异步社区
人民邮电出版社

表示学习

讲师：Houye

目录

CONTENTS

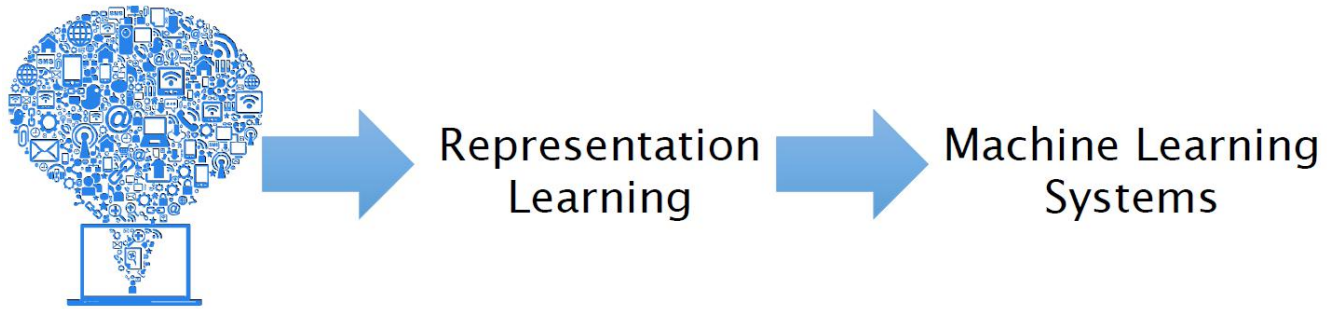
01 表示学习

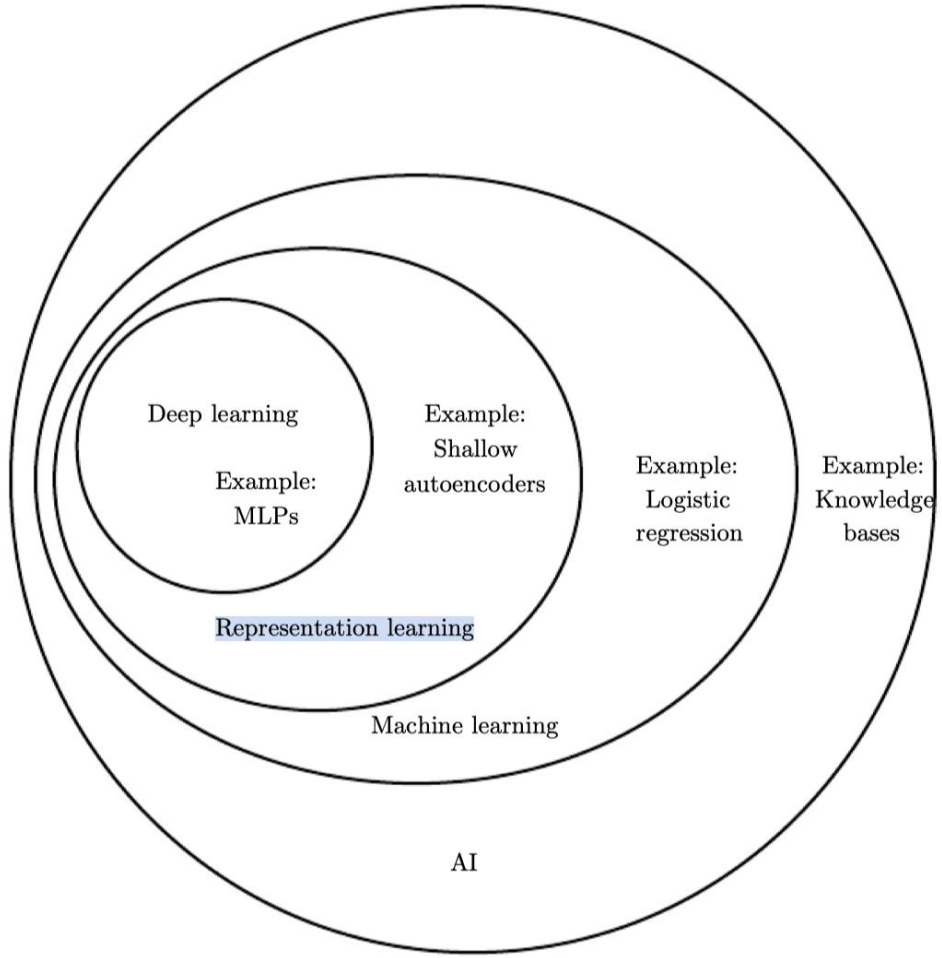
表示学习

表示 & 表示学习 & 机器学习 & 深度学习

ML = Representation + Objective + Optimization

Good Representation is Essential for
