

人工智能与数据的学术研究分析报告

Artificial Intelligence and Data Research Report

方建勇¹（余姚 浙江 315400）

[摘要] 无论在产业界，还是在大众媒体报道中，人工智能大热。对于学界来说，我们还是要保持一个平常心，毕竟这不是一个学界可以炒作的概念，尽管大家对人工智能的未来充满期待。人工智能有两个重要的基础因数来支撑，一是大到足够充分的数据，尽管数据是需要不断积累的，包括自然数据，也包括经验数据；二是算法，在不同应用领域，加上数据样本空间的量级差异，不同算法表现出来的各种效能也是有差异的。本文主要针对人工智能与数据相结合的领域，分析了学术界历年来的总体研究状况，以期在广度与深度上对该领域的研究分布作一个概要性的陈述。

[关键词] 人工智能 数据 算法 学术研究 分析报告

[Abstract] Whether in the industry, or in the mass media reports, artificial intelligence hot. For the academic community, we still have to maintain a sense of balance, after all, this is not a society can speculation concept, although everyone on the future of artificial intelligence full of expectations. Artificial intelligence has two important basic factors to support, one is large enough to full data, although the data is the need to continue to accumulate, including natural data, including empirical data; the second is the algorithm, in different applications, plus data The magnitude of the sample space is different, and the performance of the different algorithms is different. This paper mainly focuses on the field of artificial intelligence and data, and analyzes the overall research status of academic circles over the years, with a view to making a summary statement on the distribution of research in this field in breadth and depth.

[Key words] Artificial intelligence; Data; Algorithm; Academic research; Analysis report

1 引言

¹ 方建勇 男，1978 年生，1998 年考入浙江大学数学系，中国工业与应用数学学会会员，中国物流学会会员，资深 IT 项目经理，资深 IT 工程师，资深 DBA（大型数据库管理员），浙江大学历史系硕士研究生学历，浙江大学数学与应用数学专业本科毕业，理学学士学位。

无论在产业界，还是在大众媒体报道中，人工智能大热。对于学界来说，我们还是要保持一个平常心，毕竟这不是一个学界可以炒作的概念，尽管大家对人工智能的未来充满期待。人工智能有两个重要的基础因数来支撑，一是大到足够充分的数据，尽管数据是需要不断积累的，包括自然数据，也包括经验数据；二是算法，在不同应用领域，加上数据样本空间的量级差异，不同算法表现出来的各种效能也是有差异的。本文主要针对人工智能与数据相结合的领域，分析了学术界历年来的总体研究状况，以期在广度与深度上对该领域的研究分布作一个概要性的陈述。

2 概述

人工智能（Artificial Intelligence），英文缩写为 AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能是计算机科学的一个分支，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能从诞生以来，理论和技术日益成熟，应用领域也不断扩大，可以设想，未来人工智能带来的科技产品，将会是人类智慧的“容器”。人工智能可以对人的意识、思维的信息过程的模拟。人工智能不是人的智能，但能像人那样思考、也可能超过人的智能。人工智能是一门极富挑战性的科学，从事这项工作的人必须懂得计算机知识，心理学和哲学。人工智能是包括十分广泛的科学，它由不同的领域组成，如机器学习，计算机视觉等等，总的说来，人工智能研究的一个主要目标是使机器能够胜任一些通常需要人类智能才能完成的复杂工作。但不同的时代、不同的人对这种“复杂工作”的理解是不同的。²

人工智能是科学技术发展的最新成果，其发展依靠相关的云技术和大数据技术的日渐成熟。³

3 各类型学术发展趋势分析

超星发现系统收录的人工智能与数据历年发表的中文学术成果，总量为 18,641 篇，其中包括图书 198 册、期刊 7235 篇、报纸 262 篇、学位论文 7098 篇、会议论文 983 篇、标准 2 个、专利 576 项、音视频 31 个、科技成果 1389 项、年鉴 37 篇、法律法规 7 篇、案例 3 篇、信息资讯 797 篇、特色库 23 篇。按发表的时间段来看，2017 年 1761 篇⁴、2016 年 2074 篇、2015 年 1198 篇、2014 年 1011 篇、2013 年 1020 篇、2012 年 919 篇、2011 年 1036 篇、2010 年 913 篇、2000-2009 年 6556 篇、1990-1999 年 939 篇、1989 年以前 1214 篇。

