[  
[

# Word embeddings in 2017: Trends and future directions 2017年单词嵌入：趋势和未来方向

原文链接：  
原文链接：

Table of contents:  
目录：

* [Subword-level embeddings](http://ruder.io/word-embeddings-2017/index.html#subwordlevelembeddings)
* [OOV handling](http://ruder.io/word-embeddings-2017/index.html#oovhandling)
* [Evaluation](http://ruder.io/word-embeddings-2017/index.html#evaluation)
* [Multi-sense embeddings](http://ruder.io/word-embeddings-2017/index.html#multisenseembeddings)
* [Beyond words as points](http://ruder.io/word-embeddings-2017/index.html#beyondwordsaspoints)
* [Phrases and multi-word expressions](http://ruder.io/word-embeddings-2017/index.html#phrasesandmultiwordexpressions)
* [Bias](http://ruder.io/word-embeddings-2017/index.html#bias)
* [Temporal dimension](http://ruder.io/word-embeddings-2017/index.html#temporaldimension)
* [Lack of theoretical understanding](http://ruder.io/word-embeddings-2017/index.html#lackoftheoreticalunderstanding)
* [Task and domain-specific embeddings](http://ruder.io/word-embeddings-2017/index.html#taskanddomainspecificembeddings)
* [Transfer learning](http://ruder.io/word-embeddings-2017/index.html#transfer%20learning)
* [Embeddings for multiple languages](http://ruder.io/word-embeddings-2017/index.html#embeddingsformultiplelanguages)
* [Embeddings based on other contexts](http://ruder.io/word-embeddings-2017/index.html#embeddingsbasedonothercontexts)

The word2vec method based on skip-gram with negative sampling (Mikolov et al., 2013) [] was published in 2013 and had a large impact on the field, mainly through its accompanying software package, which enabled efficient training of dense word representations and a straightforward integration into downstream models. In some respects, we have come far since then: Word embeddings have established themselves as an integral part of Natural Language Processing (NLP) models. In other aspects, we might as well be in 2013 as we have not found ways to pre-train word embeddings that have managed to supersede the original word2vec.  
基于带负采样的跳过图的word2vec方法（Mikolov等人，2013年）[]于2013年发布，主要通过其附带的软件包对该领域产生了巨大影响，该软件包能够有效地训练密集的词表示，并直接集成到下游模型中。在某些方面，我们已经取得了很大进展：单词嵌入已经成为自然语言处理（NLP）模型的一个组成部分。在其他方面，我们可能会在2013年，因为我们还没有找到方法来预先训练单词嵌入，这些嵌入已经成功地取代了原来的word2vec。

This post will focus on the deficiencies of word embeddings and how recent approaches have tried to resolve them. If not otherwise stated, this post discusses pre-trained word embeddings, i.e. word representations that have been learned on a large corpus using word2vec and its variants. Pre-trained word embeddings are most effective if not millions of training examples are available (and thus transferring knowledge from a large unlabelled corpus is useful), which is true for most tasks in NLP. For an introduction to word embeddings, refer to .  
这篇文章将集中讨论单词嵌入的缺陷以及最近的方法是如何解决它们的。如果没有另外说明，这篇文章讨论了预先训练的单词嵌入，即在使用word2vec及其变体的大型语料库中学习到的单词表示。如果没有数以百万计的训练实例可用（因此从一个大的未标记语料库中转移知识是有用的），那么预训练的单词嵌入是最有效的，这对于NLP中的大多数任务都是如此。有关单词嵌入的介绍，请参阅。

## Subword-level embeddings 子字级嵌入

Word embeddings have been augmented with subword-level information for many applications such as named entity recognition (Lample et al., 2016) [], part-of-speech tagging (Plank et al., 2016) [], dependency parsing (Ballesteros et al., 2015; Yu & Vu, 2017) [, ], and language modelling (Kim et al., 2016) []. Most of these models employ a CNN or a BiLSTM that takes as input the characters of a word and outputs a character-based word representation.  
单词嵌入已经增加了许多应用程序的子单词级信息，例如命名实体识别（Lample等人，2016年）[]、词性标记（Plank等人，2016年）[]、依赖性分析（Ballesteros等人，2015年；Yu&Vu，2017年）[，]和语言建模（Kim等人，2016年）[]。这些模型大多采用CNN或BiLSTM，后者将单词的字符作为输入，并输出基于字符的单词表示。

For incorporating character information into pre-trained embeddings, however, character n-grams features have been shown to be more powerful than composition functions over individual characters (Wieting et al., 2016; Bojanowski et al., 2017) [, ]. Character n-grams – by far not a novel feature for text categorization (Cavnar et al., 1994) [] – are particularly efficient and also form the basis of Facebook’s fastText classifier (Joulin et al., 2016) []. Embeddings learned using fastText are .  
然而，为了将字符信息合并到预先训练的嵌入中，字符n-grams特征被证明比单个字符的合成函数更强大（Wieting等人，2016；Bojanowski等人，2017）[，]。字符n-grams——到目前为止还不是一个用于文本分类的新颖功能（Cavnar等人，1994年）[]——特别有效，而且也是Facebook fastText分类器的基础（Joulin等人，2016年）[]。使用fastText学习的嵌入是。

Subword units based on byte-pair encoding have been found to be particularly useful for machine translation (Sennrich et al., 2016) [] where they have replaced words as the standard input units. They are also useful for tasks with many unknown words such as entity typing (Heinzerling & Strube, 2017) [], but have not been shown to be helpful yet for standard NLP tasks, where this is not a major concern. While they can be learned easily, it is difficult to see their advantage over character-based representations for most tasks (Vania & Lopez, 2017) [].  
基于字节对编码的子字单元被发现对机器翻译特别有用（Sennrich等人，2016年）[]，在机器翻译中，它们已将单词替换为标准输入单元。它们也适用于包含许多未知单词的任务，例如实体类型（Heinzerling&Strube，2017）[]，但还没有显示出对标准NLP任务有帮助，因为这不是一个主要问题。虽然它们很容易学习，但在大多数任务中，很难看到它们相对于基于角色的表示的优势（Vania&Lopez，2017年）[]。