Good Machine Learning





Yoshua Bengio. Deep Learning of Representations. AAAI 2013 Tutorial.

表示学习顶级会议ICLR

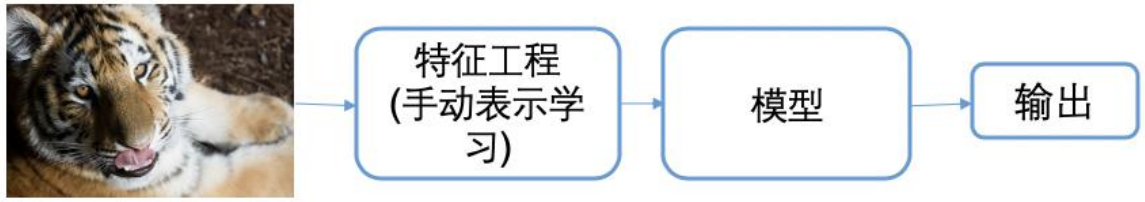
International Conference on Learning Representations

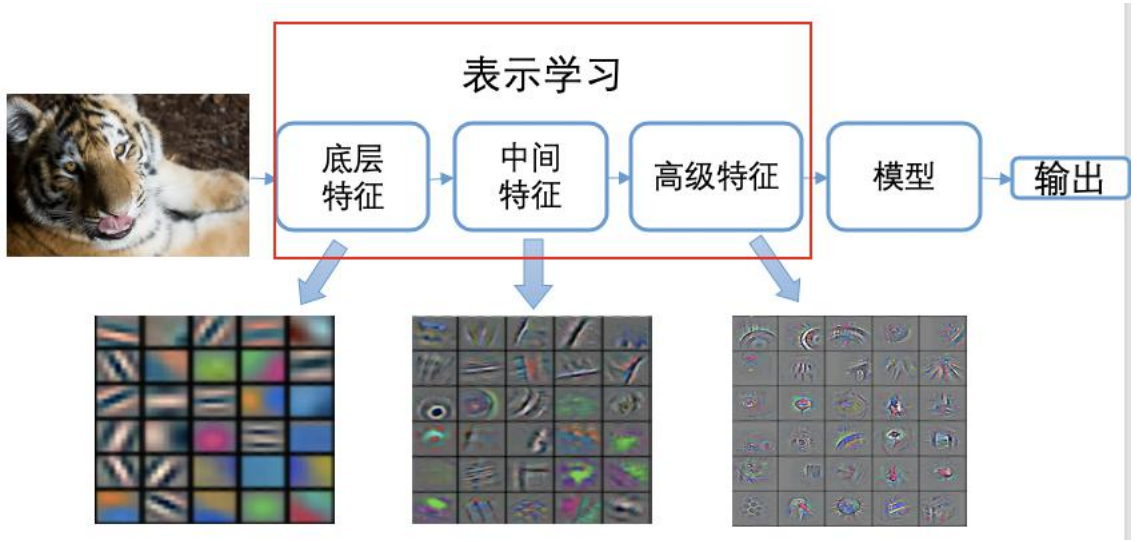
(Yoshua Bengio　　Yann LeCun)

