

·专辑:人工智能与情报学·

## 智能情报分析系统的架构设计与关键技术研究\*

化柏林 李广建

(1.北京大学信息管理系 北京 100871)

**摘要:**人工智能发展的浪潮影响着每一个领域。情报作为与智能在字面上密切相关的领域,对人工智能的发展格外关注。如何把智能渗透到情报领域,实现智能情报分析系统是新时代值得深入探讨的问题。文章在对智能与情报关系的梳理以及对现有情报分析系统述评的基础上,提出一个智能情报分析系统的框架。智能情报分析系统在资源方面具有数据集与知识库、模型库与方法库,借助知识获取、知识表示与组织、自然语言处理、人机交互、新型计算与深度学习等,实现情报需求智能感知,海量信息智能获取,多源信息动态融合、多维关联综合分析、分析结果智能解读、情报报告自动生成、面向场景适时服务。通过智能分析引擎实现资源、技术与功能之间的对接,这是智能情报分析系统的核心。

**关键词:**智能情报分析系统;情报分析;人工智能

中图分类号:TP18;G250

文献标识码:A

DOI:10.11968/tsyqb.1003-6938.2017117

### Architecture Design and Key Techniques of Intelligence Analysis System Based on Artificial Intelligence

**Abstract** The wave of artificial intelligence development affects every area. Intelligence, as a field that is closely related to intelligence by literacy, pays special attention to the development of artificial intelligence. How to infiltrate the intelligence into the intelligence analysis field and realize the intelligence analysis system is a question deserved to be discussed in depth in the new era. On the basis of reviewing the existing intelligence analysis system, this paper proposes a framework of intelligence analysis system based on AI. The intelligence analysis system has the dataset, knowledge base, model base and method base. With intelligent knowledge acquisition, knowledge representation and organization, natural language processing, human-computer interaction, new computing and deep learning, Intelligent acquisition of massive information, dynamic fusion of multi-source information, comprehensive analysis of multidimensional correlations, intelligent interpretation of analysis results, automatic generation of intelligence reports and timely service oriented to scenes. Resources, technologies and functions could be joined through the intelligent analysis engine, which is the core of intelligence analysis system.

**Key words** intelligence analysis System;intelligence analysis ;artificial intelligence

革命性的智能技术,一边撕裂旧世界,一边创造对未来的无限渴望。大数据在理念与技术方法上已触及到各行各业,而人工智能正在洗刷着传统的思维与产业形态。数据、计算能力和算法叠加后产生的能力差异,正快速形成组织机构之间不可逆且不可逾越的智能鸿沟。过去的两年人类见证了 AlphaGo 系列的成功,深度学习在图像识别、智能语音、机器翻译、智能控制等领域取得了长足进步,越来越多的

领域开始涉及并运用深度学习等人工智能技术。未来,AI 将不断扩大渗透领域,技术应用也将逐步走向成熟,全球经济将迎来“AI-first”突破,可解释性和透明度将逐步得到改善。

在海量数据急剧增长、软件工具日益丰富、竞争环境日趋激烈的今天,科学研究、国家发展与企业运营越来越强调数据与智能在分析过程中的作用。业务问题呈现出复杂化、多维化、关联化、实时化的特

\* 本文系国家社会科学基金重点项目“大数据环境下的计算型情报分析方法与技术研究”(项目编号:14ATQ005)与国家社会科学基金面上项目“基于多源数据融合的情报用户需求探测研究”(项目编号:17BTQ066)研究成果之一。

收稿日期:2017-12-19;责任编辑:魏志鹏

点,企业组织过去所依赖的基于局部数据、传统工具而进行的业务决策方式必将跌落鸿沟。如何在情报系统中更多地引入智能技术,提高情报分析的水平与质量,成为新时代情报发展的关键<sup>[1]</sup>。为了更好地实现情报信息的支持,需要借助于大数据产业情报平台或智能情报分析系统,通过系统快速准确地获取所需要的基础数据、动态信息以及最新情报,跟踪科技前沿,梳理发展脉络,把握问题关键,分析竞争格局与发展态势,仿真模拟以及未来分析,发现新颖点与新机会,从而发现其中的市场机会和风险预警,占据市场竞争中的优势或主导地位。

## 1 智能与情报的关系

智能包括人类智能与人工智能。在现代情报分析中,既要有人类智能,如专家智慧、专业背景、师傅带徒弟与长期“工匠”的积累,又要有人工智能,利用神经网络、机器学习等智能算法对大规模的数据进行信息与知识的融合并使其转化为情报,加强情报分析的智能性,是智能情报的难点与重点。

所谓“智能”,通常是指人们在认识与改造客观世界的活动中,由思维过程和脑力劳动所体现的能力,包括感知能力、思维能力和行为能力<sup>[2]</sup>。人类智能的核心在于知识,智能表现为知识获取能力、知识处理能力和知识运用能力。人们对人工智能有许多不同的观点和实现方法,包括符号主义、连接主义和行为主义等,或者叫做逻辑学派、仿生学派和生理学派。所谓“情报”,在《辞海》中将其释义为:军中集种种报告,并预见之机兆,定敌情如何,而报于上官者。目前,国内外学者关于情报概念的探讨主要有以下三种观点:(1)情报的信息论,认为情报是一种特定的信息<sup>[3]</sup>,主要指可行动的信息;(2)情报的知识论,即情报是一种知识。英国情报学家布鲁克斯认为:情报是使人原有的知识结构发生变化的那一小部分知识<sup>[4]</sup>。钱学森<sup>[5]</sup>院士认为:情报是为了解决一个特定的问题所需要的、具有及时性和针对性的知识,是激活了、活化了的的知识;(3)情报的认知论,即情报是一种认知。池建文<sup>[6]</sup>研究员认为:情报是关于他方事实的本来面貌和深层次原因的判断,是新信息与旧知识融合所形成的新认知。