Another choice for using pre-trained embeddings that integrate character information is to leverage a state-of-the-art language model (Jozefowicz et al., 2016) [] trained on a large in-domain corpus, e.g. the 1 Billion Word Benchmark (a pre-trained Tensorflow model can be found ). While language modelling has been found to be useful for different tasks as auxiliary objective (Rei, 2017) [], pre-trained language model embeddings have also been used to augment word embeddings (Peters et al., 2017) []. As we start to better understand how to pre-train and initialize our models, pre-trained language model embeddings are poised to become more effective. They might even supersede word2vec as the go-to choice for initializing word embeddings by virtue of having become more expressive and easier to train due to better frameworks and more computational resources over the last years.  
使用集成字符信息的预训练嵌入的另一种选择是利用最先进的语言模型（Jozefowicz等人，2016年）[]在大型领域内语料库上训练，例如10亿字基准（可以找到预训练的Tensorflow模型）。虽然语言建模已经被认为是有用的不同任务作为辅助目标（RII，2017）[]，预训练语言模型嵌入也被用来增加Word嵌入（彼得斯等人，2017）[]。随着我们开始更好地理解如何预先训练和初始化模型，预先训练的语言模型嵌入将变得更加有效。它们甚至可能取代word2vec，成为初始化word嵌入的首选，因为在过去几年中，由于更好的框架和更多的计算资源，它们变得更具表现力和更易于训练。

## OOV handling OOV处理

One of the main problems of using pre-trained word embeddings is that they are unable to deal with out-of-vocabulary (OOV) words, i.e. words that have not been seen during training. Typically, such words are set to the UNK token and are assigned the same vector, which is an ineffective choice if the number of OOV words is large. Subword-level embeddings as discussed in the last section are one way to mitigate this issue. Another way, which is effective for reading comprehension (Dhingra et al., 2017) [] is to assign OOV words their pre-trained word embedding, if one is available.  
使用预先训练的单词嵌入的一个主要问题是，它们无法处理词汇表外（OOV）单词，即在训练过程中没有看到的单词。通常，这些单词被设置为UNK标记，并被分配相同的向量，如果OOV单词的数量很大，这是一个无效的选择。上一节讨论的子字级嵌入是缓解此问题的一种方法。另一种对阅读理解有效的方法（Dhingra等人，2017）[]是将OOV单词分配给他们预先训练好的单词嵌入（如果有的话）。

Recently, different approaches have been proposed for generating embeddings for OOV words on-the-fly. Herbelot and Baroni (2017) [] initialize the embedding of OOV words as the sum of their context words and then rapidly refine only the OOV embedding with a high learning rate. Their approach is successful for a dataset that explicitly requires to model nonce words, but it is unclear if it can be scaled up to work reliably for more typical NLP tasks. Another interesting approach for generating OOV word embeddings is to train a character-based model to explicitly re-create pre-trained embeddings (Pinter et al., 2017) []. This is particularly useful in low-resource scenarios, where a large corpus is inaccessible and only pre-trained embeddings are available.  
最近，人们提出了不同的方法来动态生成OOV单词的嵌入。Herbelot和Baroni（2017）[]将OOV词的嵌入初始化为上下文词的和，然后快速地只对OOV嵌入进行优化，具有较高的学习率。他们的方法对于一个明确要求对nonce单词建模的数据集来说是成功的，但是还不清楚它是否能够被扩展以可靠地用于更典型的NLP任务。生成OOV单词嵌入的另一个有趣方法是训练基于字符的模型，以显式地重新创建预先训练的嵌入（Pinter等人，2017年）[]。这在低资源的场景中尤其有用，在这种场景中，大型语料库是不可访问的，并且只有经过预训练的嵌入是可用的。

## Evaluation 评价

Evaluation of pre-trained embeddings has been a contentious issue since their inception as the commonly used evaluation via word similarity or analogy datasets has been shown to only correlate weakly with downstream performance (Tsvetkov et al., 2015) []. The exclusively focused on better ways to evaluate pre-trained embeddings. As it stands, the consensus seems to be that – while pre-trained embeddings can be evaluated on intrinsic tasks such as word similarity for comparison against previous approaches – the best way to evaluate them is extrinsic evaluation on downstream tasks.  
预训练嵌入的评估自其诞生以来一直是一个有争议的问题，因为通过词相似度或类比数据集的常用评估已被证明仅与下游性能弱相关（Tsvetkov等人，2015年）[]。他们专注于更好的方法来评估预先训练的嵌入。从目前的情况来看，人们的共识似乎是，尽管可以对经过预训练的嵌入进行评估，例如对单词相似度进行评估，以便与以前的方法进行比较，但评估它们的最佳方法是对下游任务进行外部评估。

## Multi-sense embeddings 多意义嵌入

A commonly cited criticism of word embeddings is that they are unable to capture polysemy. outlined the work in recent years that focused on learning separate embeddings for multiple senses of a word (Neelakantan et al., 2014; Iacobacci et al., 2015; Pilehvar & Collier, 2016) [, , ]. However, most existing approaches for learning multi-sense embeddings solely evaluate on word similarity. Pilehvar et al. (2017) [] are one of the first to show results on topic categorization as a downstream task; while multi-sense embeddings outperform randomly initialized word embeddings in their experiments, they are outperformed by pre-trained word embeddings.  
对单词嵌入的一个常见批评是它们不能捕捉多义词。概述了近年来的工作，重点是学习单词多个感官的独立嵌入（Neelakantan等人，2014；Iacobcci等人，2015；Pilehvar&Collier，2016）[，]。然而，大多数现有的学习多感兴趣嵌入的方法只对单词相似度进行评估。Pilehvar等人。（2017）[]是最早将主题分类作为下游任务显示结果的人之一；虽然多意义嵌入在实验中优于随机初始化的词嵌入，但它们优于预先训练的词嵌入。

