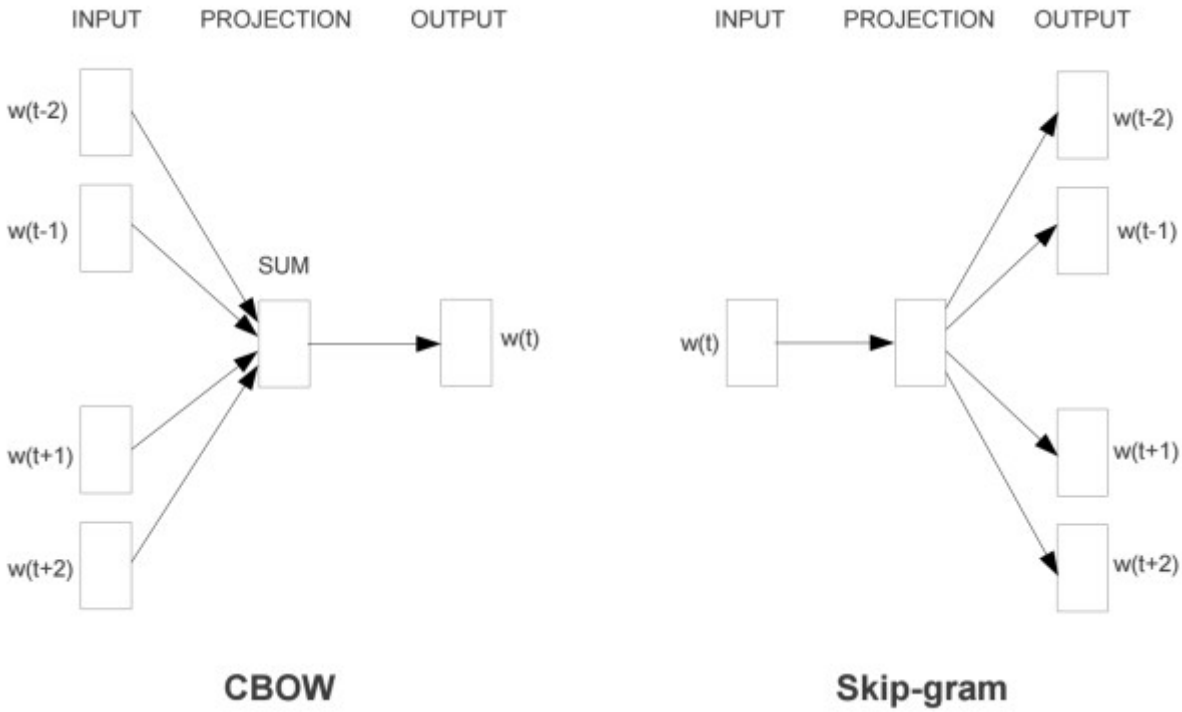
1 词表示的三种方式

在计算机中有三种方式去表示词/词的有用信息：

- 类似WordNet的资源库
- 离散化的符号例如“one-hot encoding”
- 利用“distributional similarity”分布相似性

2 word2vec

- Word2Vec模型中，主要有Skip-Gram和CBOW两种模型，从直观上理解，Skip-Gram是给定input word来预测上下文。而CBOW是给定上下文，来预测input word。两个模型得建议模型结构如下。



- Word2Vec的整个建模过程实际上与自编码器（auto-encoder）的思想很相似，即先基于训练数据构建一个神经网络，当这个模型训练好以后，我们并不会用这个训练好的模型处理新的任务，我们真正需要的是这个模型通过训练数据所学得的参数，例如隐层的权重矩阵。

- 由于Word2Vec模型是一个超级大的神经网络（权重矩阵规模非常大）。
  Word2Vec 的作者在它的第二篇论文中强调了这些问题，下面是作者在第二篇论文中的三个创新：

> 1.将常见的单词组合（word pairs）或者词组作为单个“words”来处理。
>
> 2.对高频次单词进行抽样来减少训练样本的个数。
>
> 3.对优化目标采用“negative sampling”方法，这样每个训练样本的训练只会更新一小部分的模型权重，从而降低计算负担。

- Word2Vec的两个训练技巧
  1.负采样
  2.层级softmax