

skip_gram_formulas

词语丢弃率（除去一些高频词）

$$P(w_i) = 1 - \sqrt{\frac{t}{f(w_i)}}$$

损失函数

$$-\log \sigma(u_{w_o}^\top v_{w_l}) - \sum_i^N \mathbb{E}_{w_i \sim P_n(w)} \log \sigma(-u_{w_i}^\top v_{w_l})$$

负采样相关公式：单词分布情况

$$P(w_i) = \frac{f(w_i)^{3/4}}{\sum_{j=0}^n \left(f(w_j)^{3/4}\right)}$$

大纲

准备工作

- 文本准备
- 参数设置
- 文本预处理
- 单词分布
- 训练数据准备

模型定义

损失函数定义

训练过程

可视化过程