Given the stellar results Neural Machine Translation systems using word embeddings have achieved in recent years (Johnson et al., 2016) [], it seems that the current generation of models is expressive enough to contextualize and disambiguate words in context without having to rely on a dedicated disambiguation pipeline or multi-sense embeddings. However, we still need better ways to understand whether our models are actually able to sufficiently disambiguate words and how to improve this disambiguation behaviour if necessary.  
考虑到恒星结果，近年来已经实现了使用词嵌入的神经机器翻译系统（约翰逊等人，2016）[]，看来当前的模型是足够表达的，而不必依赖于专用的消歧流水线或多感嵌入，从而在上下文中语境化和消除歧义。然而，我们仍然需要更好的方法来理解我们的模型是否能够足够地消除单词的歧义，以及在必要时如何改进这种消除歧义的行为。

## Beyond words as points 言外之意

While we might not need separate embeddings for every sense of each word for good downstream performance, reducing each word to a point in a vector space is unarguably overly simplistic and causes us to miss out on nuances that might be useful for downstream tasks. An interesting direction is thus to employ other representations that are better able to capture these facets. Vilnis & McCallum (2015) [] propose to model each word as a probability distribution rather than a point vector, which allows us to represent probability mass and uncertainty across certain dimensions. Athiwaratkun & Wilson (2017) [] extend this approach to a multimodal distribution that allows to deal with polysemy, entailment, uncertainty, and enhances interpretability.  
虽然我们可能不需要为每个单词的每个意义单独嵌入以获得良好的下游性能，但将每个单词减少到向量空间中的某个点无疑过于简单，并导致我们忽略可能对下游任务有用的细微差别。因此，一个有趣的方向是采用能够更好地捕获这些方面的其他表示。Vilnis&McCallum（2015）[]建议将每个单词建模为概率分布，而不是点向量，这允许我们表示特定维度的概率质量和不确定性。Athiwaratkun&Wilson（2017）[]将此方法扩展到允许处理多义词、蕴涵、不确定性和增强可解释性的多模态分布。

Rather than altering the representation, the embedding space can also be changed to better represent certain features. Nickel and Kiela (2017) [], for instance, embed words in a hyperbolic space, to learn hierarchical representations. Finding other ways to represent words that incorporate linguistic assumptions or better deal with the characteristics of downstream tasks is a compelling research direction.  
嵌入空间也可以改变，以更好地表示某些特征，而不是改变表示。例如，Nickel和Kiela（2017）[]将单词嵌入双曲空间，以学习层次表示法。寻找其他方法来表达包含语言假设或更好地处理下游任务特征的单词是一个引人注目的研究方向。

## Phrases and multi-word expressions 短语和多词表达式

In addition to not being able to capture multiple senses of words, word embeddings also fail to capture the meanings of phrases and multi-word expressions, which can be a function of the meaning of their constituent words, or have an entirely new meaning. Phrase embeddings have been proposed already in the original word2vec paper (Mikolov et al., 2013) [] and there has been consistent work on learning better compositional and non-compositional phrase embeddings (Yu & Dredze, 2015; Hashimoto & Tsuruoka, 2016) [, ]. However, similar to multi-sense embeddings, explicitly modelling phrases has so far not shown significant improvements on downstream tasks that would justify the additional complexity. Analogously, a better understanding of how phrases are modelled in neural networks would pave the way to methods that augment the capabilities of our models to capture compositionality and non-compositionality of expressions.  
词的嵌入除了不能捕捉词的多种意义外，还不能捕捉词组和多词表达式的意义，这些意义可能是词组和多词表达式意义的函数，也可能具有全新的意义。短语嵌入已经在最初的word2vec论文中提出（Mikolov等人，2013年）[]，并且在学习更好的组合和非组合短语嵌入方面已经有了一致的工作（Yu&Dredze，2015；Hashimoto&Tsuruoka，2016）[，]。然而，类似于多感嵌入，明确建模短语迄今没有显示出明显的改善下游任务，将证明额外的复杂性。类似地，更好地理解短语是如何在神经网络中建模的，将为增强我们的模型捕获表达式的组成性和非组成性的能力的方法铺平道路。

## Bias 偏倚

Bias in our models is becoming a larger issue and we are only starting to understand its implications for training and evaluating our models. Even word embeddings trained on Google News articles exhibit female/male gender stereotypes to a disturbing extent (Bolukbasi et al., 2016) []. Understanding what other biases word embeddings capture and finding better ways to remove theses biases will be key to developing fair algorithms for natural language processing.  
我们的模型中的偏见正在成为一个更大的问题，我们只是开始了解它对培训和评估我们的模型的影响。即使是在谷歌新闻文章上训练的单词嵌入，也会在令人不安的程度上展示女性/男性的性别刻板印象（Bolukbasi等人，2016年）[]。理解单词嵌入捕获的其他偏见，并找到更好的方法来消除这些偏见，将是开发自然语言处理公平算法的关键。

## Temporal dimension 时间维度

Words are a mirror of the zeitgeist and their meanings are subject to continuous change; current representations of words might differ substantially from the way these words where used in the past and will be used in the future. An interesting direction is thus to take into account the temporal dimension and the diachronic nature of words. This can allows us to reveal laws of semantic change (Hamilton et al., 2016; Bamler & Mandt, 2017; Dubossarsky et al., 2017) [, , ], to model temporal word analogy or relatedness (Szymanski, 2017; Rosin et al., 2017) [, ], or to capture the dynamics of semantic relations (Kutuzov et al., 2017) [].  
词汇是时代精神的一面镜子，其含义会不断变化；当前词汇的表示方式可能与过去和将来使用这些词汇的方式大不相同。因此，一个有趣的方向是考虑到词的时间维度和历时性。这可以让我们揭示语义变化的规律（Hamilton等人，2016；Bamler&Mandt，2017；Dubossarsky等人，2017）[，]，建立时间词类比或关联性模型（Szymanski，2017；Rosin等人，2017）[，]，或捕捉语义关系的动态（Kutuzov等人，2017）[]。

