

gitbook.cn

GitChat - 《自然语言处理研究报告》的文章

11 ~ 14 分

从知识产业角度来看，自然语言处理软件占有重要的地位，专家系统、数据库、知识库，计算机辅助设计系统（CAD）、计算机辅助教学系统

（Cal）、计算机辅助决策系统、办公室自动化管理系统、智能机器人等，全都需要自然语言做人机界面。长远看来，具有篇章理解能力的自然语言理解系统可用于机器自动翻译、情报检索、自动标引及自动文摘等领域，有着广阔的应用前景。

随着自然语言处理研究的不断深入和发展，应用领域越来越广。

文本方面的应用主要有：基于自然语言理解的智能搜索引擎和智能检索、智能机器翻译、自动摘要与文本综合、文本分类与文件整理、智能自动作文系统、自动判卷系统、信息过滤与垃圾邮件处理、文学研究与古文研究、语法校对、文本数据挖掘与智能决策以及基于自然语言的计算机程序设计等。

语音方面的应用主要有：机器同声传译、智能远程教学与答疑、语音控制、智能客户服务、机器聊天与智能参谋、智能交通信息服务（ATIS）、智能解说与体育新闻实时解说、语音挖掘与多媒体挖掘、多媒体信息提取与文本转化以及对残疾人智能帮助系统等。

此外,建立在自然语言处理技术基础之上的心理学、认知学、哲学、混沌学说的共同发展,将使人们对智能的起源问题有新的认识。如果把计算机网络和未来的网络看作是由机器组成的机器社会,那么一种属于机器的智能可能会因为人类的参与以及机器社会中各元素的相互作用而自然诞生。这样,机器必将能够通过“图灵测试”,达到“会思考”的层次。而有关智能机器的研究也会诞生一系列新的领域,比如,机器心理学和机器认知学等。

其中,机器心理学主要研究机器的心理反应和意图。美国圣迭戈神经科学研究所研制的机器人 DarwinV II,能够根据其感知对外部事物进行分类,并根据经验和知识采取相应的对策。然而,机器心理学的研究不能局限于此,人们还需要对机器的意识、知觉、思想、情感、情绪、创造力、机器社会、机器交流等方面进行研究,而这一切还需要计算机科学、心理学、神经科学的同步发展。

我们选取一些自然语言处理应用较为频繁的场景进

行介绍。

• 知识图谱

知识图谱能够描述复杂的关联关系，它的应用极为广泛，最为人所知的就是被用在搜索引擎中丰富搜索结果，并为搜索结果提供结构化结果体现关联，这也是 google 提出知识图谱的初衷。同时微软小冰、苹果 siri 等聊天机器人中也加入了知识图谱的应用，IBM Watson 是问答系统中应用知识图谱较为典型的例子。按照应用方式，可以将知识图谱的应用分为语义搜索、知识问答、以及基于知识的大数据分析和决策等。

语义搜索利用建立大规模知识库对搜索关键词和文档内容进行语义标注，改善搜索结果，如谷歌、百度等在搜索结果中嵌入知识图谱。知识问答是基于知识库的问答，通过对提问句子的语义分析，在将其解析为结构化的询问，在已有的知识库中获取答案。在大数据的分析和决策方面，知识图谱起到了辅助作用，典型应用是美国 Netflix 公司利用其订阅用户的注册信息以及观看行为构建的知识图谱反映出英剧版《纸牌屋》很受欢迎，于是拍摄了美剧《纸牌屋》，大受追捧。



微软小冰
小黄鸡
公子小白



IBM Watson Health





• 机器翻译

机器翻译是自然语言处理最为人知的应用场景，一般是将机器翻译作为某个应用的组成部分，例如跨语言的搜索引流等。目前以 IBM、谷歌、微软为代表的国外科研机构和企业均相继成立机器翻译团队，专门从事智能翻译研究。如 IBM 于2009年9月推出 ViaVoice Translator 机器翻译软件，为自动化翻译奠定了基础；2011年开始，伴随着语音识别、机器翻译技术、DNN（深度神经网络）技术的快速发展和经济全球化的需求，口语自动翻译研究成为当今信息处理领域新的研究热点；Google 于2011年1月正式在其 Android 系统上推出了升级版的机器翻译服务；微软的 Skype 于2014年12月宣布推出实时机器翻译的预览版、支持英语和西班牙语的实时翻译，并宣布支持40多种语言的文本实时翻译功能。





Break through language barriers.