表示学习

表示学习

- 手动表示
  - 人工筛选一些特征对图片/文本/语音进行表示
  - 人工智能：有多少人工，就有多少智能。
- 自动表示学习(end-to-end)
  - 用算法来自动学习图片/文本/语音表示
  - 深度学习的兴起就是因为其自动学习表征的能力。





定义:

表示学习旨在学习一个映射，将离散的目标(一个单词，一张图片)映射为向量空间的一个点(d维向量)

注：通常说的X Embedding， X2vec也是表示学习

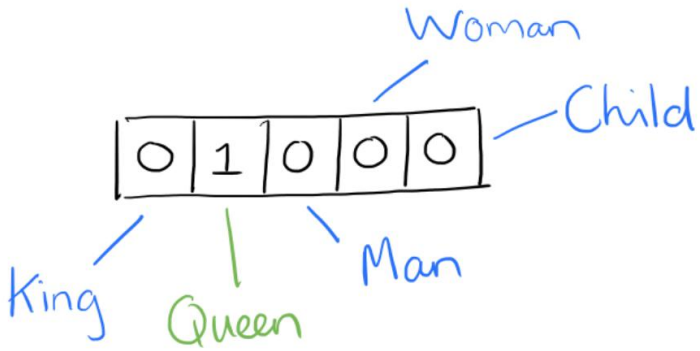
单词表示学习：Word Embedding = Word2vec = Word Representation Learning
图表示学习： Graph Embedding = Graph2vec = Graph Representation Learning

表示学习

(自然语言处理)单词表示学习的2种常见方式

One-hot Encoding

- 对单词进行编号表示，假设有V个单词，则可以用1，2，3，...，V来表示它们。
- 单词序号 转为 V维向量表示，该向量只有一个位置(节点序号)为1，剩下全为0.
- 优点：
  - 无需模型学习就可以生成单词表示，简单高效。
- 缺点
  - 随着单词数量V的增加，向量的大小不断变大，计算代价越来越高。
  - One-hot Encoding只是简单标识/区分了单词，无法反应单词背后的含义。