## Lack of theoretical understanding 缺乏理论理解

Besides the insight that word2vec with skip-gram negative sampling implicitly factorizes a PMI matrix (Levy & Goldberg, 2014) [], there has been comparatively little work on gaining a better theoretical understanding of the word embedding space and its properties, e.g. that summation captures analogy relations. Arora et al. (2016) [] propose a new generative model for word embeddings, which treats corpus generation as a random walk of a discourse vector and establishes some theoretical motivations regarding the analogy behaviour. Gittens et al. (2017) [] provide a more thorough theoretical justification of additive compositionality and show that skip-gram word vectors are optimal in an information-theoretic sense. Mimno & Thompson (2017) [] furthermore reveal an interesting relation between word embeddings and the embeddings of context words, i.e. that they are not evenly dispersed across the vector space, but occupy a narrow cone that is diametrically opposite to the context word embeddings. Despite these additional insights, our understanding regarding the location and properties of word embeddings is still lacking and more theoretical work is necessary.  
除了具有跳过gram负抽样的word2vec隐式分解PMI矩阵（Levy&Goldberg，2014）[]的见解外，对于更好地理解单词嵌入空间及其属性（例如，该求和捕获类比关系）的理论研究相对较少。Arora等人。（2016）[]提出了一种新的词汇嵌入生成模型，将语料库生成视为话语向量的随机游走，并建立了一些关于类比行为的理论动机。Gittens等人。（2017）[]提供了关于加法成分性的更彻底的理论证明，并表明从信息论意义上讲，跳过gram词向量是最优的。Mimno&Thompson（2017）[]进一步揭示了单词嵌入和上下文单词嵌入之间的有趣关系，即它们不是均匀地分布在向量空间中，而是占据一个与上下文单词嵌入截然相反的窄圆锥体。尽管有这些额外的见解，我们对单词嵌入的位置和属性的理解仍然缺乏，需要更多的理论工作。

## Task and domain-specific embeddings 任务和领域特定的嵌入

One of the major downsides of using pre-trained embeddings is that the news data used for training them is often very different from the data on which we would like to use them. In most cases, however, we do not have access to millions of unlabelled documents in our target domain that would allow for pre-training good embeddings from scratch. We would thus like to be able to adapt embeddings pre-trained on large news corpora, so that they capture the characteristics of our target domain, but still retain all relevant existing knowledge. Lu & Zheng (2017) [] proposed a regularized skip-gram model for learning such cross-domain embeddings. In the future, we will need even better ways to adapt pre-trained embeddings to new domains or to incorporate the knowledge from multiple relevant domains.  
使用预先训练的嵌入的一个主要缺点是，用于训练它们的新闻数据通常与我们希望使用它们的数据非常不同。然而，在大多数情况下，我们无法访问目标域中数百万个未标记的文档，这些文档允许从头开始对良好的嵌入进行预培训。因此，我们希望能够适应预先训练的大型新闻语料库的嵌入，以便他们捕获我们的目标域的特征，但仍然保留所有相关的现有知识。Lu&Zheng（2017）[]提出了一个用于学习此类跨域嵌入的正则化跳过图模型。在未来，我们将需要更好的方法来适应预先训练的嵌入到新的领域或合并来自多个相关领域的知识。

Rather than adapting to a new domain, we can also use existing knowledge encoded in semantic lexicons to augment pre-trained embeddings with information that is relevant for our task. An effective way to inject such relations into the embedding space is retro-fitting (Faruqui et al., 2015) [], which has been expanded to other resources such as ConceptNet (Speer et al., 2017) [] and extended with an intelligent selection of positive and negative examples (Mrkšić et al., 2017) []. Injecting additional prior knowledge into word embeddings such as monotonicity (You et al., 2017) [], word similarity (Niebler et al., 2017) [], task-related grading or intensity, or logical relations is an important research direction that will allow to make our models more robust.  
而不是适应一个新的领域，我们也可以使用现有的知识编码的语义词典，以增加与我们的任务相关的信息预先训练的嵌入。将这种关系注入嵌入空间的一种有效方法是回溯拟合（Faruqui等人，2015年）[]，该方法已扩展到其他资源，如ConceptNet（Speer等人，2017年）[]，并扩展为智能选择积极和消极的例子（Mrkšić等人，2017年）[]。将额外的先验知识注入到单词嵌入中，例如单调性（You et al.，2017）[]、单词相似性（Niebler et al.，2017）[]、任务相关的分级或强度或逻辑关系，是一个重要的研究方向，这将使我们的模型更加健壮。

Word embeddings are useful for a wide variety of applications beyond NLP such as information retrieval, recommendation, and link prediction in knowledge bases, which all have their own task-specific approaches. Wu et al. (2017) [] propose a general-purpose model that is compatible with many of these applications and can serve as a strong baseline.  
Word嵌入在NLP之外的许多应用中都很有用，例如信息检索、推荐和知识库中的链接预测，这些都有自己的特定于任务的方法。吴等人。（2017）[]提出一个通用模型，该模型与许多此类应用程序兼容，并可作为一个强大的基准。

## Transfer learning 迁移学习

Rather than adapting word embeddings to any particular task, recent work has sought to create contextualized word vectors by augmenting word embeddings with embeddings based on the hidden states of models pre-trained for certain tasks, such as machine translation (McCann et al., 2017) [] or language modelling (Peters et al., 2018) []. Together with fine-tuning pre-trained models (Howard and Ruder, 2018) [], this is one of the most promising research directions.  
最近的工作并没有使单词嵌入适应任何特定的任务，而是试图通过基于为特定任务预先训练的模型的隐藏状态（如机器翻译（McCann et al.，2017）[]或语言建模（Peters et al.，2018）[]）的嵌入来增加单词嵌入，从而创建上下文化的单词向量。与微调预训练模型（Howard and Ruder，2018）[]一起，这是最有希望的研究方向之一。

## Embeddings for multiple languages 多种语言的嵌入

As NLP models are being increasingly employed and evaluated on multiple languages, creating multilingual word embeddings is becoming a more important issue and has received increased interest over recent years. A promising direction is to develop methods that learn cross-lingual representations with as few parallel data as possible, so that they can be easily applied to learn representations even for low-resource languages. For a recent survey in this area, refer to Ruder et al. (2017) [].  
随着NLP模型在多种语言上的应用和评估越来越多，创建多语言单词嵌入成为一个更重要的问题，近年来受到越来越多的关注。一个有希望的方向是开发尽可能少的并行数据来学习跨语言表示的方法，这样即使对于低资源语言，它们也可以很容易地应用于学习表示。有关此区域的最新调查，请参阅Ruder等人。（2017年）[]。

