

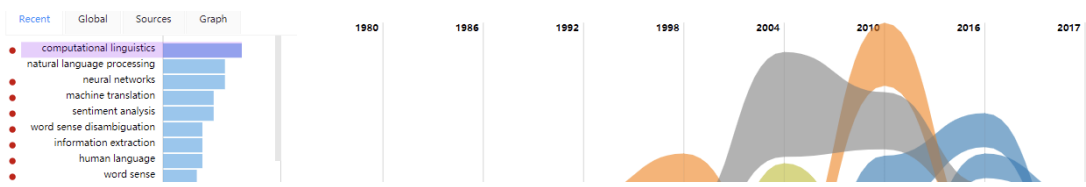
gitbook.cn

GitChat - 《自然语言处理研究报告》的文章

9 ~ 11 分

随着深度学习时代的来临，神经网络成为一种强大的机器学习工具，自然语言处理取得了许多突破性发展，情绪分析、自动问答、机器翻译等领域都飞速发展。

下图分别是 AMiner 计算出的自然语言处理近期热点和全球热点。通过对1994-2017年间自然语言处理领域论文的挖掘，总结出二十多年来，自然语言处理的领域关键词主要集中在计算机语言、神经网络、情感分析、机器翻译、词义消歧、信息提取、知识库和文本分析等领域。旨在基于历史的科研成果数据的基础上，对自然语言处理热度甚至发展趋势进行研究。图中，每个彩色分支表示一个关键词领域，其宽度表示该关键词的研究热度，各关键词在每一年份（纵轴）的位置是按照这一时间点上所有关键词的热度高低进行排序。



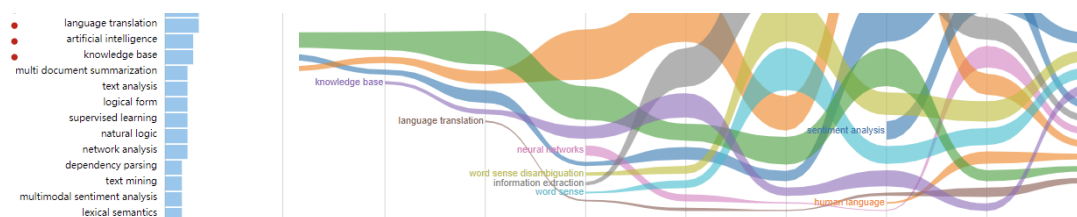


图 14 自然语言处理近期热点图

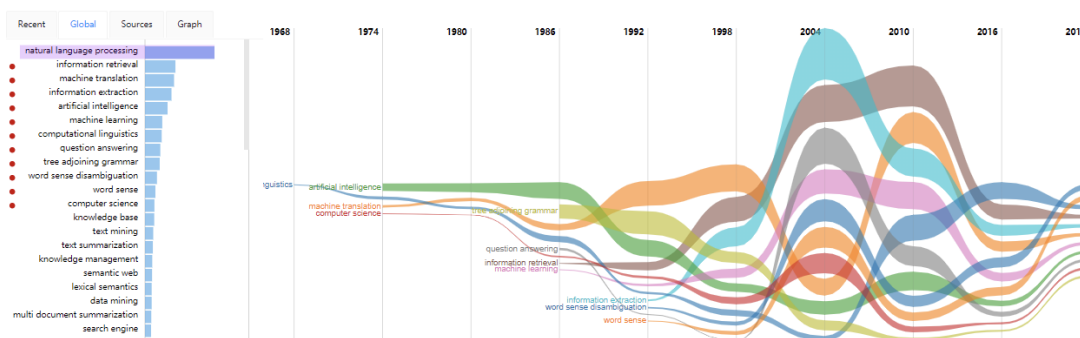


图 15 自然语言处理全球热点图

图14显示，情绪分析、词义消歧、知识库和计算机语言学将是最近的热点发展趋势，图15显示词义消歧、词义理解、计算机语言学、信息检索和信息提取将是自然语言处理全球热点。

我们同时在微博@ArnetMiner 中发起了关于自然语言处理未来发展趋势的投票，得到了如下结果。

文本理解与推理：浅层分析到深度理解

135 (28.1%)

对话机器人：实用化、场景化 83 (17.3%)

NLP 行业：与专业领域结合 74 (15.4%)

学习模式：先验语言知识与深度学习结合

45 (9.4%)

文本情感分析：事实性文本到情感性文本

43 (9%)

语言知识：人工构建到自动构建 25（5.2%）

信息检索：跨语言、多媒体 23（4.8%）

文本生成：规范文本到自由文本 15（3.1%）

NLP 平台化：封闭到开放 13（2.7%）

对抗训练思想的应用 9（1.9%）

共有465人次参与了投票，文本理解与推理由浅层分析到深度理解有135人次支持，占比28.1%，对话机器人实用化、场景化，NLP 行业与专业领域结合，学习模式由先验语言知识与深度学习结合以及文本情感分析由传统媒体到社交媒体依次排列，分别占比17.3%、15.4%、9.4%和9%。我们依据排列由高到低选取其中几项展开介绍。

- **文本理解与推理：浅层分析向深度理解迈进**

Google 等公司已经推出了以阅读理解作为深入探索自然语言理解的平台。文本理解和推理是自然语言处理的重要部分，现在的机器软件已经可以根据文本的语境上下文分辨代词等指示词，这是文本理解与推理从浅层分析向深度理解迈进的重要一步。

- **对话机器人：实用化、场景化**

从最初2012年到2014年的语音助手，到2014年起逐渐出现的聊天机器人微软小冰、百度小度，再到2016年哈工 SCIR-笨笨，对话机器人越来越智能。最初的语音助手可以听得到但是听不懂，之后的对话机器人可以听得懂但是实用性却不强，现在对话

机器人更多的是和场景结合，即做特定场景时有用的人机对话。

- **NLP+行业：与专业领域深度结合**

银行、电器、医药、司法、教育等领域对自然语言处理的需求都非常多。自然语言处理与各行各业的结合越来越紧密，专业化的服务趋势逐渐增强。刘挺教授预测，自然语言处理首先会在信息准备充分，并且服务方式本身就是知识和信息的领域产生突破，例如医疗、金融、教育和司法领域。

- **学习模式：先验语言知识与深度学习结合**

自然语言处理中学习模式有一个较为明显的变化。在浅层到深层的学习模式中，浅层学习是分步骤的，深度学习的方法贯穿在浅层分析的每个步骤中，由各个步骤连接而成。而直接的深度学习则是直接的端到端，人为贡献的知识在深度学习中所占的比重大幅度减小。但如何将深度学习应用于自然语言处理需要进行更多的研究和探索，针对不同任务的不同字词表示，将先验知识和深度学习相结合是未来的一个发展趋势。

- **文本情感分析：事实性文本到情感文本**

之前的研究主要是新闻领域的事实性文本，现在情感文本分析更受重视，并且在商业和政府舆情上可以得到很好地应用。如2017年新浪微舆情和哈工大推出“情绪地图”，网民可以登录新浪舆情官方网站

查询任何关键词的“情绪地图”，这是语义情绪分析在舆情分析产业的首次正式应用。