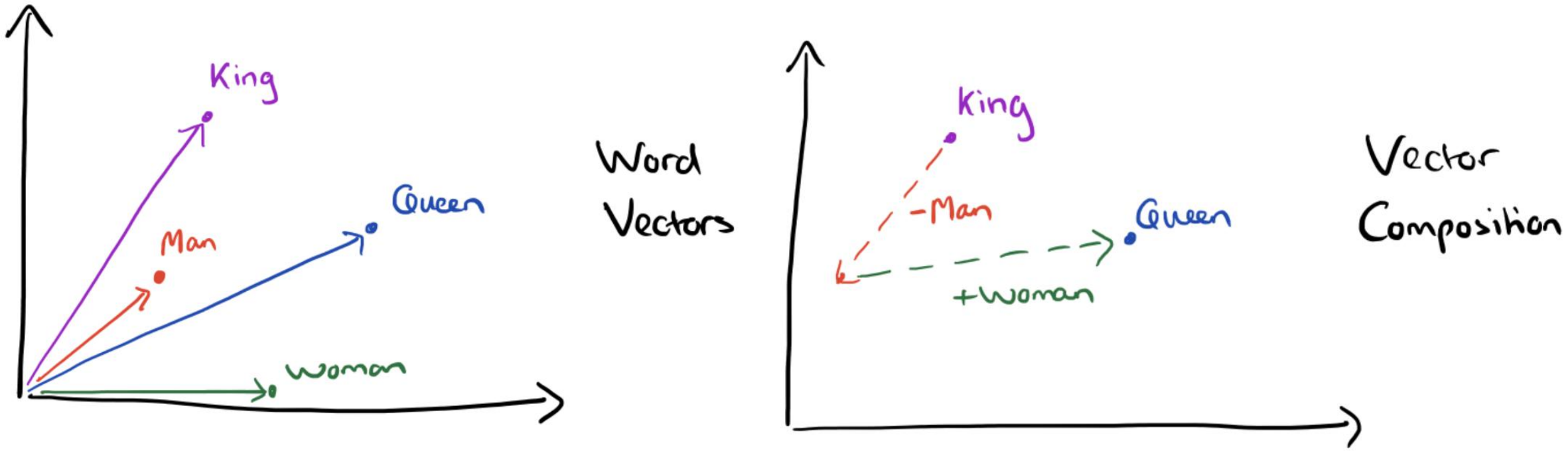


Word2vec (Word Representation Learning)

- 对单词进行编号表示，假设有V个单词，则可以用1，2，3，...，V来表示它们。
- 通过映射函数(通常是神经网络)将单词映射为d维度向量表示，并随着模型优化过程进行学习。所有单词的表示构成一个矩阵，大小为$V * d$。
- 优点：
  - d是一个超参数，和单词总数V没有关系。可以处理超大词表。
  - d维表示可以很好的反映单词含义。相似的单词，其d维表示在向量空间比较近。
- 缺点：
  - d的大小需要手工指定。太大，浪费资源。太小，无法充分描述单词的特点。
  - 需要模型训练过程。

单词表示学习Word2vec确实可以反映单词(Word)蕴含的意思吗?

King – Man = Queen – Woman



02 Word2vec

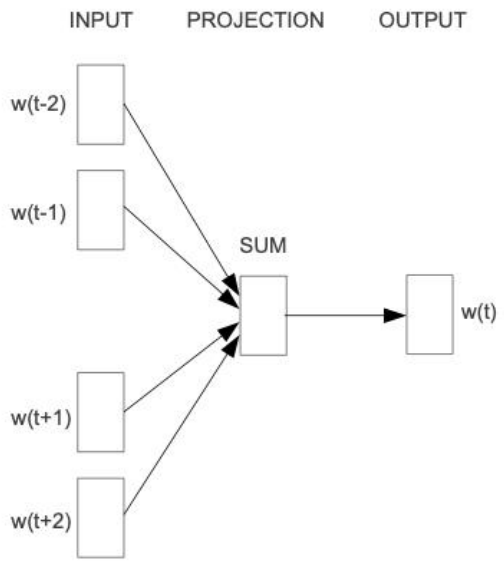
Word2vec

Efficient estimation of word representations in vector space

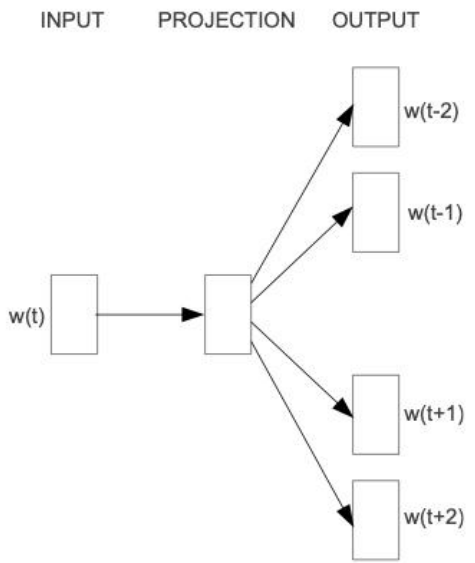
T Mikolov, K Chen, G Corrado, J Dean - arXiv preprint arXiv:1301.3781, 2013 - arxiv.org

We propose two novel model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. We observe large improvements in accuracy at much lower computational cost, ie it takes less than a day to learn high quality word vectors from a 1.6 billion words data set. Furthermore, we show that these vectors …

☆ 99 Cited by 17405 Related articles All 38 versions »