## Embeddings based on other contexts 基于其他上下文的嵌入

Word embeddings are typically learned only based on the window of surrounding context words. Levy & Goldberg (2014) [] have shown that dependency structures can be used as context to capture more syntactic word relations; Köhn (2015) [] finds that such dependency-based embeddings perform best for a particular multilingual evaluation method that clusters embeddings along different syntactic features.  
单词嵌入通常只基于周围上下文单词的窗口来学习。Levy&Goldberg（2014）[]已经表明依赖结构可以用作上下文来捕获更多的句法词关系；Kóhn（2015）[]发现这种基于依赖的嵌入对于沿着不同句法特征对嵌入进行聚类的特定多语言评估方法最有效。

Melamud et al. (2016) [] observe that different context types work well for different downstream tasks and that simple concatenation of word embeddings learned with different context types can yield further performance gains. Given the recent success of incorporating graph structures into neural models for different tasks as – for instance – exhibited by graph-convolutional neural networks (Bastings et al., 2017; Marcheggiani & Titov, 2017) [, ], we can conjecture that incorporating such structures for learning embeddings for downstream tasks may also be beneficial.  
Melamud等人。（2016）[]观察到不同的上下文类型对于不同的下游任务很好地工作，并且使用不同上下文类型学习的单词嵌入的简单连接可以产生进一步的性能提升。鉴于最近成功地将图形结构融入不同任务的神经模型中，例如图卷积神经网络（Bastings et al.，2017；Marcheggiani&Titov，2017）[，]），我们可以推测，将这些结构融入到下游任务的学习嵌入中也可能是有益的。

Besides selecting context words differently, additional context may also be used in other ways: Tissier et al. (2017) [] incorporate co-occurrence information from dictionary definitions into the negative sampling process to move related works closer together and prevent them from being used as negative samples. We can think of topical or relatedness information derived from other contexts such as article headlines or Wikipedia intro paragraphs that could similarly be used to make the representations more applicable to a particular downstream task.  
除了选择不同的语境词外，附加语境也可以以其他方式使用：Tissier等人。（2017）[]将字典定义中的共现信息纳入负采样过程，以使相关工作更紧密地结合在一起，并防止它们被用作负采样。我们可以考虑从其他上下文（如文章标题或维基百科简介段落）获得的主题或相关信息，这些信息同样可以用于使表示更适用于特定的下游任务。

## Conclusion 结论

It is nice to see that as a community we are progressing from applying word embeddings to every possible problem to gaining a more principled, nuanced, and practical understanding of them. This post was meant to highlight some of the current trends and future directions for learning word embeddings that I found most compelling. I’ve undoubtedly failed to mention many other areas that are equally important and noteworthy. Please let me know in the comments below what I missed, where I made a mistake or misrepresented a method, or just which aspect of word embeddings you find particularly exciting or unexplored.  
很高兴地看到，作为一个社区，我们正在从应用单词嵌入到每一个可能的问题，以获得一个更原则，更微妙，更实际的理解。这篇文章的目的是强调一些当前的趋势和未来的方向，学习单词嵌入，我发现最引人注目。毫无疑问，我没有提到其他许多同样重要和值得注意的领域。请在下面的评论中让我知道我遗漏了什么，我在哪里犯了错误或歪曲了一个方法，或者你觉得单词嵌入的哪个方面特别令人兴奋或未经探索。

# Hacker News 黑客新闻

Refer to the for some more insights on word embeddings.  
有关单词嵌入的更多信息，请参阅。

## Other blog posts on word embeddings 其他关于word嵌入的博客文章

If you want to learn more about word embeddings, these other blog posts on word embeddings are also available:  
如果您想了解更多关于word嵌入的信息，还可以在以下博客中找到其他关于word嵌入的文章：

* [On word embeddings - Part 1](http://sebastianruder.com/word-embeddings-1/index.html)
* [On word embeddings - Part 2: Approximating the softmax](http://sebastianruder.com/word-embeddings-softmax/index.html)
* [On word embeddings - Part 3: The secret ingredients of word2vec](http://sebastianruder.com/secret-word2vec/index.html)
* [Unofficial Part 4: A survey of cross-lingual embedding models](http://sebastianruder.com/cross-lingual-embeddings/index.html)

