

教育应用中的自然语言处理

孙琳



个人介绍

- 剑桥大学计算机系博士
 - 自然语言处理
- 校宝在线 CTO
 - Empower education with technology
- 研究：
 - 剑桥大学语言实验室研究员（Affiliated）
 - 发表论文20多篇，
 - 坚持研究：2016在Journal of Computational Linguistics上发表论文
 - 多个学术杂志和会议的审稿人

概览

- 人工智能当中的一个分支
 - 自然语言处理在教育中的应用
- 自然语言处理Natural language processing (NLP)简介
- NLP在教育中的应用
- 未来展望

自然语言处理简介

- 自然语言处理是人工智能的一个重要分支
 - 跟计算机视觉，语音等构成了人工智能研究
 - 也可以被叫做计算语言学(Computational linguistics)
- 尝试让计算机把人类的语言（自然语言）转化为可执行的计算机命令
 - 中文 \leftrightarrow 机器语言
 - 关注的核心是“语言”



自然语言处理 — 难点

- 自然语言处理不是编程语言
 - 自然语言的逻辑性有多强(ruo)?
 - 一句话6种理解方法（重读红色词）：
 - 我从来没说偷过钱
 - 1. 我从来没说偷过钱 => 别人说过，我没说过
 - 2. 我从来没说偷过钱 => 就是从没说过
 - 3. 我从来没说偷过钱 => 我没说，但是我用其他方式表达过
 - 4. 我从来没说偷过钱 => 我说过有人偷过，但不是他
 - 5. 我从来没说偷过钱=> 不是偷，可能借过
 - 6. 我从来没说偷过钱=> 偷的不是钱，是别的东西
 - 7. 继续玩：我从来没说偷过我的钱=>他偷过钱，但不是我的钱

自然语言处理 — 难点

- 不光逻辑，需要极强的知识库
 - they指的是什么？
 - We gave monkeys the bananas because they were hungry. (they指猴子)
 - We gave monkeys the bananas because they were over-ripe. (they指香蕉)
- 组合用法
- 比喻用法和开放性
 - “灌水”，“潜水” (网络用语)