CBOW



Skip-gram

Word2vec

Word2vec的核心假设：**出现在相邻位置的单词的含义具有相似性。**

以单词“stars(星星)”为例

- 星星通常与shining(闪亮)，bright(明亮)，night(夜晚)一起出现，它们的含义也比较相似。
- 星星很少与tree(树木)一起出现(几率较低，但大于0)，它们的含义也差异较大。

Construct vector representations

	shining	bright	trees	dark	look
stars	38	45	2	27	12

```
he curtains open and the stars shining in on the barely
ars and the cold , close stars " . And neither of the w
rough the night with the stars shining so brightly , it
made in the light of the stars . It all boils down , wr
 surely under the bright stars , thrilled by ice-white
sun , the seasons of the stars ? Home , alone , Jay pla
m is dazzling snow , the stars have risen full and cold
un and the temple of the stars , driving out of the hug
 in the dark and now the stars rise , full and amber a
bird on the shape of the stars over the trees in front
 But I could n't see the stars or the moon , only the
 they love the sun , the stars and the stars . None of
r the light of the shiny stars . The plash of flowing w
man 's first look at the stars ; various exhibits , aer
rief information on both stars and constellations, inc
```

Similarity in meaning as vector similarity

Word2vec的核心假设：**出现在相邻位置的单词的含义具有相似性。** 等价于，**单词word与其上下文context可以互相预测** (上下文context指的就是出现在相邻位置的的单词。)

给定单词序列$[w_1, w_2, \ldots, w_{t-c}, \ldots, w_t, \ldots, w_{t+c}, \ldots]$并限定上下文范围$c$， 单词$w_t$的上下文context是$[w_{t-c}, \ldots, w_{t+c}] \ / w_t$.

CBOW： 基于上下文context来预测当前词word，即 $\Pr(w_t \mid \text{context}(w_t))$

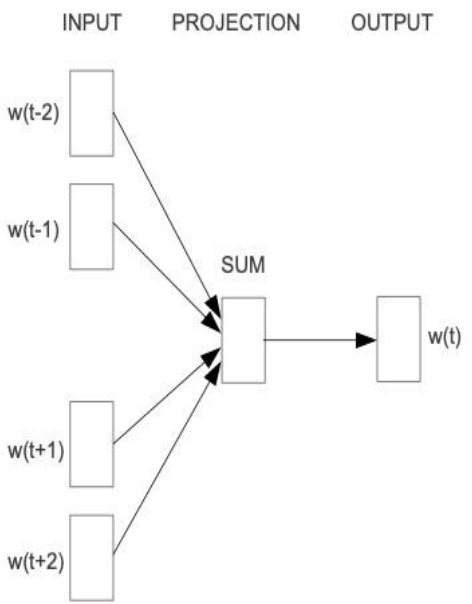
- 基于[of, the, shiny, the plash, of]来预测stars

Skip-gram：基于当前词word来预测其上下文context，即 $\Pr(\text{context}(w_t) | w_t)$

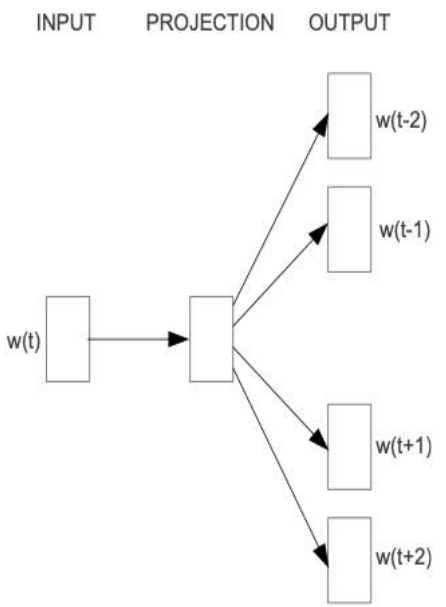
- 基于stars来预测[of, the, shiny, the plash, of]