## References 工具书类

1. Cavnar, W. B., Trenkle, J. M., & Mi, A. A. (1994). N-Gram-Based Text Categorization. Ann Arbor MI 48113.2, 161–175.   
   Cavnar，W.B.，Trenkle，J.M.和Mi，A.A.（1994年）。基于N-Gram的文本分类。密歇根州安阿伯48113.2，161–175。
2. Wieting, J., Bansal, M., Gimpel, K., & Livescu, K. (2016). Charagram: Embedding Words and Sentences via Character n-grams. Retrieved from   
   Wieting，J.，Bansal，M.，Gimpel，K.，和Livescu，K.（2016年）。字谜：通过字谜嵌入单词和句子。检索自
3. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics. Retrieved from   
   Bojanowski，P.，Grave，E.，Joulin，A.，和Mikolov，T.（2017年）。用子词信息丰富词向量。计算语言学协会学报。检索自
4. Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of Tricks for Efficient Text Classification. arXiv Preprint arXiv:1607.01759. Retrieved from   
   Joulin，A.，Grave，E.，Bojanowski，P.，和Mikolov，T.（2016年）。有效的文本分类技巧包。arXiv预印本arXiv:1607.01759。检索自
5. Rei, M. (2017). Semi-supervised Multitask Learning for Sequence Labeling. In Proceedings of ACL 2017.   
   Rei，M.（2017年）。序列标记的半监督多任务学习。在2017年美国公民自由联盟诉讼中。
6. Peters, M. E., Ammar, W., Bhagavatula, C., & Power, R. (2017). Semi-supervised sequence tagging with bidirectional language models. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 1756–1765).   
   Peters，M.E.，Ammar，W.，Bhagavatula，C.，和Power，R.（2017年）。基于双向语言模型的半监督序列标注。计算语言学协会第55届年会论文集（第1756-1765页）。
7. Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., & Wu, Y. (2016). Exploring the Limits of Language Modeling. arXiv Preprint arXiv:1602.02410. Retrieved from   
   Jozefowicz，R.，Vinyals，O.，Schuster，M.，Shazeer，N.和Wu，Y.（2016年）。探索语言建模的局限性。arXiv预印本arXiv:1602.02410。检索自
8. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural Architectures for Named Entity Recognition. In NAACL-HLT 2016.   
   Lample，G.，Ballesteros，M.，Subramanian，S.，Kawakami，K.，和Dyer，C.（2016年）。命名实体识别的神经结构。在NAACL-HLT 2016。
9. Plank, B., Søgaard, A., & Goldberg, Y. (2016). Multilingual Part-of-Speech Tagging with Bidirectional Long Short-Term Memory Models and Auxiliary Loss. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.   
   Plack，B.，Søgaard，A.，和Goldberg，Y.（2016年）。具有双向长短期记忆模型和辅助损失的多语言词性标注。计算语言学协会第54届年会论文集。
10. Yu, X., & Vu, N. T. (2017). Character Composition Model with Convolutional Neural Networks for Dependency Parsing on Morphologically Rich Languages. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 672–678).   
    Yu，X.和Vu，N.T.（2017年）。基于卷积神经网络的形态丰富语言依赖性分析字符合成模型。计算语言学协会第55届年会论文集（第672-678页）。
11. Kim, Y., Jernite, Y., Sontag, D., & Rush, A. M. (2016). Character-Aware Neural Language Models. AAAI. Retrieved from   
    Kim，Y.，Jernite，Y.，Sontag，D.和Rush，A.M.（2016年）。字符感知神经语言模型。啊。检索自
12. Sennrich, R., Haddow, B., & Birch, A. (2016). Neural Machine Translation of Rare Words with Subword Units. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016). Retrieved from   
    Sennrich，R.，Haddow，B.，和Birch，A.（2016年）。具有子词单位的罕见词的神经机器翻译。《计算语言学协会第54届年会论文集》（ACL 2016）。检索自
13. Heinzerling, B., & Strube, M. (2017). BPEmb: Tokenization-free Pre-trained Subword Embeddings in 275 Languages. Retrieved from   
    Heinzerling，B.和Strube，M.（2017年）。BPEmb：275种语言中无标记化的预训练子词嵌入。检索自
14. Dhingra, B., Liu, H., Salakhutdinov, R., & Cohen, W. W. (2017). A Comparative Study of Word Embeddings for Reading Comprehension. arXiv preprint arXiv:1703.00993.   
    Dhingra，B.，Liu，H.，Salakhutdinov，R.，和Cohen，W.W.（2017年）。词汇嵌入在阅读理解中的比较研究。arXiv预印本arXiv:1703.00993。
15. Herbelot, A., & Baroni, M. (2017). High-risk learning: acquiring new word vectors from tiny data. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing.   
    Herbelot，A.和Baroni，M.（2017年）。高风险学习：从微小的数据中获取新的词向量。2017年自然语言处理经验方法会议记录。
16. Pinter, Y., Guthrie, R., & Eisenstein, J. (2017). Mimicking Word Embeddings using Subword RNNs. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Retrieved from   
    Pinter，Y.，Guthrie，R.和Eisenstein，J.（2017年）。使用子单词RNNs模拟单词嵌入。2017年自然语言处理经验方法会议记录。检索自
17. Ballesteros, M., Dyer, C., & Smith, N. A. (2015). Improved Transition-Based Parsing by Modeling Characters instead of Words with LSTMs. In Proceedings of EMNLP 2015.   
    Ballesteros，M.，Dyer，C.，和Smith，N.A.（2015年）。通过使用LSTMs对字符而不是单词进行建模，改进了基于转换的语法分析。2015年EMNLP会议记录。
18. Neelakantan, A., Shankar, J., Passos, A., & Mccallum, A. (2014). Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space. In Proceedings fo (pp. 1059–1069).   
    Neelakantan，A.，Shankar，J.，Passos，A.，和Mccallum，A.（2014年）。向量空间中每个字多个嵌入的有效非参数估计。在第1059-1069页。
19. Iacobacci, I., Pilehvar, M. T., & Navigli, R. (2015). SensEmbed: Learning Sense Embeddings for Word and Relational Similarity. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (pp. 95–105).   
    Iacobcci，I.，Pilehvar，M.T.和Navigli，R.（2015）。词义嵌入：学习词义和关系相似性的词义嵌入。计算语言学协会第53届年会和第七届自然语言处理国际联席会议记录（第95-105页）。
20. Pilehvar, M. T., & Collier, N. (2016). De-Conflated Semantic Representations. In Proceedings of EMNLP.   
    Pilehvar，M.T.和Collier，N.（2016年）。将语义表示分离。在EMNLP程序中。