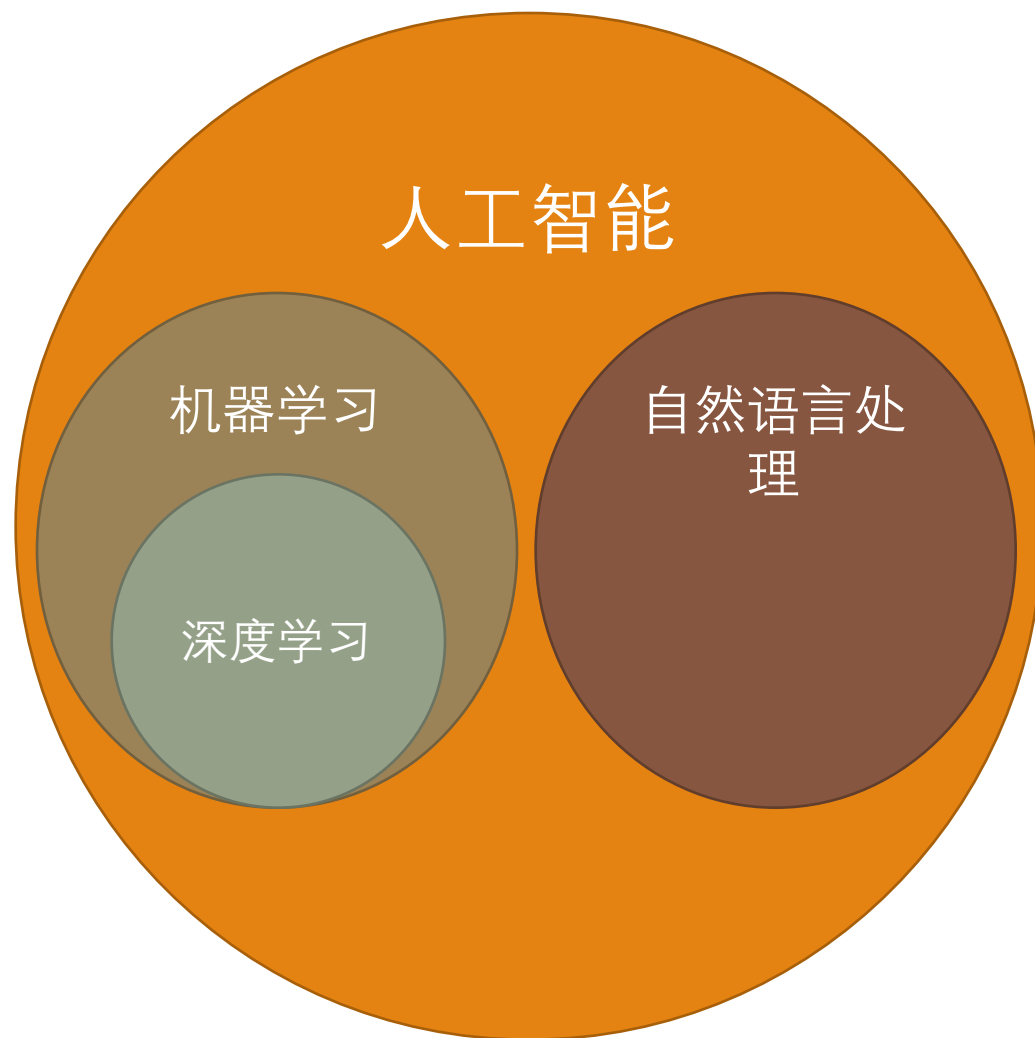
自然语言处理简介

- 常用的方法:
 - 机器学习（深度学习）
 - 规则和逻辑
 - 语言学！
- 常见自然语言处理应用:
 - 搜索引擎
 - 机器翻译
 - 语音识别
 - 问答系统

自然语言处理历史

- 自然语言处理是人工智能领域的开创领域之一
 - 早于视觉，语音和机器学习
 - 始于冷战“太空竞赛”
- 跟人工智能一样，命途多舛
 - 经历多次人工智能寒冬
 - 好不容易迎来了人工智能和深度学习的春天：
 - 摩尔定律停了
 - 深度学习目前也未能解决大部分NLP的问题

自然语言处理和深度学习



自然语言处理和深度学习

- 提供的新的思路和方法
- 深度学习早已经被用到NLP所有的领域当中
- NLP上大部分任务没看到视觉和语音的那种巨大改进
 - 几乎都有改进，但是不大
 - 黑盒问题和语言学理论的关系

自然语言处理 — 现状

- 语音识别系统 (95%)
- 机器翻译系统 (80%)
- 词性标注
 - 句子级别正确率57%
 - 2009-2017 词级别正确率提高了不到1%
- 句法分析
 - 停在了2006年 (Constituent parsing)
 - Google SyntaxNet (Dependency parsing) — 四年只有2%的提升
- 多轮对话 (60%)

NLP的教育应用

跟教育科技契合度最高的AI技术

1. 跟语言教学相关的应用

- 外语和母语教育（自动评分，辅导口语写作等）

2. 教育文本处理

- 教材的编订（CLC），文本阅读分级，文本简化，生成题目

3. 对话系统，使用自然语言进行教学



NLP的教育应用—作文评分

- ETS E-Rater
 - TOEFL(2009), GMAT(1998), GRE (2008)
 - 50% 完全符合, 70%接近
 - 跟人工打分接近
- 校宝在线 1Course
 - 达到ETS E-Rater水平
 - 详细的反馈

| 系统 | 数据 | 人工评分r | 机器评分r |
|----------------|--------|-------|-------|
| IEA (2011) | 美国高中 | 0.91 | 0.91 |
| E-rater (2012) | 托福独立写作 | 0.69 | 0.75 |
| 校宝在线(2014) | 托福独立写作 | 0.88 | 0.89 |