Stars的上下文，c=3

r the light of the shiny stars . The plash of flowing w



CBOW



Skip-gram

Word2vec

CBOW: $\Pr(w_t \mid \text{context}(w_t))$

- 将所有单词映射为一个d维表示，$w_t \rightarrow v_{w_t}$
- 获取w_t上下文表示$\tilde{\boldsymbol{v}}_{w_t}$ (实际就是上下文里的单词的表示进行平均)

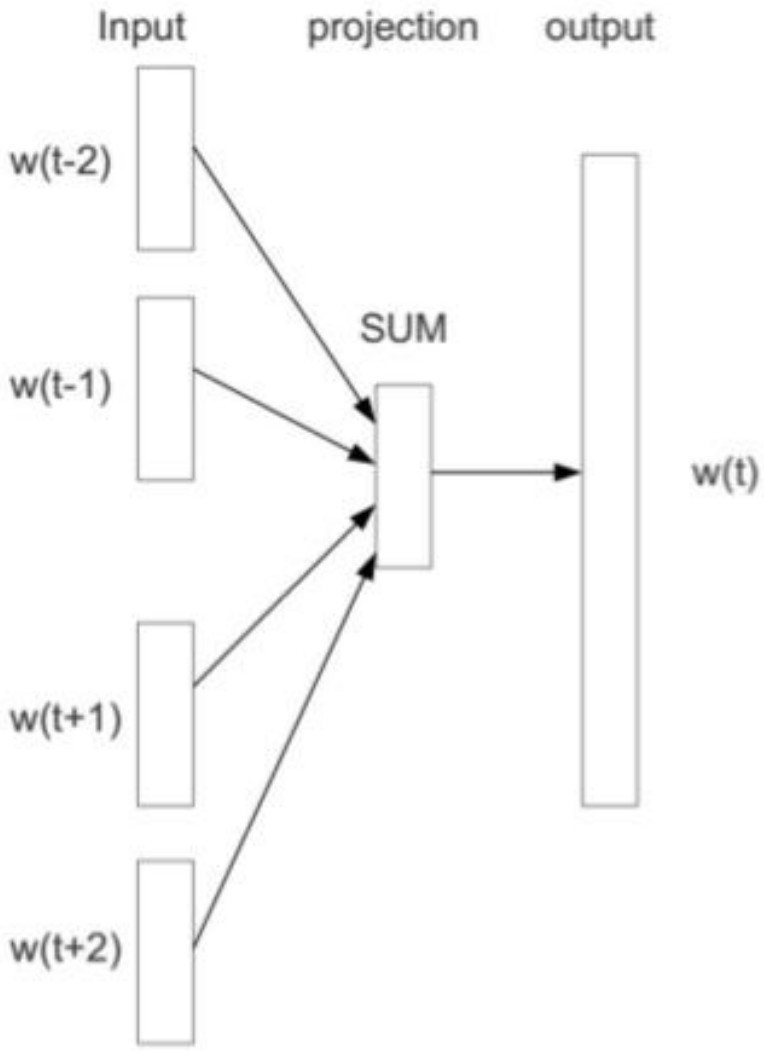
$$\tilde{\boldsymbol{v}}_{w_t} = \frac{\sum_{i=t-c}^{t+c} v_{w_i}}{2c} \quad (i \neq t)$$

- 预测.

$$\Pr(w_t \mid \text{context}(w_t)) = \frac{\exp(\text{sim}(\tilde{v}_{w_t}, v_{w_t}))}{\sum_{w'} \exp(\text{sim}(\tilde{v}_{w_t}, v_{w'_t}))}$$

这里基于一个假设:

$$\text{sim}(\tilde{\boldsymbol{v}}_{w_t}, \boldsymbol{v}_{w_t}) > \text{sim}(\tilde{\boldsymbol{v}}_{w_t}, \boldsymbol{v}_{w'_t})$$