21. Tsvetkov, Y., Faruqui, M., Ling, W., Lample, G., & Dyer, C. (2015). Evaluation of Word Vector Representations by Subspace Alignment. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17-21 September 2015, 2049–2054.   
    Tsvetkov，Y.，Faruqui，M.，Ling，W.，Lample，G.，和Dyer，C.（2015）。基于子空间对齐的词向量表示方法评价。2015年自然语言处理经验方法会议记录，葡萄牙里斯本，2015年9月17-21日，2049-2054。
22. Pilehvar, M. T., Camacho-Collados, J., Navigli, R., & Collier, N. (2017). Towards a Seamless Integration of Word Senses into Downstream NLP Applications. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1857–1869).   
    Pilehvar，M.T.，Camacho Collados，J.，Navigli，R.，和Collier，N.（2017年）。将词义无缝集成到下游NLP应用程序中。《计算语言学协会第55届年会论文集》（第一卷：长篇论文）（第1857-1869页）。
23. Johnson, M., Schuster, M., Le, Q. V, Krikun, M., Wu, Y., Chen, Z., … Dean, J. (2016). Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. arXiv Preprint arXiv:1611.0455.   
    Johnson，M.，Schuster，M.，Le，Q.V.，Krikun，M.，Wu，Y.，Chen，Z.，…Dean，J.（2016年）。谷歌的多语言神经机器翻译系统：实现零镜头翻译。arXiv预印本arXiv:1611.0455。
24. Vilnis, L., & McCallum, A. (2015). Word Representations via Gaussian Embedding. ICLR. Retrieved from   
    Vilnis，L.和McCallum，A.（2015年）。通过高斯嵌入的单词表示。ICLR。检索自
25. Athiwaratkun, B., & Wilson, A. G. (2017). Multimodal Word Distributions. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017).   
    Athiwaratkun，B.和Wilson，A.G.（2017年）。多模态单词分布。《计算语言学协会第55届年会论文集》（ACL 2017）。
26. Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In 30th Conference on Neural Information Processing Systems (NIPS 2016). Retrieved from   
    Bolukbasi，T.，Chang，K.-W.，Zou，J.，Saligrama，V.和Kalai，A.（2016年）。男人对电脑程序员就像女人对家庭主妇一样？单词嵌入。第30届神经信息处理系统会议（NIPS 2016）。检索自
27. Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (pp. 1489–1501).   
    Hamilton，W.L.，Leskovec，J.，和Jurafsky，D.（2016年）。历时嵌入揭示了语义变化的统计规律。计算语言学协会第54届年会论文集（第1489-1501页）。
28. Bamler, R., & Mandt, S. (2017). Dynamic Word Embeddings via Skip-Gram Filtering. In Proceedings of ICML 2017. Retrieved from   
    Bamler，R.和Mandt，S.（2017年）。通过跳格过滤实现动态字嵌入。在ICML 2017年会议记录中。检索自
29. Dubossarsky, H., Grossman, E., & Weinshall, D. (2017). Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models. In Conference on Empirical Methods in Natural Language Processing (pp. 1147–1156). Retrieved from   
    Dubossarsky，H.，Grossman，E.，和Weinshall，D.（2017年）。输出控制：词义变化的规律和词表示模型中固有的偏误。自然语言处理经验方法会议（1147-1156页）。检索自
30. Szymanski, T. (2017). Temporal Word Analogies : Identifying Lexical Replacement with Diachronic Word Embeddings. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 448–453).   
    Szymanski，T.（2017年）。时间词类比：用Diachronic Word Embeddings识别词汇替换。计算语言学协会第55届年会论文集（第448-453页）。
31. Rosin, G., Radinsky, K., & Adar, E. (2017). Learning Word Relatedness over Time. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Retrieved from   
    Rosin，G.，Radinsky，K.，和Adar，E.（2017年）。随着时间的推移学习单词的相关性。2017年自然语言处理经验方法会议记录。检索自
32. Kutuzov, A., Velldal, E., & Øvrelid, L. (2017). Temporal dynamics of semantic relations in word embeddings: an application to predicting armed conflict participants. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Retrieved from   
    Kutuzov，A.，Velldal，E.，和Øvrelid，L.（2017）。词汇嵌入中语义关系的时间动态：预测武装冲突参与者的应用。2017年自然语言处理经验方法会议记录。检索自
33. Levy, O., & Goldberg, Y. (2014). Neural Word Embedding as Implicit Matrix Factorization. Advances in Neural Information Processing Systems (NIPS), 2177–2185. Retrieved from   
    Levy，O.和Goldberg，Y.（2014年）。隐式矩阵分解的神经单词嵌入。神经信息处理系统进展，2177-2185。检索自
34. Arora, S., Li, Y., Liang, Y., Ma, T., & Risteski, A. (2016). A Latent Variable Model Approach to PMI-based Word Embeddings. TACL, 4, 385–399. Retrieved from   
    Arora，S.，Li，Y.，Liang，Y.，Ma，T.，和Risteski，A.（2016年）。一种基于PMI的词嵌入的潜在变量模型方法。塔克，4385-399。检索自
35. Gittens, A., Achlioptas, D., & Mahoney, M. W. (2017). Skip-Gram – Zipf + Uniform = Vector Additivity. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 69–76).   
    Gittens，A.，Achlioptas，D.，和Mahoney，M.W.（2017年）。跳过Gram–Zipf+Uniform=向量可加性。计算语言学协会第55届年会论文集（69-76页）。
36. Mimno, D., & Thompson, L. (2017). The strange geometry of skip-gram with negative sampling. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 2863–2868).   
    Mimno，D.和Thompson，L.（2017年）。负采样跳格的奇异几何。《2017年自然语言处理经验方法会议记录》（第2863-2868页）。
37. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. NIPS.   
    Mikolov，T.，Chen，K.，Corrado，G.，和Dean，J.（2013年）。词和短语的分布表示及其组成。夹。
38. Yu, M., & Dredze, M. (2015). Learning Composition Models for Phrase Embeddings. Transactions of the ACL, 3, 227–242.   
    Yu，M.和Dredze，M.（2015年）。短语嵌入的学习组合模型。ACL的事务，3227–242。
39. Hashimoto, K., & Tsuruoka, Y. (2016). Adaptive Joint Learning of Compositional and Non-Compositional Phrase Embeddings. ACL, 205–215. Retrieved from   
    Hashimoto，K.，&Tsuruoka，Y.（2016年）。组合短语嵌入和非组合短语嵌入的自适应联合学习。前交叉韧带，205-215。检索自
40. Lu, W., & Zheng, V. W. (2017). A Simple Regularization-based Algorithm for Learning Cross-Domain Word Embeddings. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 2888–2894).   
    鲁，W.和郑，V.W.（2017）。一种简单的基于正则化的跨域单词嵌入学习算法。《2017年自然语言处理经验方法会议记录》（第2888-2894页）。