² 参见百度百科人工智能词条。

³ 马臻，车雨霏（武汉大学信息管理学院）：《大数据技术在人工智能中的应用研究》[J]，消费导刊 2016 第 5 期。

⁴ 为方便起见，这里单位统一为篇。

自 1979 年第一篇期刊论文收录以来，1984 年收录了第一篇会议论文，表明学界开始在会议上对人工智能与数据这个研究课题进行讨论，1985 年收录了第一篇学位论文，表明这个课题已经进入到高校研究群体当中，1989 年诞生了人工智能与数据研究领域的第一个科技成果和 2 项专利。1999 年是个分水岭，在这一年里，有两项人工智能与数据领域的标准诞生，期刊论文数量首次年度突破 100 篇，达到了 106 篇。2002 年，人工智能与数据领域的学位论文首次年度突破 100 篇，达到了 117 篇，随后出现爆发性增长，2006 年至 2015 年学位论文每年都超过 500 篇，其中在 2011 年达到年度峰值的 626 篇，2016 年有所下降，达到 361 篇，2017 年截至发稿时为 270 篇，表明学界对人工智能与数据领域的研究热度自 2016 年以来下降很多。学术研究热度晴雨表的一个领先指标，是会议论文的数量，2007 年、2008 年和 2009 年发表的会议论文数量分别为 91 篇、97 篇和 92 篇，从整体上来看，这是会议论文最为高产的三年，随后年发表数量陆续锐减，2010 年 35 篇、2011 年 78 篇、2012 年 48 篇、2013 年 55 篇、2014 年 28 篇、2015 年 15 篇、2016 年 18 篇，2017 年截至发稿时为 23 篇。与会议论文和学位论文发表数量发展趋势形成鲜明对照的是，期刊论文、专利和报纸在 2016 年、2017 年出现井喷态势，这些短、频、快的学术成果应为政府和媒体的推动有重要关联，2016 年期刊论文数量翻番，突破 1000 篇，达到 1172 篇，2017 年截至发稿时为 1523 篇；2016 年专利数量同样翻番，突破 100 项，达到 181 项，2017 年截至发稿时为 235 项；2016 年报纸发表量翻两番，达到 107 篇，2017 年截至发稿时为 121 篇。

表 1 人工智能与数据各类型学术发展趋势

“人工智能”“数据”-各类型学术发展趋势曲线									
序号	年份	图书 (数量)	期刊 (数量)	学位论文 (数量)	会议论文(数量)	专利 (数量)	标准 (数量)	报纸 (数量)	科技成果(数量)
1	1979	0	1	0	0	0	0	0	0
2	1980	0	2	0	0	0	0	0	0
3	1981	0	5	0	0	0	0	0	0
4	1982	0	0	0	0	0	0	0	0
5	1983	0	4	0	0	0	0	0	0
6	1984	0	11	0	3	0	0	0	0
7	1985	0	16	1	1	0	0	0	0
8	1986	0	30	0	8	0	0	0	0
9	1987	0	28	0	4	0	0	0	0

10	1988	0	30	1	6	0	0	0	0
11	1989	0	28	1	7	2	0	0	1
12	1990	0	31	0	10	0	0	0	0
13	1991	0	49	4	17	0	0	0	0
14	1992	1	44	3	9	0	0	0	0
15	1993	0	35	2	13	2	0	0	4
16	1994	0	64	2	18	1	0	0	0
17	1995	1	60	7	11	1	0	0	3
18	1996	1	80	24	16	0	0	0	3
19	1997	0	57	13	12	0	0	0	7
20	1998	1	91	28	24	6	0	0	5
21	1999	1	106	35	8	4	2	0	15
22	2000	5	111	30	14	3	0	0	24
23	2001	4	167	62	15	1	0	0	57
24	2002	3	156	117	26	2	0	0	58
25	2003	4	173	180	53	12	0	0	34
26	2004	4	210	252	25	8	0	0	42
27	2005	3	219	359	50	7	0	1	50
28	2006	9	224	544	50	11	0	1	43
29	2007	12	260	573	91	9	0	0	48
30	2008	10	310	529	97	20	0	0	53
31	2009	11	319	598	92	17	0	0	29
32	2010	23	280	501	35	16	0	0	41
33	2011	14	249	626	78	17	0	0	42
34	2012	9	253	544	48	33	0	0	16
35	2013	9	324	558	55	45	0	0	8
36	2014	15	331	502	28	55	0	9	10
37	2015	14	494	506	15	78	0	23	5
38	2016	17	1172	361	18	181	0	107	0
39	2017	22	1523	270	23	235	0	121	0