新的时代特点与技术环境给情报工作带来了很大的变化,数据对象与规模不断扩大,技术方法持续更新,但是不管对象与技术方法如何变化,情报工作的宗旨从未有过根本性改变,情报工作的目的就是要保障决策、解决决策过程中信息不完备的问题<sup>[7]</sup>。从任务的类型来看,情报分为情报识别、情报跟踪、情报比较、情报评价与情报预测<sup>[8]</sup>。从各类信息源中快速识别有用的信息,及时准确地捕捉最新、最快的情报动态,持续地对关键技术问题、重要国家与机构进行定期或长期跟踪,把握最新进展与动向,梳理技术路线与演化规律,分析竞争关系与态势,预测事情、格局、科学技术发展的趋势与走向,以便提前布局、未雨绸缪。

## 2 情报分析系统发展现状

情报分析系统是面向任务的,通过集成实现情报分析理论方法的技术完成特定的情报任务,技术实现方法种类和数量的选择与面向任务的研究对象、目的有关。情报分析系统发展至今,其智能水平早已不能满足应用需求。分析人员更倾向于系统能帮助发现未知的潜在模式,能从分析活动的各个环节上加以支持,提供一体化的解决方案,而非仅仅是零散的对若干分析功能的支持。这些需求也促进情报分析系统向着更智能的方向发展,主要包括对数据的深层发现、挖掘和对分析过程的智能支持等方面。情报分析系统呈现出计算性、融合性、工程化以及快速响应等特点。从行业实践来看,情报分析系统包括科技情报分析系统、企业竞争情报系统、公安情报分析系统等。

### 2.1 情报分析系统理论研究

李广建等<sup>[9]</sup>认为大数据时代下的情报研究应从单一领域情报研究转向全领域情报研究、综合利用多种数据源、注重新型信息资源的分析、强调情报研究的严谨性和情报研究的智能化五个方面;张志强<sup>[10]</sup>认为科技文献情报机构需要构建适应数据密集型科学范式需求的新型范式,着力开展专业型、计算型、战略型、政策型和方法型等“五型融合”的科技情报研究新范式;贺德方<sup>[11]</sup>通过分析大数据兴起、科研范式变革、方法工具演进、用户需求变化的时代特点,提出利用工程化思维开展科技情报研究的新范

式。乔晓东等<sup>[12]</sup>认为,大数据时代,情报处理和分析工作应以工程化和系统化的思路开展,提出了技术情报工程的理念和以技术基础设施为支撑的情报分析体系;苏新宁等<sup>[13]</sup>指出面向突发事件应急决策的快速响应的情报体系是一个以大数据环境为基、情报技术为力、情报流控制为策、应急决策为标的新型情报体系,并从组织结构、功能组成、组织功能关系三个方面构建了新型情报体系;李纲等<sup>[14-15]</sup>认为突发事件的监测与识别是应急决策情报体系的一个重要环节,一个以情报为核心的突发事件监测与识别的理论架构应该以情报收集、情报分析、情报评估与利用为过程,实现全源情报、实时情报、精准情报。情报分析呈现出工程体系化、海量数据高度依赖化、多源融合、集成计算、快速响应等特点,这些特点都是实现智能的基础与前提。

## 2.2 现有情报分析系统的主要类型

(1)科技情报分析系统。现代科技情报工作的态势解析与决策服务应借鉴大数据研究的最新进展,加快变革科技信息的采集、获取、挖掘、分析及影响方式,及时和变革性地提升科技知识的产生,使其更为直接地辅助科技决策。由中科院文献情报中心开发的“研究所一线科技信息监测服务平台”(http://stis.las.ac.cn)从海量的网络科技信息资源中发现最新最重要的科技资源,自动计算分析科技资源的情报价值、识别科技资源中包含的重要科技对象、重要科技术语,快速发现研究领域的研究热点等。北京市科技情报研究所开发的大数据科技情报服务平台,通过词库生成器、科技情报分析系统、大矩阵和创新点自动发现系统,在城市交通情报服务、突发事件应急决策、网络舆情监测管理、战略性新兴产业和文化创意产业等领域进行了应用实践。中国科学技术信息研究所与万方数据联合开发的科技创新辅助决策支持系统 STADS(后演化为科技创新小助手),可以根据用户的输入快速生成包含有各种复杂图表的 PDF 报告。

中科院自动化研究所王飞跃团队融合社会科学、情报科学、信息科学等学科理论提出基于 ACP 方法的开源情报解析理论框架,并以天网(科情信息监测)、天眼(科情信息分析与挖掘)、天鹰(科情信息