41. Faruqui, M., Dodge, J., Jauhar, S. K., Dyer, C., Hovy, E., & Smith, N. A. (2015). Retrofitting Word Vectors to Semantic Lexicons. In NAACL 2015.   
    Faruqui，M.，Dodge，J.，Jauhar，S.K.，Dyer，C.，Hovy，E.，和Smith，N.A.（2015年）。将单词向量改写为语义词典。在2015年NAACL。
42. Mrkšić, N., Vulić, I., Séaghdha, D. Ó., Leviant, I., Reichart, R., Gašić, M., … Young, S. (2017). Semantic Specialisation of Distributional Word Vector Spaces using Monolingual and Cross-Lingual Constraints. TACL. Retrieved from   
    Mrkšiç，N.，Vuliç，i.，Séaghdha，D.Ó.，Leviant，i.，Reichart，R.，Gašiç，M.，…Young，S.（2017年）。使用单语和跨语言约束的分布词向量空间语义专业化。塔克。检索自
43. Ruder, S., Vulić, I., & Søgaard, A. (2017). A Survey of Cross-lingual Word Embedding Models Sebastian. arXiv preprint arXiv:1706.04902. Retrieved from   
    Ruder，S.，Vulić，I.，和Søgaard，A.（2017年）。塞巴斯蒂安跨语言单词嵌入模型综述。arXiv预印本arXiv:1706.04902。检索自
44. Levy, O., & Goldberg, Y. (2014). Dependency-Based Word Embeddings. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), 302–308.   
    Levy，O.和Goldberg，Y.（2014年）。基于依赖关系的单词嵌入。计算语言学协会第52届年会论文集，302-308。
45. Köhn, A. (2015). What’s in an Embedding? Analyzing Word Embeddings through Multilingual Evaluation. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17-21 September 2015, (2014), 2067–2073.   
    Kóhn，A.（2015年）。嵌入的是什么？通过多语种评价分析词的嵌入。2015年自然语言处理经验方法会议记录，葡萄牙里斯本，2015年9月17-21日，（2014），2067-2073。
46. Melamud, O., McClosky, D., Patwardhan, S., & Bansal, M. (2016). The Role of Context Types and Dimensionality in Learning Word Embeddings. In Proceedings of NAACL-HLT 2016 (pp. 1030–1040). Retrieved from   
    Melamud，O.，McClosky，D.，Patwardhan，S.，和Bansal，M.（2016年）。语境类型和维度在词汇嵌入学习中的作用。NAACL-HLT 2016年会议记录（第1030-1040页）。检索自
47. Bastings, J., Titov, I., Aziz, W., Marcheggiani, D., & Sima’an, K. (2017). Graph Convolutional Encoders for Syntax-aware Neural Machine Translation. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing.   
    Bastings，J.，Titov，I.，Aziz，W.，Marcheggiani，D.和Sima'an，K.（2017年）。用于语法感知神经机器翻译的图卷积编码器。2017年自然语言处理经验方法会议记录。
48. Marcheggiani, D., & Titov, I. (2017). Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing.   
    Marcheggiani，D.和Titov，I.（2017年）。用图卷积网络编码句子进行语义角色标注。2017年自然语言处理经验方法会议记录。
49. Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. Proceedings of the International Conference on Learning Representations (ICLR 2013), 1–12.   
    Mikolov，T.，Corrado，G.，Chen，K.，和Dean，J.（2013年）。向量空间中单词表示的有效估计。学习表征国际会议记录（ICLR 2013），1–12。
50. Vania, C., & Lopez, A. (2017). From Characters to Words to in Between: Do We Capture Morphology? In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 2016–2027).   
    Vania，C.和Lopez，A.（2017年）。从文字到文字再到介于两者之间：我们能捕捉到形态学吗？计算语言学协会第55届年会论文集（2016-2027页）。
51. You, S., Ding, D., Canini, K., Pfeifer, J., & Gupta, M. (2017). Deep Lattice Networks and Partial Monotonic Functions. In 31st Conference on Neural Information Processing Systems (NIPS 2017). Retrieved from   
    You，S.，Ding，D.，Canini，K.，Pfeifer，J.，和Gupta，M.（2017年）。深格网络与部分单调函数。第31届神经信息处理系统会议（NIPS 2017）。检索自
52. Nickel, M., & Kiela, D. (2017). Poincaré Embeddings for Learning Hierarchical Representations. arXiv Preprint arXiv:1705.08039. Retrieved from   
    Nickel，M.和Kiela，D.（2017年）。用于学习层次表示的Poincaré嵌入。arXiv预印本arXiv:1705.08039。检索自
53. Niebler, T., Becker, M., Pölitz, C., & Hotho, A. (2017). Learning Semantic Relatedness From Human Feedback Using Metric Learning. In Proceedings of ISWC 2017. Retrieved from   
    Niebler，T.，Becker，M.，Pólitz，C.，和Hotho，A.（2017年）。使用度量学习从人类反馈中学习语义相关。在ISWC 2017年的诉讼中。检索自
54. Wu, L., Fisch, A., Chopra, S., Adams, K., Bordes, A., & Weston, J. (2017). StarSpace: Embed All The Things! arXiv preprint arXiv:1709.03856.   
    Wu，L.，Fisch，A.，Chopra，S.，Adams，K.，Bordes，A.，和Weston，J.（2017年）。星际空间：嵌入所有的东西！arXiv预印本arXiv:1709.03856。
55. Speer, R., Chin, J., & Havasi, C. (2017). ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In AAAI 31 (pp. 4444–4451). Retrieved from   
    Speer，R.，Chin，J.，和Havasi，C.（2017年）。ConceptNet 5.5：一个开放的多语言通用知识图。AAAI 31（第4444-4451页）。检索自
56. Tissier, J., Gravier, C., & Habrard, A. (2017). Dict2Vec : Learning Word Embeddings using Lexical Dictionaries. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Retrieved from   
    Tissier，J.，Gravier，C.和Habrard，A.（2017年）。DAT2VEC：使用词汇词典学习单词嵌入。2017年自然语言处理经验方法会议记录。检索自
57. Mccann, B., Bradbury, J., Xiong, C., & Socher, R. (2017). Learned in Translation: Contextualized Word Vectors. In Advances in Neural Information Processing Systems.   
    Mccann，B.，Bradbury，J.，Xiong，C.，和Socher，R.（2017年）。在翻译中学习：语境化的词载体。神经信息处理系统进展。
58. Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. Proceedings of NAACL-HLT 2018. [↩](http://ruder.io/word-embeddings-2017/index.html#fnref:58)
59. Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. In Proceedings of ACL 2018. Retrieved from <http://arxiv.org/abs/1801.06146> [↩](http://ruder.io/word-embeddings-2017/index.html#fnref:59)

Cover image credit: Hamilton et al. (2016)