

语言是什么？

计算机如何理解语言？

什么是自然语言处理？

基本的转换

分词，词干提取，词形还原

用于提取
词根去词缀 去除 = ed 等词缀变化

N-grams (N元模型)

转换方法

TF-IDF

词嵌入

1. 语言模型和特征学习技术具有统一性。
2. 词语映射到实向量。词 \rightarrow 特征空间

表示方法

1. 词袋失去词与词顺序。
2. 稀疏性

嵌入矩阵

词 - one hot

循环神经网络(RNN)

Recurrent Neural Network

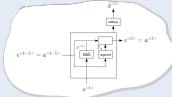
一对多 音乐生成(输入一个生成词接下来词)
多对1 分类问题
多对多 机器翻译

RNN的局限性

单句 / 无法捕捉 长期依赖关系

门控循环单元 (GRU)

增加更新门



LSTM

更新门 + 忘记门

可保留 $t-1$ 时刻旧值
并在 t 时刻向其添加值