* 学习词袋模型概念：离散、高维、稀疏。
* 学习分布式表示概念：连续、低维、稠密。
* 理解word2vec词向量原理并实践，来表示文本。
* word2vec 中的数学原理详解 [word2vec1](https://blog.csdn.net/itplus/article/details/37969519)
* word2vec原理推导与代码分析 [word2vec2](http://www.hankcs.com/nlp/word2vec.html)

分布表示（distributional representation）：分布（distributional）描述的是上

下文的概率分布，因此用上下文描述语义的表示方法（基于分布假说的方

法）都可以称作分布表示。与之相对的是形式语义表示。

分布式表示（distributed representation）：分布式（distributed）描述的是把

信息分布式地存储在向量的各个维度中，与之相对的是局部表示（local

representation），如词的独热表示（one-hot representation），在高维向量中

只有一个维度描述了词的语义。一般来说，通过矩阵降维或神经网络降维

可以将语义分散存储到向量的各个维度中，因此，这类方法得到的低维向

量一般都可以称作分布式表示。

Word2Vec的目标是生成带有语义的单词的向量表示，用于进一步的NLP任务。每个单词向量通常有几百个维度，语料库中每个唯一的单词在空间中被分配一个向量。例如，单词“happy”可以表示为4维向量[0.24、0.45、0.11、0.49]，“sad”具有向量[0.88、0.78、0.45、0.91]。

这种从单词到向量的转换也被称为单词嵌入（word embedding）。这种转换的原因是机器学习算法可以对数字（在向量中的）而不是单词进行线性代数运算。

为了实现Word2Vec，有两种风格可以选择，Continuous Bag-of-Words(CBOW)或Skip-gram(SG)。简单来说，CBOW尝试从相邻单词（上下文单词）猜测输出（目标单词），而Skip-Gram从目标单词猜测上下文单词。实际上，Word2Vec是基于分布假说，其认为每个单词的上下文都在其附近的单词中。因此，通过查看它的相邻单词我们可以尝试对目标单词进行预测。

根据Mikolov（引用于这篇文章），以下是Skip-gram和CBOW之间的区别：

Skip-gram：能够很好地处理少量的训练数据，而且能够很好地表示不常见的单词或短语

CBOW：比skip-gram训练快几倍，对出现频率高的单词的准确度稍微更好一些