尤其值得之注意的是，在“一带一路”这一发展背景下，合作沟通会涉及60多个国家、53种语言，此时机器翻译的技术应用显得尤为重要，语言的畅通是“一带一路”战略得以实施的重要基础。而机器翻译涉及到语义分析、上下文环境等诸多挑战，其发展道路还有很长一段路要走。

• 聊天机器人

聊天机器人是指能通过聊天 app、聊天窗口或语音唤醒 app 进行交流的计算机程序，是被用来解决客户问题的智能数字化助手，其特点是成本低、高效且持续工作。例如 siri，小娜等对话机器人是一个应用场景。除此之外，聊天机器人在一些电商网站有着很实用的价值，可以充当客服角色，例如京东客服 jimi，有很多基本的问题，其实并不需要真的联系人工客服来解决。通过应用智能问答系统，可以排除掉大量的用户问题，比如商品的质量投诉、商品的基本信息查询等程式化问题，在这些特定的场景中，特别是会被问到高度可预测的问题中，利用聊天机器人可以节省大量的人工成本。

聊天机器人





• 文本分类

文本分类是指根据文档的内容或者属性，将大量的文档归到一个或多个类别的过程。这一技术的关键问题是如何构建一个分类函数或分类模型，并利用这一分类模型将未知文档映射到给定的类别空间。

按照其领域分类不同的期刊、新闻报道，甚至多文档分类也是可能的。文本分类的一个重要应用之处是垃圾电子邮件检测，除此之外，腾讯、新浪、搜狐之类的门户网站每天产生的信息纷繁复杂，依靠人工整理分类是一项耗时巨大的工作且很不现实，此时文本分类技术的应用就显得极为重要。

• 搜索引擎

自然语言处理技术例如词义消歧、句法分析、指代消解等技术在搜索引擎中常常被使用。搜索引擎的职责不单单是帮助用户找到答案，还能帮助用户找到所求，连接人与实体世界的服务。搜索引擎最基本的模式是自动化地聚合足够多的内容，对之进行

解析、处理和组织，响应用户的搜索请求找到对应结果返回。每一个环节，都需要用到自然语言处理。用百度举例，比如用户可以搜“天气”、“日历”、“机票”及“汇率”这样的模糊需求，会直接在搜索结果呈现结果。用户还可以搜索“范冰冰演过的电视剧”这样的复杂问题，百度都可以准确地回答。

一方面，有了自然语言处理技术才使得搜索引擎能够快速精准的返回用户的搜索结果，几乎所有的自然语言处理技术都在搜索引擎中有应用的影子；另一方面，搜索引擎（例如谷歌商业帝国和百度巨头）在商业上的成功，也促进了自然语言处理技术的进步。



• 推荐系统

第一个推荐系统是1992年 Goldberg 提出的 Tapestry，这是一个个性化邮件推荐系统，第一次提出了协同过滤的思想，利用用户的标注和行为信息对邮件进行重排序。推荐系统依赖数据、算法、人机交互等环节的相互配合，应用了数据挖掘技术、信息检索技术以及计算统计学等技术使用推荐系统的目的是联系用户和信息，帮助用户发现对自己有价值的信息，同时让信息能够展示在对它感兴趣的用户面前，精准推荐，用来解决信息过载和用户无明确需求的问题。

推荐系统在音乐电影的推荐、电子商务产品推荐、个性化阅读、社交网络好友推荐等场景发挥着重要的作用，美国 Netflix 2/3的电影是因为被推荐而观看，Google news 利用推荐系统提升了38%的点击率，Amazon 的销售中推荐占比高达35%。