管理与决策支持)为指导流程,构建面向大数据和开源信息的科情监测与协作创新平台。在异构、多模态的科技情报资源之上,生成科情获取传感网络、科情语义网络、科研协作社会网络 3 个核心网络,实现智能采集、处理、分析与挖掘科技情报等功能<sup>[16]</sup>。由清华大学和北京搜狗公司合作开展的科技情报大数据挖掘及服务平台,以新一代科技情报分析与挖掘平台 AMiner 为基础,运用动态策略选择的语义集成方法和多维依赖关系的语义标注方法,建立了亿级节点规模的科技知识图谱。运用基于话题的影响力模型对异构科技情报网络进行深度挖掘,对科技信息网络中学者、学术活动和知识概念进行隐含语义建模,建立了超过 2.3 亿学术论文/专利和 1.36 亿学者的科技智库,提供面向科技文献、专利和科技新闻的语义搜索、语义分析、成果评价等知识服务<sup>[17]</sup>。

(2)面向企业的竞争情报系统。竞争情报系统(Competitive Intelligence System, CIS)是基于信息采集、全文检索、文本挖掘等核心技术,对企业自身、竞争对手和企业外部环境的情报信息进行收集、存储、处理、分析的应用管理系统,为企业提供战略决策支持,从而提高企业的核心竞争力。企业竞争情报系统是以人的智能为主导、以有形智能软件为技术载体,将实时资讯、经营数据监测、市场动态监测、竞争对手调研、行业分析研究等竞争情报内容服务,系统地集成成为一个企业竞争情报整体解决方案,以人机协作工作体系和竞争情报管理体系,实现对企业所处整体竞争环境的全面监测、分析与掌控,以增强企业竞争力为目标的人机结合的企业竞争战略决策支持与咨询系统。典型的竞争情报系统有 Knowledge Works、Text-Analyst,百度的 eCIS、北京拓尔思公司的 CIS 系统、谷尼企业竞争情报系统、才思竞争情报系统等。

(3)基于大数据的公安情报系统。公安情报分析系统以公安情报分析业务为指导,实时汇总各类社会信息、公安信息、互联网信息等,通过分析模型、人群碰撞与分析精灵等工具,通过关联分析、可视化分析等方法,使社会各类基本信息产生有价值的情报信息,并智能化的产生研判结果。大数据公安情报分析系统一般具有可视化智能轨迹分析查询、异常活动人员研判分析、智能统计分析、警情动态监测及分



析预警、智能情报检索、关联查询、碰撞比对,实现人、事、物、组织和地点等五要素的信息数据分析和研判,并在这些信息数据之间建立起内在关联,可帮助公安干警梳理各类分散的、独立的情报线索。典型的大数据公安情报系统有与美国 CIA、FBI 合作的大数据平台 Palantir、浪潮大数据警务云平台、江苏中科惠软公司开发的智慧公安网情报信息化作战平台等,北京拓尔思公司开发的大数据舆情分析平台 TRS SMAS 也有些类似功能。

### 2.3 相关述评

从上述概述可以看出,情报分析更多地借助于系统支撑,不断地融入大数据与智能技术已经成为一个趋势。情报分析系统智能化既有情报行业本身发展的内在需求,也具备社会发展应用的外在动力,计算机领域与大数据领域开始有一些团队构建智能情报分析系统成为一个新的特点,这个现象值得关注与深思。

情报分析系统的研究大部分是模型构建或系统设计,而实际构建并有良好运行结果的系统还比较少,也缺乏对系统运行效果的评估与评价。在模型构建或系统设计时多以系统论为指导。竞争情报系统的探讨比较充分,实际开发或应用的单位也更多一些,其次是以公安情报或反恐情报分析系统。

智能分析首先会在信息加工处理与分析阶段实现高度自动化与智能,如何在情报分析系统的输入端与输出端更多地引入自动化与智能化会是下一步发展的重点。智能情报分析系统应从情报需求特点

出发,以科学技术与市场应用情报服务为导向,以技术跟踪、情报分析与智能预警技术为支撑,设计一套具有较强针对性和实用性,结合行为过程和技术过程,融合网络搜索、数据挖掘、信息分析、人工智能等多学科领域知识和方法,能够实时、长期、准确地对科技前沿、政策法规、社会网络活动、最新动态以及发展趋势进行跟踪、监测、采集、统计、分析和预警评价的科技情报服务体系。

### 3 智能情报分析系统的功能结构

智能情报分析系统在资源方面具有数据集与知识库、模型库与方法库,在技术维度需要知识获取技术、知识表示与组织技术、自然语言处理技术、人机交互技术、新型计算机技术与深度学习技术,在资源的基础上通过这一系列技术实现以下功能:情报需求智能感知,海量信息智能获取,多源信息动态融合、多维关联综合分析、分析结果智能解读、情报报告自动生成、面向场景适时服务。在资源、技术与功能之间通过智能分析引擎实现连接与驱动,智能分析引擎包括认知模型、业务逻辑规则系统、算法调度引擎、验证与评估器四个构件,是整个系统的核心与动力(见图1)。

#### 3.1 问题情境快速建模

在大数据环境下,越来越重视用户场景,只有分析出用户的场景,并针对场景的变化实时更新服务的内容与方式,才能更好地理解并合理地满足用户需求,用户对产品与服务的接受程度、满意程度从而

