如何在计算机中获得有用的意思?

- 使用 WordNet 这样的有上位关系的分类方法

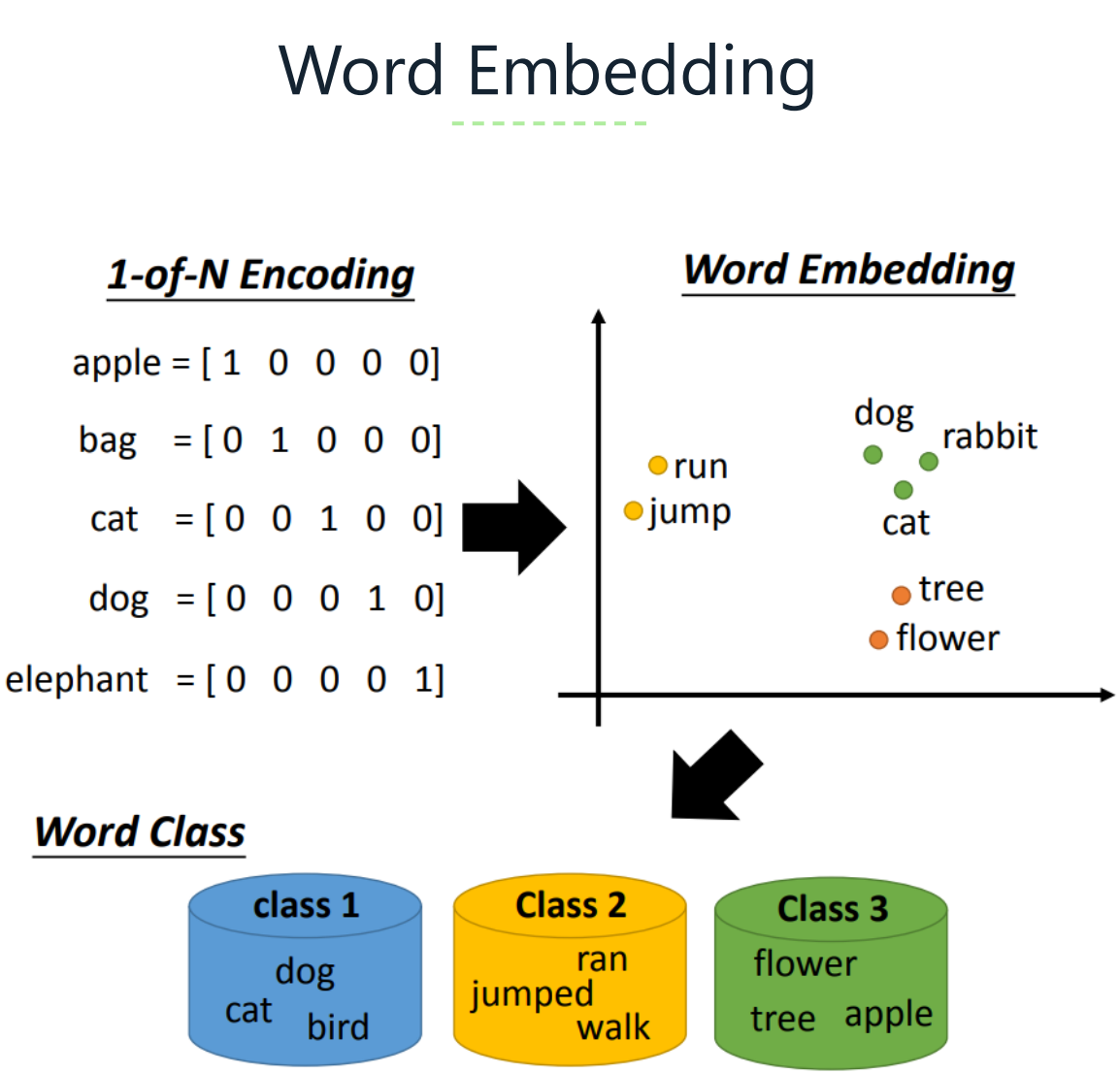
离散表达的问题

- 作为资源很不错，但是缺少细微差别
- 缺少新的词（不可能保持最新状态）
- 是主观的
- 需要使用人力来创造和更新
- 很难计算准确的词之间的相似度

绝大多数基于规则和统计的NLP工作将单词作为原子符号：hotel 啥的

从符号表达到分布式表达

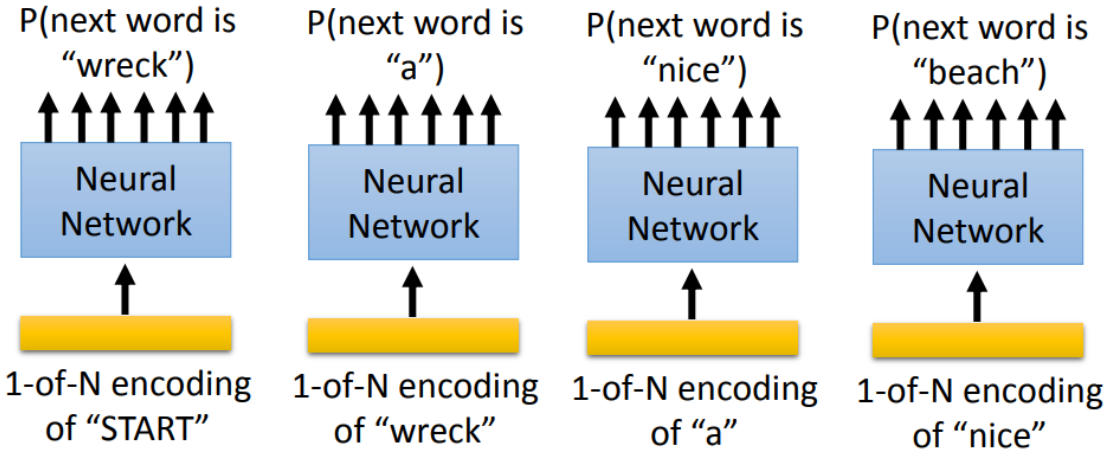
- 使用 one-hot 编码
- 问题是 one-hot 向量中没有自然语义中相似度的概念（比如 motel 和 hotel 相交等于 0）
- 可以单独处理相似性，但是不如直接找一个可以表达相似性的编码方式



