

CCL2018参会总结

刘焕勇

中国科学院软件研究所

2018-10-24

主旨报告-脑科学与人工智能

- 如何从脑科学研究中找到更好的人工智能语言模型？
- 目前脑科学研究中研究的脑活动是短期时间段内的一种活动，而认知科学是一个长时间的过程，所以这种时间上的差异，是否真的能够类推过去
- 不同物种的大脑其实有很大差异，目前脑科学的研究只在小鼠上做实验，这种实验的结果是否对人工智能有实际意义？
- 大脑处理语音和音乐信号，为音乐为什么有需求给出了答案，音乐可以取悦人的大脑

主旨报告-理性主义与经验主义

- 从裁掉语言学家到联合语言学家
- 语言计量中的应用
 - 词汇语义变迁
 - 词汇首次出现

深度学习在NLP

- 词向量
 - One-hot、Svd、Wordvec、Glove、ELMO
- 词向量修正
 - 基于可靠词汇约束的词向量修正
 - 基于情绪约束的词向量修正
 - 基于上下文约束的词向量修正
- 语义向量
 - Transformer
 - Self-Attention

语义体系&语义资源库

- AMR体系
- Wordsketch
- 复合名词短语的语义关系体系，名词短语成分中名词之间的关系
- 依存语义体系

评测-客服领域用户意图分类

- 三类业务类型分类
- 若干小类意图分类
- 数据集：10万

评测-中文隐喻识别与情感分析

- 中文动词隐喻的识别
 - 动词隐喻、名词隐喻、负例三分类问题
 - 训练集4394条，测试集1100条
 - F1值: 0.832727
- 中文隐喻的情感分析
 - 乐、好、怒、哀、惧、恶、惊共七分类问题
 - 训练集3630条，测试集909条
 - Macro-F值: 0.41180

评测-中文文本蕴涵

- 中文文本蕴含分类
 - 输入“前提句Premise”和“假设句Hypothesis”，判断蕴含类别括：
 - 蕴含(E, entailment)：P能合理推理得到H, 反之不一定；
 - 矛盾(C, contradiction)：P和H不能同时成立；
 - 无关(N, neutral)： P和H不存在必然的联系。
- 训练级：
 - 8万训练集，1万开发级，1万测试集，共10万数据集

评测-中文幽默计算

- 中文幽默类型识别

- 谐音幽默、谐义幽默、反转幽默共3类标签分类任务
- 训练级7298条，测试集1825条
- Macro-F值: 0.65748

- 中文幽默等级划分

- 强幽默、弱幽默共2类标签分类任务
- 训练级8046条，测试集2012条
- F1值: 0.62723

评测-“讯飞杯”中文机器阅读理解评测

- 输入
 - 篇章和问题
- 输出
 - 篇章中连续片段作为答案
- 数据
 - 中文维基百科类文本，人工标注，共19071个问题
 - 设置训练集、开发集、测试集、挑战集

有趣的题目

- 引语识别
- 中文冗余表达识别
- 美猴王心魔已死-基于情感分析的西游记分析
- 领域情感词vs通用情感词的效果对比
- 足球新闻自动生成
 - 神经网络抽取重要句子+数据模板生成+篇章结构自定义（新闻四个部分）
- 证券市场股东画像构建
 - 计量上的特征：时序特征，金额特征
- 主题模型古典乐器诗词文本挖掘
 - LDA+NMF, topic2author
- 文本可读性研究综述
 - 文本可读性特征+公式、分类、排序

有趣的题目

- 反问句识别
 - 反问句卷积池化+反问句显式标记
- 自编码器的通用性文本表征
 - 基于公开数据集训练自编码器
- 双关语识别
 - 词向量+同义词融合
- 长篇小说计量统计和分析
 - 词汇丰富度、颜色词（文章色彩度）、情感（文章氛围）
- 基于语言知识资源进行语义理解与推理
 - 基于北大实体概念词典

总结-评测

- 目前的评测任务，都最终转换成一个分类问题
 - 转换成一个分类问题，这是否合适？人脑的处理形式是不是也是分类？
 - 若当做一个分类问题来做，**模型和评测任务**本身是不是要深入去考虑
 - 任务中的评测数据集大部分都**较小**，是否具有泛化能力？
 - 模型越复杂是不是效果越好？
 - 文本蕴涵中第一名成绩比第二名高出4个百分点，但只用了一个CNN

总结-模型

- 水平不够，想法很有趣，创新堪忧
 - 有很多新奇的论文题目，想法很多，但都不够深入
 - 实验室自己标点小规模语料，然后用目前通用的模型，做各个模型之间的效果对比，文章就出来了，换汤不换药
- 对深度学习模型在各个NLP任务中的应用缺乏思考
 - 应该考虑深度学习模型自身的问题，而非只是拿来主义
 - 应该具体深入的去考虑NLP任务本身，分析NLP任务的整个情况
 - 例如文本蕴涵任务，LSTM+attention效果还比不过CNN，这到底是模型的问题，还是任务本身的问题？这个其实需要去考虑

总结-语言资源

- 语言资源库这次是出现了三个
 - 北大发布的中文实体概念词典
 - 北语标的2万条名词性短语语义库
 - 北语标的蕴涵语料库
- 语言资源库的构建四点感想
 - 一是语言资源库标准规范的统一性
 - 二是语言资源库的构建动机，应该以解决语言理解问题而提出
 - 三是语言资源库的共享性，语言资源库应该共享
 - 四是语言资源库的连通。语言资源之间应该打成通路，如目前的HowNet, Cilin, 中文实体概念词典，现代汉语语法词典等，应该联通，联合发力