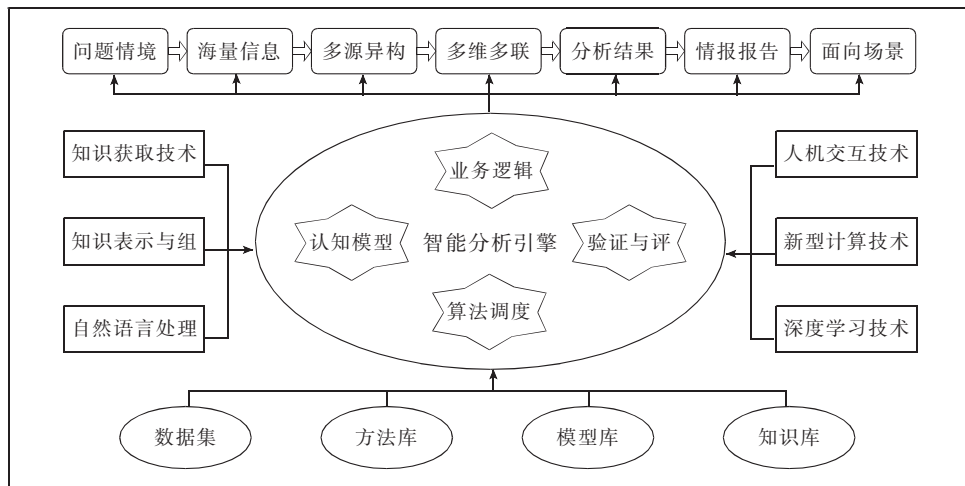


图1 智能情报分析系统体系架构图

会有大幅度地提高。情报问题与任务往往具有很强的动态性与个性化,这个特点导致很难生产出一套通用的情报分析系统。大数据环境下,可以通过多种渠道收集用户的数据,基于用户数据为用户(包括机构用户与个人用户)所处的情报问题与任务进行快速建模。针对不同的环境与时间节点,在用户基本属性的基础上,可迅速构建情报问题动态要素及指标,梳理并计算要素之间的相关关系与制约关系,确定要素之间的权重、时间先后关系等,形成情报问题任务与情境的模型。从注重分析过程向问题与情景方面转化,是智能情报分析系统发展的一个趋势。在情报系统中会有很多关于情景问题与应用场景的研究、相关数据的支撑、模型的建立、决策影响因素的制约关系计算。

### 3.2 智能采集信息

情报分析系统需要实时关注最新信息,需要广泛地从网络上采集各类信息。智能信息采集是情报分析系统的基础与关键,智能信息采集需要确定信息源与信息类型,自动探测信息分布,评估任务所需信息量,然后进行自动采集,对采集下来的信息进行过滤、筛选与甄别。自动采集虽然技术上比较成熟,但由于网络采集信息需要依赖于网页特点,网站存在改版更新、网页设计布局时常变化等特点,所以如何识别网页特征、自适应配置采集规则、动态调整模板与参数需要体现出很强的智能性。

### 3.3 多源异构动态融合

在智能情报分析系统中,采集网络信息只是一个方面,很多情报任务还需要从某些专业数据库中加载数据以及从自有的信息资源中识别出关联信息。要描述当前的竞争环境、技术前沿,传统的单一的数据是不完备的,需要综合论文、专利、政策、标准、产品、市场、网络挖掘、新闻舆情等不同来源、不同结构、不同内容揭示的数据进行集成融合,形成多源、异构、跨域的大数据,通过汇集这些多源的数据,实现对各类数据的即时交叉比对生成准确及时的各类信息,为问题诊断、态势评估与形势分析提供全面的数据支撑与情报参考。把多源异构的信息进行动态融合,用数据的多源性减少信息的不确定性、保证情报产品的客观性,从而提高决策分析的科学性。

### 3.4 多维关联与可视化智能分析

智能情报分析系统可以从多源异构大数据中发现一些清晰的线索,弄清行业或学科发展的来龙去脉,利用论文、专利等数据快速绘制技术路线图、产品生命周期图、学科发展脉络图。把不同的数据组合起来,进行关联映射与分析,既可以形成一些群组,也可以发现一些有用的关系与模式,确定关键人物、事件、联系和模式,计算数据之间的共性与关联关系,运用关联规则、聚类分析、社会网络分析、向量空间模型等一系列分析方法,对于整个行业或问题所涉及到的整体全貌进行呈现。智能情报分析系统可以实现资源之间的相互链接与揭示、概念之间的关系发现,支持各种主题分析、聚类分析。

可视化智能分析将各类数据以图形的方式展现出来,通过社会网络图、云图、雷达图、热力图等图形的方式对数据和数据间的关联进行描述和展现,运用众多图形分析的方法(如关联分析、网络分析、路径分析、时间序列分析、空间分析等)来发现和揭示数据中隐含的公共要素和关联。实现重要与关键信息的标注、时间演化与发展路线图,支持演化分析、技术发展路径分析;支持发展趋势预测分析,决策场景要素分析,对决策所涉及的影响要素与要素之间的关系进行直观的可视化展示。

