同语料库的实验结果表明，我们的动态模型推断出更易于解释的词嵌入轨迹，并且比基于在时间片上单独训练的静态模型的竞争方法导致更高的预测可能性。