

word2vec

将词/字 \rightarrow 向量

例: 今天天气真好, 我们一起去郊游吧!

① 分词 (jieba 工具包)

今天 / 天气 / 真好, 我们 / 一起 / 去 / 郊游 / 吧!

② 准备三大参数

word2id		id2word		word2onehot
今天 0		0 今天		今天 [1 0 0 0 0 ...]
天气 1		1 天气		天气 [0 1 0 0 0 ...]
真好 2		2 真好		真好 [0 0 1 0 0 ...]
⋮		⋮		⋮

③ 设计模型

常见的模型有 CBOW, skip-grow

今天天气真好 word-size = 5000, embedding-size = 50

skip-grow 训练过程: 当前值 \rightarrow 其它值