### 3.5 分析结果智能解读

在结果解读方面,传统的信息分析是从空间上发现数据之间的关系,包括关联规则挖掘、聚类分析等,或者从时间上发现数据的变化以及变化的规律,包括时间序列分析、预测分析等。能够进行复杂的可视化展示,但是对结果如何解读,数据所呈现出的规律、数据所折射出现象以及数据背后的原因都需要专家结合数据分析经验以及行业知识进行解读判断。智能情报分析系统可以从某种程度上代替人做一些智能分析的工作。如数据年度增长图或产品生命周期图,传统的做法是专家根据数据特点划分为几个阶段,整个发展或增长过程符合哪一种增长曲线。智能情报分析可以利用机器学习的方式发现数据的规律,探寻相关因素分析数据现象及现象背后的原因,用自然语言的方式给出可供选择的方案。不断地累积行业规则、分析模式,形成一套流程化的情

报分析方案,辅助生成情报报告。

### 3.6 情报报告辅助生成

不管专家撰写的情报报告还是智能情报分析系统生成的情报报告,其报告的内容一定是体现了智能成分(专家智慧或人工智能),其报告的核心与关键内容是一般人员或情报用户不能从其他方式或渠道直接获取的,这样的情报报告才会让情报用户眼前一亮,才会更好地吸引情报用户。要想产出有独特视角、独特观点和高质量的研究报告,必须依赖专家的特色理论以及专家对本专业领域情报信息敏锐的洞察力和分析力,不同的机构、不同的研究人员对同一数据资源可能会得出不同的分析结论<sup>[18]</sup>,这种不同的结论对于多角度透视事物的本质,以及增加理性思考与判断都有很重要的作用。在用户建模、数据搜集与加载、多维分析与可视化展示、报告模板生成方面等有规律的地方可以快速生成,情报分析人员只需要适当的参与即可。

### 3.7 面向场景适时服务

不同的时段与不同的场景会有不同的需求,有时需要及时跟踪最新动态,有时需要对最新态势进行评估分析,有时需要情报人员梳理出技术发展路线,有时需要梳理出某一类人员的名单。因此作为情报研究人员,需要借助情报系统收集整合各类数据信息,对数据信息进行多分类、多特征的标签,便于根据不同任务从不同的视角数据信息进行切片、下钻、分类汇总。如找出某一问题的所有技术路线,按照时间绘出演化路线,展示各角色与角色之间的关系。这三种需求分别需要从主题内容、时间轴以及角色分布三个维度进行分析与汇总。情报问题是动态演化的,情报分析系统要针对时代特点、政策环境变化、领导活动动态适时调整情报描述的视角与内容,只有针对问题、数据基础随时变化、面向场景适时服务才能吸引情报分析人员持续地使用。

## 4 智能情报分析系统的核心组件

智能情报分析系统需要以大量的数据为基础,融合大数据与人工智能技术,基于动态知识图谱和具体的业务场景模型,支持复杂业务问题的自动识别、判断并做出前瞻或实时决策的智能化产品系统。

要实现情报系统的智能化,需要具备以下条件:海量的数据集与全领域知识库支撑,丰富的模型库与方法库、机器学习算法、高效的分析引擎与推理机。

### 4.1 数据集与知识库

智能情报分析系统需要将门类庞杂、种类繁多的海量数据进行整合,建立统一的大数据语义知识网搜索平台,全面而深入挖掘信息之间的关联关系,将分散的孤立的信息集中到一起,对不同来源的数据进行交叉验证,互补融合。对数据进行规范化以及内容的揭示、语义的计算。为情报分析系统及时方便地根据任务加载数据,需要建立完整的元数据目录、数据接口以及数据使用日志。

在智能情报分析系统中,不管是机器学习的方式分析数据之间关系获取知识,还是人工构建行业知识,都是为了让系统具有更好的知识支撑。情报分析的过程主要是基于数据的规律,对数据进行加工处理、分析挖掘,而这些规律就是知识,包括有效数据的判别与筛选知识、数据结构的相关知识、多源数据的融合知识、信息甄别知识、相关性判断知识、计量分析知识,以及自然科学和社会科学等专业领域知识。梳理情报任务所面向行业的知识体系,将各类数据库整合为统一的知识图谱,以国际统一的本体网络语言 OWL 表示知识,并利用知识图谱技术实现行业情报的推理与应用。

### 4.2 模型库与方法库

在智能情报分析系统中不可缺少模型。模型是对现实世界与复杂问题的抽象,从抽象层次上描述系统的静态特征、动态行为和约束条件,主要包括业务模型、数据模型、算法模型与系统模型等。基于多维度的数据挖掘、统计分析,进行算法模型的建立和调优。要创建足够透明的深度学习模型以解释它们的预测,特别是当这些模型的结果被用来影响或告知人类决策时。

方法库有着对方法的详细的描述,包括方法的定义、分类、输入与输出、处理流程、应用范围、主要功能、优缺点等。基于流程的情报方法体系包括情报收集方法、信息融合方法、数据清洗方法、信息分析方法、情报研究方法以及情报传递方法等。其中信息分析方法包括计量分析方法、模式分析方法、关联分



析方法、聚类分析方法、网络分析方法、演化分析方法、共现分析方法、异常分析方法等,从数据挖掘的角度又包括关联规则挖掘、分类与聚类、回归分析、连接分析、决策树、粗糙集、神经网络、遗传算法等。

