本课件仅用于教学使用。未经许可，任何单位、组织和个人不得将课件用于该课程教学之外的用途(包括但不限于盈利等)，也不得上传至可公开访问的网络环境

新媒体大数据分析

New Media Big Data Analysis

作业：文献调研

黄振亚，朱孟潇，张凯，唐秀秀

课程主页：
http://staff.ustc.edu.cn/~mxzhu/Course/NNM2026.html

助教：谭施越 朱嘉骏
bigdata_20262026@163.com

课程要求与考核方式

- 课程要求
  - 1. 课堂出勤，平时作业
  - 2. 文献调研报告
  - 3. 实验任务（需要编程），之后布置给大家
  - 4. 课程交流与课程汇报

- 考核方式
  - 课堂出勤（30%）+调研报告（30%）+实践报告（40%）
  - 结课：第18周

任务一：文献调研

- 结合本学期上课的内容，调研相关的文献并撰写报告
  - 文献可以从模型、应用、平台、产品等任一角度进行调研，调研的内容尽可能广，并将学到的内容和心得以报告形式提交
  - 建议：大数据（数据分析）+ 本人学科专业
  - 语言要求：中英文不限（包括调研的内容）
  - 格式要求：专业、美观
- 报告内容：
  - 题目
  - 调研结果综述
  - 学习心得与思考
  - 参考文献

任务一：文献调研

- 什么是文献调研

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2019.2914054, IEEE Transactions on Pattern Analysis and Machine Intelligence

The Gap of Semantic Parsing: A Survey on Automatic Math Word Problem Solvers

Dongxiang Zhang, Lei Wang, Luming Zhang, Bing Tian Dai and Heng Tao Shen

Abstract—Solving mathematical word problems (MWPs) automatically is challenging, primarily due to the semantic gap between human-readable words and machine-understandable logics. Despite the long history dated back to the 1960s, MWPs have regained intensive attention in the past few years with the advancement of Artificial Intelligence (AI). Solving MWPs successfully is considered as a milestone towards general AI. Many systems have claimed promising results in self-crafted and small-scale datasets. However, when applied on large and diverse datasets, none of the proposed methods in the literature achieves high precision, revealing that current MWP solvers still have much room for improvement. This motivated us to present a comprehensive survey to deliver a clear and complete picture of automatic math problem solvers. In this survey, we emphasize on algebraic word problems, summarize their extracted features and proposed techniques to bridge the semantic gap, and compare their performance in the publicly accessible datasets. We also cover automatic solvers for other types of math problems such as geometric problems that require the understanding of diagrams. Finally, we identify several emerging research directions for the readers with interests in MWPs.

Index Terms—math word problem, semantic parser, reasoning, survey, natural language processing, machine learning

任务一：文献调研

- 什么是文献调研

A Survey of Knowledge Tracing

Qi Liu, Shuanghong Shen, Zhenya Huang, Enhong Chen, *Senior Member, IEEE* and Yonghe Zheng

Abstract—High-quality education is one of the keys to achieving a more sustainable world. In contrast to traditional face-to-face classroom education, online education enables us to record and research a large amount of learning data for offering intelligent educational services. Knowledge Tracing (KT), which aims to monitor students' evolving knowledge state in learning, is the fundamental task to support these intelligent services. In recent years, an increasing amount of research is focused on this emerging field and considerable progress has been made. In this survey, we categorize existing KT models from a technical perspective and investigate these models in a systematic manner. Subsequently, we review abundant variants of KT models that consider more strict learning assumptions from three phases: before, during, and after learning. To better facilitate researchers and practitioners working on this field, we open source two algorithm libraries: EduData for downloading and preprocessing KT-related datasets, and EduKTM with extensible and unified implementation of existing mainstream KT models. Moreover, the development of KT cannot be separated from its applications, therefore we further present typical KT applications in different scenarios. Finally, we discuss some potential directions for future research in this fast-growing field.

Index Terms—Knowledge Tracing; Intelligent Education; Educational Data Mining; Adaptive Learning; User Modeling

9 Feb 2023

任务一：文献调研

- 可以通过以下途径了解更多更前沿的研究领域
  - 期刊及其子刊：
    - Nature, Science, IEEE/ACM Transactions等
  - 国际/国内会议
    - 《中国计算机学会推荐国际学术会议和期刊目录》
    - 《清华大学计算机学科推荐学术会议和期刊列表》
    - 《中国科大计算机学院学位分委会认定的顶级会议与期刊》
    - 《信息与智能学部学位分委员会认定的A档/B档会议与期刊》
    - 例如，KDD, WSDM, ICDM, AAAI, IJCAI, SIGIR, CIKM, NIPS, ICML, VLDB, SIGMOD, ICDE, ACL,EMNLP, WWW
  - 各种科技论坛、公众号
    - 机器之心，新智元，专知
  - 在线课程

任务一：文献调研

- 例如，从模型角度出发：
  - 采样方法
    - MCMC - Metropolis-Hastings
    - MCMC - SLICE SAMPLING
  - 分类
    - 决策树
    - 支持向量机(SVM)
    - 贝叶斯分类器
  - 聚类
    - K均值算法(K-means)
    - 基于密度的聚类(DBScan)
    - 层次聚类
  - 深度学习
    - 卷积神经网络(CNN)
    - 循环神经网络(RNN)
  - 强化学习
    - 马尔科夫决策过程(MDP)
    - 深度Q网络(DQN)
  - 大模型
    - 文本大模型（ChatGPT）
    - 视觉大模型（GPT4V）
    - 国产大模型（讯飞星火等）

任务一：文献调研

- 例如，从应用角度出发
  - **教育**
    - EKT: Exercise-aware Knowledge Tracing for Student Performance Prediction
    - DisenQNet: Disentangled Representation Learning for Educational Questions
  - **金融**
    - Product Supply Optimization for Crowdfunding Campaigns
  - **推荐系统**
    - Relevance meets Coverage: A Unified Framework to Generate Diversified Recommendations
    - Personalized Travel Package Recommendation
  - **社交网络**
    - An Influence Propagation View of PageRank
    - From Footprint to Evidence: An Exploratory Study of Mining Social Data for Credit Scoring
  - ChatGPT，AlphaGO等
  - 大模型在XXX中的应用

文献调研-评分标准

- **多文献**调研报告
  - 独立写作程度如何、是否抄袭，不许大模型生成
  - 是否有明确主题、主题是否符合课程内容、表述是否合适
  - 是否有背景和现状综述
  - **对现状综述的总结是否有框架、有逻辑**
  - 对各文献的单独表述是否清晰具体
  - **是否对各文献涉及模型和方法做有意义的比较分析**
  - 逻辑表达是否通顺
  - 参考文献是否完整
  - 提交是否及时

调研报告

- 时间节点
  - **10月19日完成调研主题和方案（1页纸），邮箱发送助教检查**
  - **11月30日完成调研报告，并提交**
- **文献调研报告 1份**
  - **独立完成，每人一份**
  - **用给的LaTex模板写报告，篇幅至少6页**

助教：谭施越 朱嘉骏
bigdata_20262026@163.com

