**Word representations: A simple and general method**

**for semi-supervised learning**

**Joseph Turian**

De ́partement d’Informatique et Recherche Ope ́rationnelle (DIRO) Universite ́ de Montre ́al Montre ́al, Que ́bec, Canada, H3T 1J4 lastname@iro.umontreal.ca

**Lev Ratinov**

Department of Computer Science University of Illinois at Urbana-Champaign Urbana, IL 61801 ratinov2@uiuc.edu

**Yoshua Bengio**

De ́partement d’Informatique et Recherche Ope ́rationnelle (DIRO) Universite ́ de Montre ́al Montre ́al, Que ́bec, Canada, H3T 1J4 bengioy@iro. umontreal.ca

**Abstract**

If we take an existing supervised NLP system, a simple and general way to improve accuracy is to use unsupervised word representations as extra word features. We evaluate Brown clusters, Collobert and Weston (2008) embeddings, and HLBL (Mnih & Hinton, 2009) embeddings of words on both NER and chunking. We use near state-of-the-art supervised baselines, and find that each of the three word representations improves the accuracy of these baselines. We find further

improvements by combining different word representations. You can download our word features, for off-the-shelf use in existing NLP systems, as well as our code, here:

**摘要**

如果我们采用现有的有监督的NLP系统，提高准确度的一种简单而通用的方法是使用无监督的单词表示作为额外的单词特征。我们评估了Brown集，Collobert和Weston（2008）嵌入，以及HLBL（Mnih＆Hinton，2009）在NER和块上的词向量，并使用最先进的监督基线，发现在三个单词表示中的每一个都提高了这些基线的准确性。我们通过组合不同的词向量得到了进一步的改进。您可以下载我们的词特征，在现有NLP系统改进使用，以及下载我们的代码：http://metaoptimize.com/projects/wordreprs/