Word2vec

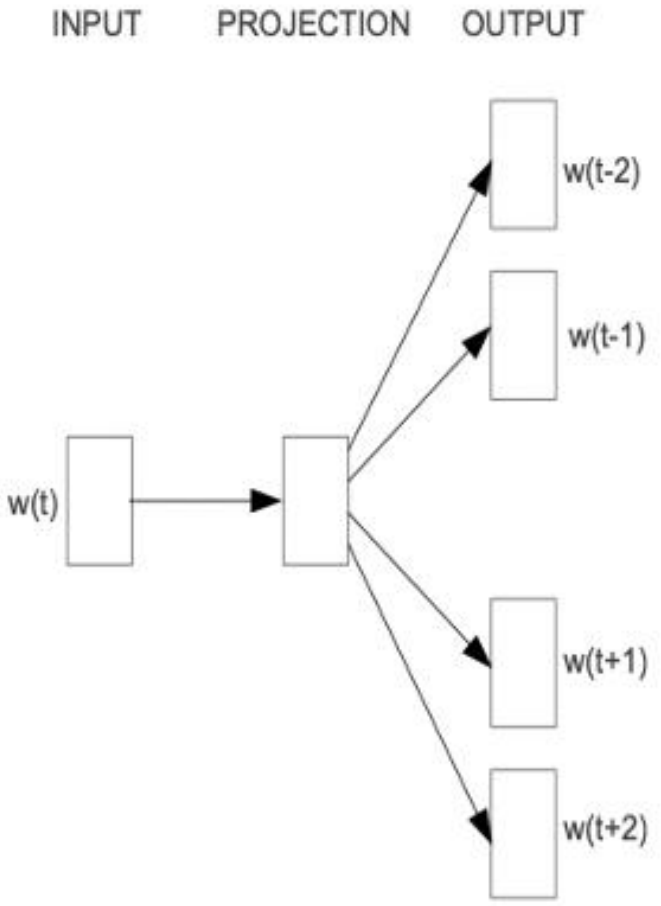
Skip-gram $\Pr(\text{context}(w_t)|w_t)$

- 将所有单词映射为一个d维表示，$w_t \rightarrow v_{w_t}$
- 获取w_t上下文中某个单词的表示表示$\boldsymbol{v}_{w_i}$，$w_i \in context(w_t)$
- 预测. 这里分2步
  - 预测上下文中的一个单词

$$\Pr(w_i \mid w_t) = \frac{\exp(\text{sim}(v_{w_t}, v_{w_i}))}{\sum_{w'} \exp\left(\text{sim}\left(v_{w_t}, v_{w_i'}\right)\right)}$$

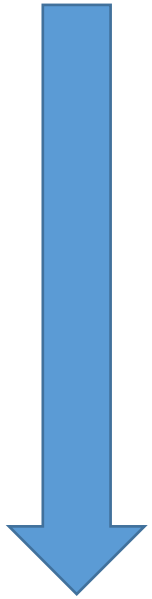
  - 预测整个上下文

$$\Pr(\text{context}(w_t)|w_t) = \prod_{w_i \in context(w_t)} \Pr(w_i \mid w_t)$$



Skip-gram

Word2vec



Node2vec

he curtains open and the stars shining in on the barely
ars and the cold , close stars " . And neither of the w
rough the night with the stars shining so brightly , it
made in the light of the stars . It all boils down , wr
surely under the bright stars , thrilled by ice-white
sun , the seasons of the stars ? Home , alone , Jay pla
m is dazzling snow , the stars have risen full and cold
un and the temple of the stars , driving out of the hug
in the dark and now the stars rise , full and amber a
bird on the shape of the stars over the trees in front
But I could n't see the stars or the moon , only the
they love the sun , the stars and the stars . None of
r the light of the shiny stars . The plash of flowing w
man 's first look at the stars ; various exhibits , aer
rief information on both stars and constellations, inc