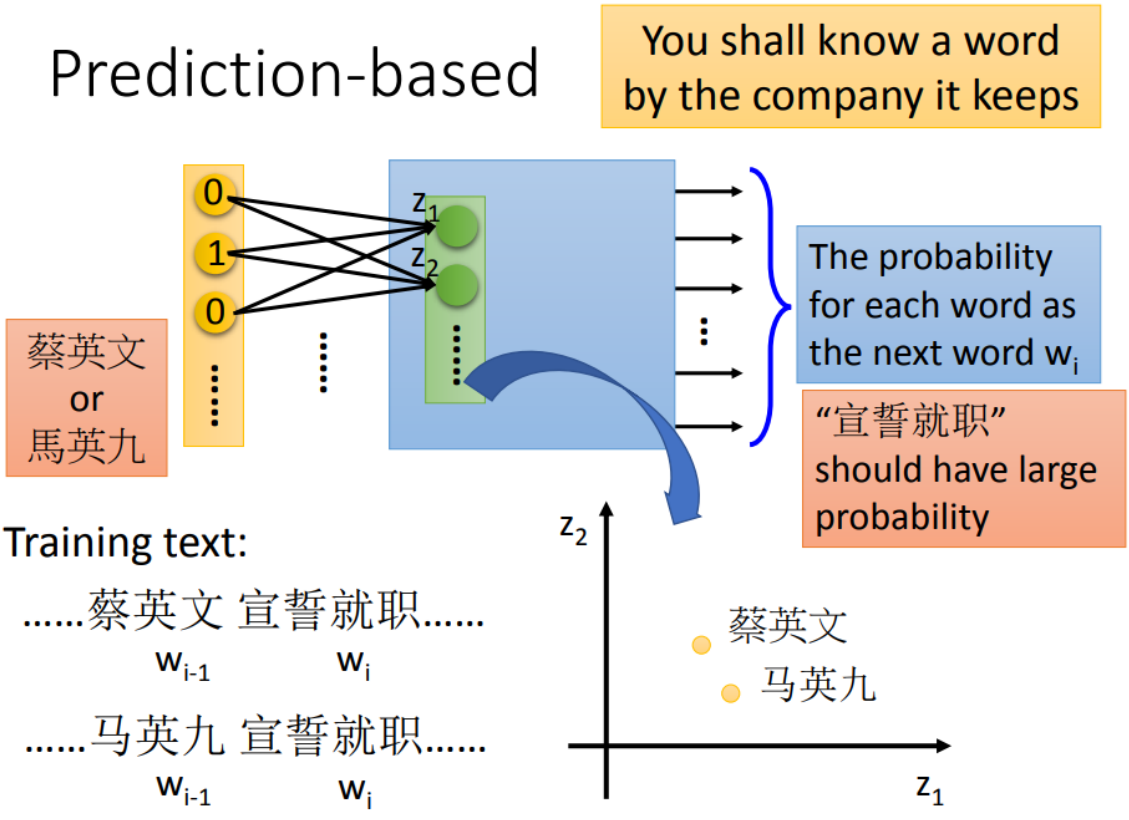
一个词可以通过他的上下文来理解

- count based
  - 如果两个词经常一起出现，那么他们的向量表示的应该相近
- prediction based

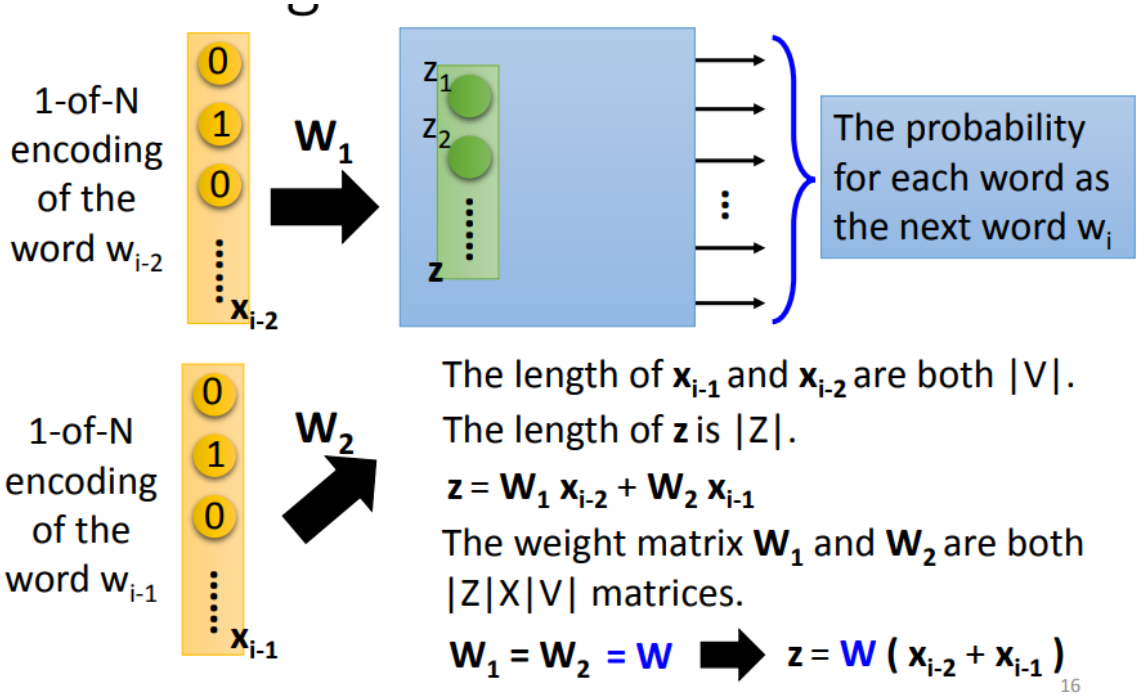
P("wreck a nice beach")
=P(wreck|START)P(a|wreck)P(nice|a)P(beach|nice)