### 4.3 机器学习算法

当前人工智能之所以如此火,与前几年对大数据的研究与重视分不开的,有了海量的数据基础,还需要一个好的学习机制与算法,能从大量的数据中进行学习,总结出规律与特征,通过不断地优化与迭代,逐步逼近目标,这就是机器学习。机器学习按照学习方法分为机械式学习、指导式学习、示例学习、类比学习与解释学习。根据学习能力分为有监督学习、无监督学习,以及介于两者之间的强化学习。其中,监督学习方法利用一组已知类别的样本调整分类器的参数不断优化并提高性能,主要包括卷积神经网络、Hopfield 网络、径向基函数网络、贝叶斯学习、决策树、线性分类;无监督学习根据未标记过的训练样本解决模式识别中的各种问题,包括对抗生成网络、前馈神经网络、关联规则学习、分层聚类、聚类分析、k-最近邻算法;半监督学习包括生成模型、低密度分离、基于图形的方法、联合训练等方法。

AlphaGo 结合了深度学习、强化学习和蒙特卡洛树搜索这三种算法,深度强化学习(DQN)是深度学习与强化学习的结合,就是用深度学习网络自动学习动态场景的特征,然后通过强化学习对应场景特征的决策动作序列。深度学习主要包括深度信念网络、深度卷积神经网络、深度递归神经网络、分层时间记忆、深度玻尔兹曼机等等。在深度学习的基础上又进行了扩展,包括增强学习、增量学习、迁移学习等<sup>[19]</sup>。这些好的学习算法能够起作用,主要依赖于丰富的数据集作为训练语料。图像识别依赖于已经标注好的丰富的图片库,AlphaGo 也是依赖于大量的训练,即使 Alpha zero 不和对方博弈,通过自己博弈,也是训练了 500 万盘棋,从中不断地学习特征与优化算法。

### 4.4 智能分析引擎

智能情报分析系统具有丰富的数据集与知识库,通过模型库与方法库进行分析,但调用何种模型与方法,执行效果如何等,需要有一个智能分析引擎

进行调动与驱动。智能分析引擎是情报分析系统的关键与核心,由认知模型、业务逻辑规则系统、算法调度引擎、验证与评估器组成。

认知模型包括用户画像与场景分析引擎,通过分析情报用户的类型、调取情报用户静态属性以及监测情报用户动态信息,整合用户需求兴趣偏好信息,从不同维度为用户赋予合适的标签,提取用户的共性群体特征,并揭示情报用户的个体特征与偏好,准确地刻画出“千人千面”,以便准确地提供个性化服务与精准推荐。情报需求由情报需求类型、情报需求主题以及情报需求情境三个维度构成。

通过认知模型确定用户需求与场景之后,利用业务逻辑系统调取相关资源。情报需求是动态监测还是领域分析?是技术分析还是政策分析?通过业务逻辑规则系统进行确定。根据业务领域搜索相关的知识,对多源异构知识进行融合,针对不同行业,按照行业特点,制定精细的行业推荐规则。搜索的任务包括根据任务描述搜索相应的数据信息,根据分析工作的需要搜索相应的支撑知识。搜索算法包括深度搜索、广度搜索、启发式搜索:A\* 搜索 state-space representation, pruning strategy, beam search, game tree search, alpha-beta 搜索。

算法调度引擎就是要建立数据、问题方法之间的关系。情报分析方法众多,如何选择方法有时会成为一个问题,而使用智能 agent,根据应用场景与问题的变化动态选择合适的研究方法,就是体现情报智能性的一个标志。验证与评估器主要负责对监控分析过程,对相应的执行阶段进行评估,根据评估结果给出下一步的动作,是继续分析还是回溯反馈等。

## 5 智能情报分析系统的关键技术

智能情报分析系统除了分布式云计算技术、大规模并行计算技术、高维数据关联分析技术、数据可视化展示技术、用户画像与个性化推荐技术等通用大数据技术外,要使情报系统更好地运转,还需要知识获取技术、知识表示与组织技术、自然语言处理技术、人机交互技术以及新型计算技术。

### 5.1 知识获取技术

知识库构建有四种方法:(1) 知识工程师方法。

由知识工程师编写领域知识规则录入知识库或者人工构建本体,是传统的知识获取方法;(2)资料转化方法。把主题词表、本体等转化为知识图谱,形成机器可读的知识。知识图谱本质上是一种语义网络,图中的结点代表实体(entity)或者概念(concept),边代表实体/概念之间的各种语义关系;(3)知识抽取方法。知识抽取是指从各种类型的数据和信息资源中获得各种知识的过程,从多种媒体资源(如文本、图像、视频、音频等)中抽取出知识,从数据集中发现重要模式的过程等<sup>[20]</sup>。通过知识抽取从多种数据源获得采用某种知识表示形式的、完整、正确、无歧义的知识元及其语义关系,进而作为后续知识融合的输入,这种知识抽取和组织完全依赖于对知识间的各种关系的认识、挖掘和组织;(4)知识发现方法。从大量的数据中发现数据之间的关联规则,并把知识表示成计算机可以理解、可以推理的本体,与深度学习等智能算法结合起来。知识抽取重在把人们已经显性表达出来的知识结构化、计算机化。知识发现重在发现对人们新颖的、事先未知的知识。

## 5.2 知识表示与组织技术

如何表示并组织好知识,让计算机可以很好地识别与利用知识是智能情报分析系统的关键。知识图谱就是当代最通用的语义知识表示形式化框架。知识图谱的节点就是语义学里面说的“符号根基(symbol grounding)”,即语言符号与真实或想象空间中的对象的对接,在计算机中体现为语言符号与数字化对象的对接。边则是语义学里面说的“角色指派(role assignment)”,在计算机中体现为每个数字化对象与其他数字化对象之间的语义关系标签。

