NLP技术可行性分析报告

复旦大学

2021.08.31

材料纲要：

1. 电信业务场景难点分析（数据、结构和语义）

数据混乱，结构不清晰，类别过多

1. 目前文本分类的主流算法分析

列举分析目前bert等词向量做不好分类任务的原因

引出数据结构化的意义

1. 类脑智能算法在NLP领域的应用研究

3.1类脑智能算法简介

以脑造脑

3.2语言建模与结构化表示

语言的分布式表示模型

3.3基于类脑智能算法的预训练模型

以三篇论文为主线，果蝇跑nlp，分析深度学习算法与大脑信息处理的不同

以神经元活动为基础，对Bert进行调参

3.4 低资源自然语言处理算法

少样本学习

3.5 提升模型可解释性研究

可解释性学习

1. 类脑智能算法的预期效果

#用电信数据做论证，现在能达到的水平和预期效果，预估

#口语化语料的处理，多分类任务

#结合，模型更新，避免标注

#类别信息

两段式研究方法：非结构化数据——句法解析——语义解析——语义理解

4.1句法解析在结构分析中的作用

4.2句法解析在语义理解中的作用

4.3句法解析的难点，应用和问题和原因

4.4我们的理论改进

* 1. 句法解析可能的有效性（基于传统句法解析、替代性方案）
  2. 句法解析可能的有效性（基于我们理论的模型）

4.5 对语义理解的思考及后续的研究

5.报告总述

1. 电信业务场景难点分析
2. 文本分类主流算法

文本分类是文本处理中的经典问题，应用非常广泛。例如垃圾过滤、新闻分类、词性标注等等。文本分类的核心方法为首先提取分类数据的特征，然后选择最优的匹配，从而分类。但是文本也有自己的特点，根据文本的特点，文本分类的一般流程为：1.预处理；2.文本表示及特征选择；3.构造分类器；4.分类。

通常来讲，文本分类任务是指在给定的分类体系中，将文本指定分到某个或某几个类别中。被分类的对象有短文本，例如句子、标题、商品评论等等，长文本，如文章等。分类体系一般人工划分，例如：1）政治、体育、军事 2）正能量、负能量 3）好评、中性、差评。因此，对应的分类模式可以分为：二分类与多分类问题。

2.1主流算法

2.2列举分析目前bert等词向量做不好分类任务的原因

2.3引出数据结构化的意义

1. 类脑智能算法在NLP领域的研究

3.1类脑智能算法简介

以脑造脑

3.2语言建模与结构化表示

语言的分布式表示模型

3.3基于类脑智能算法的预训练模型

以三篇论文为主线，果蝇跑nlp，分析深度学习算法与大脑信息处理的不同

以神经元活动为基础，对Bert进行调参

3.4 低资源自然语言处理算法

少样本学习

3.5 提升模型可解释性研究

可解释性学习

1. 类脑智能算法的预期效果

4.1句法解析在结构分析中的作用

4.2句法解析在语义理解中的作用

4.3句法解析的难点，应用和问题和原因

4.4我们的理论改进

* 1. 句法解析可能的有效性（基于传统句法解析、替代性方案）
  2. 句法解析可能的有效性（基于我们理论的模型）

4.5 对语义理解的思考及后续的研究