P(b|a): the probability of NN predicting the next word.



取出第一层神经元的输入，用它来表示一个词

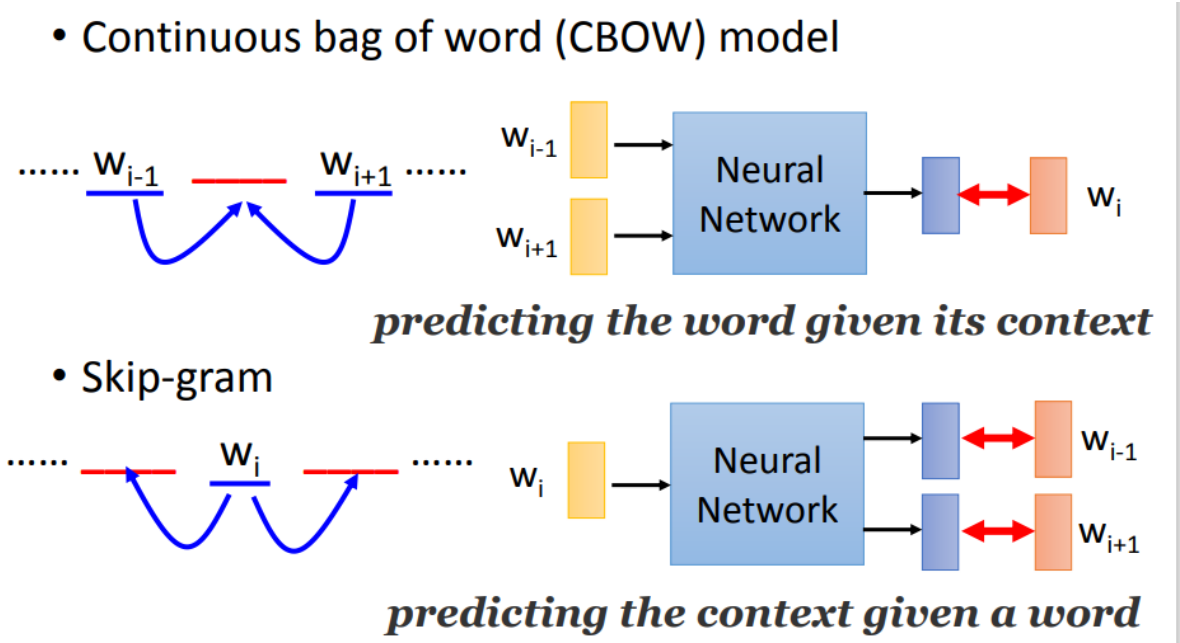


参数共享



不同的架构

- 给出上下文，预测当前的词
- 给出当前的词，预测上下文



Beyond Bag of Word

相同长度的向量表达不同长度的词序列

- 向量用来表示词序列的意思
- 一个词序列可以是一个文档或段落
- 序列中词的顺序是不能被忽略的