语义结构表示框架中,现有的知识图谱可以描述实体、关系、属性(状态)及其值这三类要素,但是对于情感、程度变化、因果条件、逻辑模态等,现有的知识图谱结构并不能很好地表达,需要进行改造建立事理图谱才能适应这些语义要素的表示。知识图谱研究对象为名词性实体及其关系,事理图谱研究对象是谓词性事件及其关系。知识图谱的主要形式是实体属性和关系,事理图谱则是事理逻辑关系以及概率转移信息。事件间的演化关系多数是不确定的,而实体之间的关系基本是稳定的。

## 5.3 人机交互技术

运用自然语言问答、大数据可视化、知识图谱化、地图GIS化等手段,提供大量方便的人机交互接口,实现人与机器的完美融合以及人与人之间的协同工作。智能情报分析系统对情报分析人员提供全文位的精准检索与面向问题的智能问答,支持面向文档的主题检索、属性特征检索以及面向问题的语义检索,支持面向技术方案、产品市场、机构人员的精准检索。辅助情报分析人员快速准确地找到问题的关键,问题的支撑数据、基本面数据、核心情报数据。通过检索引擎,为情报人员提供从海量数据中寻找蛛丝马迹的服务。通过人机接口,提供语音输入,基本实现面向领域问题的人机对话等。通过可视化技术实时展示各种信息分析结果,包括知识图谱、社会网络关系、大数据分析结果等。通过集成研讨厅支持多面板同时展示,支持研讨厅大屏、办公电脑中屏以及移动端小屏多级展示,可以同时展示情报分析各要素及要素之间的关系,可以展示历史演进过程,可以对未来的场景以及场景变化进行仿真模拟。

## 5.4 自然语言处理技术

智能情报分析系统所处理的数据对象很多是文本格式的,如论文、专利、新闻、政策等。智能情报分析系统的有些支撑资源也是文本格式的,如知识库;情报分析系统的自然语言检索接口、人机对话接口也是以自然语言为媒介的,而这些问题的解决都需要自然语言处理技术的支撑。自然语言处理已经成为一种应用赋能技术,随着实体知识库的构建、知识抽取和自动写作在特定领域的实用化和对话机器人从对接语料到对接知识图谱的换代,通过新一代人工智能创新创业团队,全面渗透到人工智能应用的各个角落。自然语言处理从浅层到深层面临范式转换,还处在对接情感计算与常识计算的战略性要地的关键位置。

自然语言处理技术的应用场景甚广,大致可分为分析型、生成型和交互型三类。舆情监控系统是典型的分析型系统;报告撰写系统是典型的生成型系统;各类聊天机器人是典型的交互型系统。从计算的角度,文本的向量化是跨越统计和联结两大阵营的。基于统计的向量化方法是潜在语义索引(LSI)模型,



基于联结的向量化方法是词嵌入(Word embedding)模型,前者具有保距离特性,后者具有保运算特性而且可与任务派生的优化目标深度耦合。目前,嵌入技术已经不局限于词,也可以整结构、整句嵌入了。但是总体上,当前技术能够大规模处理的,仍然只是具有“浅层句法”或者“简单标记”的NLP任务。更复杂语言现象的理解、更复杂语义关系的抽取,仍然任重道远<sup>[21]</sup>。

### 5.5 新型计算技术

新型计算技术包括语义计算、属性计算、情景计算、差异计算以及信号计算。由主题计算转向属性为主的分面计算以及面向任务的场景计算为主。信息检索、知识主题图、热点分析、文本聚类、信息抽取、潜在语义索引、神经网络、LDA模型,虽然带有一定的语义计算,但本质上还是以关键词为核心要素的计算,是通过词语为特征构建相应的向量实现语义计算。目前的神经网络主要对特征起反应,尚不能对语义起反应,也就是说,即使机器做出了正确的选择,但可能并不理解业务本身。在计算机图像识别中,可以通过大量的训练样本及学习,感觉到数据的特征并接近目标,识别出新图片中的猫,但计算机并不能很好地理解究竟什么是猫,这一点与人的认知还是有一定区别的。在谷歌的机器翻译中,采用Seq2Seq与注意力机制,虽然大多数情况下翻译的也很好,但这种端到端的学习难以解决语义的问题。深度学习需要大量的训练数据,其机制与结果的解释性并不强,神经网络的隐含层就像黑箱一样,内容不为人所知。在计算的粒度上,主题计算更多地关注名词或动词等反映内容的实体词,而涉及到情感、趋势、倾向、个性化等分析往往需要更多地关注特征计算或属性计算,需要更多地关注形容词或副词等修饰词。如何面向问题与用户场景去加载数据、构建模型,这些情景计算技术也是智能情报分析系统的关键技术之一。