4 关键词分析

表 2 人工智能与数据关键词频次分析表

频次区间	关键词（频次）
3000 以上	人工智能(3163)

1000-2999	数据挖掘(1652)
500-999	专家系统(727)、神经网络(722)
200-499	故障诊断(462)、支持向量机(397)、遗传算法(374)、机器学习(363)、粗糙集(352)、关联规则(314)、数据库(295)、人工神经网络(280)、模式识别(257)、数据仓库(254)、聚类(241)、Bp神经网络(229)、人工智能技术(206)
100-199	物联网(198)、数据挖掘技术(177)、知识发现(176)、特征提取(172)、决策支持系统(172)、决策树(161)、知识库(150)、属性约简(150)、人脸识别(148)、继电保护(144)、系统设计(138)、信息技术(132)、聚类分析(129)、贝叶斯网络(129)、分类(128)、信息系统(125)、电力系统(124)、本体(120)、图像处理(119)、预测(116)、虚拟现实(116)、特征选择(116)、数据处理(114)、决策支持(107)、系统研究(106)、智能化(105)、云计算(104)、计算机应用(101)
87-99	数据采集(92)、互联网(90)、入侵检测(87)
87 以下	未统计

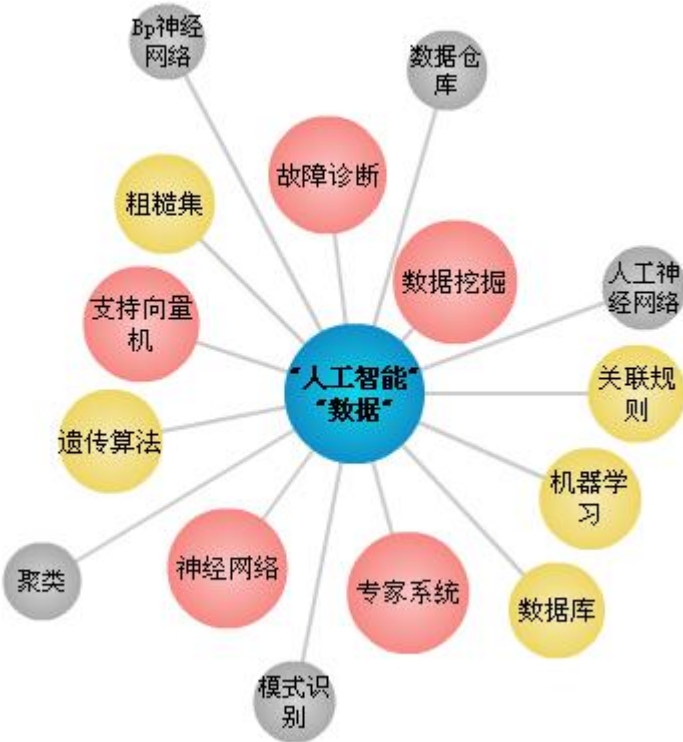


图 1 人工智能与数据关键词频次泡型图

5 作者机构分析

表 3 作者机构频次分析表

频次区间	作者机构（频次）
400 以上	浙江大学(423)
300-399	华中科技大学(372)
200-299	吉林大学(295)、中国科学院(257)、电子科技大学(228)、重庆大学(227)、哈尔滨工业大学(220)、北京航空航天大学(214)、武汉大学(203)
150-199	上海交通大学(194)、华南理工大学(190)、西安交通大学(190)、东北大学(173)、西安电子科技大学(166)、合肥工业大学(153)
100-149	华北电力大学(145)、大连理工大学(143)、西南交通大学(138)、东南大学(135)、山东大学(135)、天津大学(132)、中国科学技术大学(130)、中山大学(130)、北京科技大学(128)、北京工业大学(125)、中南大学(122)、南京理工大学(114)、清华大学(113)、北京交通大学(113)、武汉理工大学(109)、南京大学(103)、安徽大学(100)
70-99	北京理工大学(97)、哈尔滨工程大学(97)、南京航空航天大学(97)、厦门大学(96)、北京邮电大学(92)、湖南大学(90)、复旦大学(88)、广东工业大学(87)、同济大学(85)、四川大学(82)、江南大学(79)、北京大学(76)、上海大学(76)、西北工业大学(74)、华中理工大学(73)、中国地质大学（武汉）(71)、燕山大学(70)
70 以下	未统计

6 作者分析

表 4 作者发表篇数分析表

