

- NLP中的一些神经网络
- 【论文讨论】简单翻译了一下论文中提到的特征
- 预处理中因为包安不上导致的一个奇思妙想
- 句子语义相似度预测（基本规则的探索）——baseline

快速发帖

【论文讨论】简单翻译了一下论文中提到的特征

138-王泽群-NLP LV2

1 · 3 年前

作文评分的特征

标签（空格分隔）： 机器学习 NLP 好未来

我简单翻译了一下，可能有些我没看明白.....大家不要喷谢谢！
也请不要转载=。=内部交流即可
来自https://www.researchgate.net/publication/278383803_Task-Independent_Features_for_Automated_Essay_Grading

Length Features

文章长度 与作文的给定写作时间有直接关系

作文都是在有限时间内写完的，所以长度可以有效表征作者的写作能力进而表征作文的质量。

计算token数和sentence数

平均句子长度和词汇长度 与任务弱相关

表征作者驾驭复杂文本的能力。

Occurrence Features

逗号、引号、感叹号的出现 与任务弱相关

有价值的特征。

正式引用 和任务强相关（对给材料的作文）

直接引语和间接引语

转述动词等规则区分是否为引语

（对给材料的作文）核心概念 仅给材料时使用

识别核心概念：对材料中高tf.idf的词汇或者短语

Syntax Features

句法变化 与任务弱相关

不同的句法树的占比，平均树高
时间和原因状语从句的比例，时间和原因状语从句的引导词

Style Features

formality F 与任务强相关

$$F = \frac{\sum_{i \in A} \frac{c(i)}{n} - \sum_{j \in B} \frac{c(j)}{n} + 100}{2}$$

A= {N, ADJ, PP, DET}, B= {PR, V,ADV, UH}, and n is the number of tokens in the text

type-token-ratio

用以估计作者词汇量的大小

average word frequency 与任务弱相关

作者词汇量的丰富程度与语料库中词汇的频率关系很大.作者使用的词汇在语料库中的频率越小,作者的语言熟练度越高

We model this idea by calculating the average word frequency in the Web1T-corpus (Brants and Franz,2006).

Cohesion Features

文章结构 与任务弱相关

连接词的使用(like therefore or accordingly)表征了文章的一致性

用一个fixed list来计数连接词的使用数, 用文章的token总数来normalize

Coherence Features 与任务弱相关

measuring the topical overlap between adjacent sentences. We use similarity measures based on n-gram overlap and redundancy (e.g. of nouns).

Error Features 与任务弱相关

语法和拼写错误是一篇烂文的明显标志, 但被发现在作文评分中影响并不大.

基于规则的简单语法错误检测

Readability Features 与任务弱相关

非常简单或非常复杂的写作方式都是烂文的标志.

rely on normalized counts of words, letters, syllables or other phenomena(like abbreviations) which affect the readability.

high scoring essays might be associated with different levels of readability

Set-Dependent Features

n-grams features和 POS n-grams features

👍 赞一下

🔗 分享

评论

138-王泽群-NLP LV2 🏆 1 · 3 年前 · #1楼

该评论已被用户删除

138-王泽群-NLP LV2 🏆 1 · 3 年前 · #2楼

该评论已被用户删除



发布评论

B

I

H

“

≡

≡

≡

✂

</>

x²

😊

👁

📄

✖

帮助

LaTeX工具

表格工具

lines: 1 words: 0 Autosaved: 10:07 pm

发布



和鲸社区公众号

- 工具

K-Lab
- 内容

项目

数据集

专栏

专区
- 实战

比赛

众包

HeyWhale 和鲸

商务合作