评分 评语

3.5

- 较好地针对话题完成任务, 但是一些论点所用解释、例证和/或细节没有充分展开
- 体现一致、递进、连贯的特点, 虽偶尔出现观点的联系不清, 但不影响文章表达
- 语言有一定的熟练度, 能使用一定量不同类型的句法和词汇。在结构、词汇或习惯用法上存在一定的错误



1COURSE

NLP的教育应用—作文纠错

- 提供作文中的拼写，语法和润色方面的建议
- 提供商：
 - 1Checker (校宝)
 - Grammarly
 - Juku
 - Noun Plus
 - Whitesmoke



NLP的教育应用—作文纠错

- 华南理工大学2016年对系统性能的评测，3年未更新的1Checker个人版依然领先。
- NLP在教育应用中，最令研究人员关注的方向之一

| | Total text error | Right check | Recall ratio |
|----------|------------------|-------------|--------------|
| 1Checker | 231 | 48 | 20.8% |
| Word | 231 | 4 | 1.7% |
| JuKu | 231 | 33 | 14.3% |
| NounPlus | 231 | 28 | 12.1% |

发表在 2016 “Grammar Automatic Checking System for English Abstract of Master’s Thesis “

NLP的教育应用—简答题评分

- 针对有固定答案的简答题
 - 不能批开放型的简答题
 - 可以应用到物理，生物和化学等
 - 翻译题（词汇，句子）
- 原理：
 - 把学生的答案跟标准答案进行比较
 - 语意相关性比对
 - 跟机器翻译的原理类似

69. 你最好每天运动一小时来保持健康。

_____ do exercise for an hour a day to keep fit.

70. 你能在我外出的时候浇花吗？

_____ water the flowers while I am away?

71. 昨天他直到看完这本书才休息。

Yesterday, he _____ he finished reading the book.

72. 学得越多，你就越会意识到自己知道得有多么少。

The more you learn, _____ you know.

73. 为了使北京的空气更洁净，我们要尽量多乘坐公交车。

We should _____ in Beijing.

NLP的教育应用—简答题评分

- Oxford-UCLES
 - 高精度
 - 对于每道题，都要手写规则
- ETS C-Rater
 - Indiana's English 11 End of Course Assessment
 - 阅读理解题型
 - 机器打分84%， 人工打分90.8%

NLP的教育应用—阅读分级

- Lexile 蓝思阅读分级

- 跟词的频率和平均句子长度(LMSL)有关系

- Lexile level= $((9.82247 * \text{词频}) - (2.14634 * \text{句长}) - c + 3.3) * 180 + 200$

- 词汇频率是词汇难度的体现

- 平均句长是语法复杂度的体现

- 中文阅读分级

NLP的教育应用—词汇测试

- 自动生成词汇测试选择题：
- 步骤
 1. 找到合适的句子
 2. 找到合适的词
 3. 生成合理的迷惑项
 - 本地合理，整句不合理

Follow these tips to stay safe during a hurricane.

to stay _____ during

- open
- safe
- quiet
- active

Follow these tips to stay _____ during a hurricane.

- open
- safe
- quiet
- active

Example showing distractor applicability in a narrow vs. full context.

NLP和教育—研究方向

- Workshops : BEA Workshop, NLP Tea, AIED
- 方向
 - 自动纠错 40% 自动打分 30%
- 进度
 - 纠错: 40-60%的覆盖率, 依然遥远
 - 打分: 在特定任务上基本达到了人的水平
- 问答系统, 对话系统并不是研究人员关注的方向
- 研究机构
 - ETS, Cambridge assessment 等等

NLP和教育—我的结论

- 很难从 *目前* 深度学习在的成功来推测在自然语言处理领域的成功。
 - 在应用层面（教育）更难
- 未来仍有希望，需要NLP领域有更好的突破（对话系统，问答系统）
- 黑盒问题
 - 模型复杂度和可解释性悖论
 - 教育领域需要的是可解释的模型，便于人来学习。
 - 几十亿参数的端到端模型：语音识别，机器翻译，手写识别
 - 自动作文打分：不仅要分数准确，还要有合理详细的解释



孙琳

18668055056

lin.sun@xiaobao100.com