传统的计算更多地计算数据之间的共性与关联关系,如关联规则、聚类分析、社会网络分析、向量空间模型等都是计算数据对象之间的共性特点。这些

计算得到了很好地发展与应用,技术进步与方法研究已相对比较成熟。在信息量巨大的情况下,找出一些共性与关联并不总是很有效,由文档间的共性关联计算转向差异性对比计算将成为一个新的尝试。通过差异性对比,识别出新需求、新产品、新理论、新方法、新技术、新方案会助力现代科技情报工作<sup>[22]</sup>。在未来趋势分析以及场景分析时,有些弱的信号也能逐步变强,成为事情的主导力量或因素,这要求对信号分析也要有深入研究与持续跟踪。语义计算、属性计算、情景计算、差异计算以及信号计算这些新型计算技术也将成为智能情报分析系统能否切合需求、发挥作用的关键技术。

## 6 结语

随着数据累积量的不断加大与数据结构类型的复杂多样,新兴信息技术包括深度学习(如云计算技术、深度学习算法、各类数据挖掘算法以及自然语言技术的提高)以及用户对情报需求的要求不断提高,情报系统必然会朝着智能方向发展,这既是时代特点的展现,也是行业发展的必然要求。按照图灵所提出的标准,情报用户对收到的情报报告难以判断出报告的撰写是由人还是由机器完成的,这份报告就成功地欺骗了人,实现了真正的智能。

但是,必须清楚地认识到,在情报系统方面实现人工智能还有很长的路要走。智能取代部分工作是发展的趋势,但不会是全部。与棋谱游戏、图像识别等当前流行的人工智能领域相比,情报分析领域的目标更加复杂,在机器学习运用方面也缺乏足够的情报案例训练数据,还有情报分析需要语义空间而不是特征空间,由于这些困难的存在,使得人工智能在情报领域的广泛应用充满了挑战。

情报行业必须认清情报分析系统智能化的趋势,需要勇于接受这个挑战,把握住历史发展的机遇,及时跟踪并运用大数据与智能技术,面向社会发展与应用需求,认真研究情报行业特点与规律,运用新的技术去解决行业共性与关键问题,扎实推进情报分析水平与服务。

## 参考文献:

- [1] 徐宏宇. 新智能时代颠覆情报的未来——访中科院自动化研究所复杂系统管理与控制国家重点实验室主任王飞跃[J]. 竞争情报, 2017, 13(4):4-7.
- [2] 林崇德, 杨治良, 黄希庭. 心理学大辞典. 上海: 上海出版社, 2003:1704.
- [3] 王崇德. 关于情报学[J]. 情报理论与实践, 1996(5):1-2.
- [4] Brookes B C. The foundations of information science: Part I. Philosophical aspects[J]. Journal of Information Science and Engineering, 1980(2):125-133.
- [5] 钱学森. 科技情报工作的科学技术[J]. 情报理论与实践, 1983, 6(6):3-10.
- [6] 池建文. 论情报的两个基本问题[J]. 情报学报, 2006, 25(S1):290-293.
- [7] 王忠军, 于伟, 杨晴. 科技情报机构实践创新发展专家访谈[J]. 情报理论与实践, 2017, 40(12):145.
- [8] 王延飞, 赵柯然, 陈美华. 情报研究中的治学思考[J]. 图书情报工作, 2017, 61(16):55-59.
- [9] 李广建, 杨林. 大数据视角下的情报研究与情报研究技术[J]. 图书与情报, 2012(6):1-8.
- [10] 张志强. 论科技情报研究新范式[J]. 情报学报, 2012, 31(8):788-797.
- [11] 贺德方. 工程化思维下的科技情报研究范式——情报工程学探析[J]. 情报学报, 2014, 33(12):1-13.
- [12] 乔晓东, 朱礼军, 李颖, 等. 大数据时代的技术情报工程[J]. 情报学报, 2014, 33(12):38-53.
- [13] 苏新宁, 朱晓峰. 面向突发事件应急决策的快速响应情报体系构建[J]. 情报学报, 2014, 33(12):53-77.
- [14] 李纲, 叶光辉. 网络视角下的应急情报体系“智慧”建设主题探讨[J]. 情报理论与实践, 2014, 37(8):51-55.
- [15] 李纲, 李阳. 情报视角下的突发事件监测与识别研究[J]. 图书情报工作, 2014, 58(24):66-72.
- [16] 王飞跃. 知识产生方式和科技决策支撑的重大变革——面向大数据和开源信息的科技态势解析与决策服务[J]. 中国科学院院刊, 2012, 27(5):527-537.
- [17] 北京市科学技术委员会[EB/OL]. [2018-01-13]. <http://www.bjkw.gov.cn/wzwj/2017chushen/src/I01-2017-030.html>.
- [18] 化柏林. 论情报的本质[J]. 情报理论与实践, 2012, 35(7):1-5.
- [19] 增强学习、增量学习、迁移学习——概念性认知[EB/OL]. [2017-12-23]. <http://blog.csdn.net/zyazky/article/details/51942135>.
- [20] 张智雄, 吴振新, 刘建华, 等. 当前知识抽取的主要技术方法解析[J]. 现代图书情报技术, 2008(8):2-11.
- [21] 白硕. 人工智能的诗与远方, 一文读懂 NLP 起源、流派和技术[EB/OL]. [2018-01-11]. [http://www.sohu.com/a/215995721\\_473283](http://www.sohu.com/a/215995721_473283).
- [22] 化柏林. 科技信息大数据在情报研究服务中的应用[J]. 图书情报工作, 2017, 61(16):150-156.

作者简介: 化柏林, 男, 北京大学信息管理系助理教授, 博士; 李广建, 男, 北京大学信息管理系教授, 博士生导师。