篇数区间	发表作者（篇数）
50 以上	张天序(56)
30-49	董金祥(38)、李德华(37)、田金文(37)、潘云鹤(33)
20-29	柳健(29)、郑南宁(28)、王磊(25)、孔繁胜(25)、彭嘉雄(23)、吴飞(21)、胡汉平(21)、何志均(21)、田雨农(21)、刘伟(20)、陈奇(20)

13-19	张燕平(19)、庄越挺(19)、桑农(19)、丁明跃(19)、王伟(18)、张伟(18)、汪国有(18)、史忠植(17)、俞瑞钊(17)、王刚(16)、杨静(16)、张铃(16)、王军(15)、张勇(15)、张莉(15)、刘静(15)、刘波(15)、李强(14)、王鹏(14)、吴涛(14)、黄厚宽(14)、熊范纶(14)、李伟(13)、王勇(13)、李军(13)、张敏(13)、李斌(13)、刘勇(13)、张涛(13)、刘洋(13)、张超(13)、郭涛(13)、王士同(13)
13 以下	未统计

7 学科分类统计

人工智能与数据-中文学科分类统计		
序号	名称	数量
1	工业技术	9506
2	经济	1314
3	文化、科学、教育、体育	568
4	交通运输	449
5	医药、卫生	414
6	天文学、地球科学	385
7	数理科学和化学	302
8	农业科学	188
9	航空、航天	113
10	环境科学、安全科学	86
11	社会科学总论	77
12	政治、法律	76
13	生物科学	76
14	语言、文字	61
15	军事	56
16	哲学、宗教	51
17	自然科学总论	43
18	艺术	23
19	文学	6
20	历史、地理	5
21	综合性文献	1

8 刊种统计

人工智能与数据-刊种统计		
序号	名称	数量

1	模式识别与人工智能	682
2	计算机科学	112
3	计算机工程与应用	78
4	计算机研究与发展	64
5	吉林大学学报（工学版）	59
6	华中科技大学学报（自然科学版）	56
7	数据采集与处理	55
8	计算机工程	49
9	计算机应用	47
10	中国计算机报	44
11	计算机应用研究	43
12	计算机学报	39
13	智能系统学报	35
14	电脑知识与技术	35
15	计算机工程与设计	35
16	计算机与数字工程	35
17	金融电子化	35
18	计算机技术与发展	34
19	小型微型计算机系统	33
20	通信产业报	30
21	IT 经理世界	28
22	电脑知识与技术(学术交流)	28
23	软件学报	26
24	中国图象图形学报	26
25	中国金融电脑	26
26	现代电子技术	26
27	计算机辅助设计与图形学学报	26
28	通信世界	26
29	黑龙江科技信息	25
30	中国信息化	25
31	中国传媒科技	25
32	红外与激光工程	25
33	微电子学与计算机	23
34	计算机光盘软件与应用	22
35	青年记者	21
36	电子世界	21
37	福建电脑	21
38	中国经济信息	21
39	科学导报	20
40	计算机应用与软件	20
41	计算机测量与控制	20

42	微计算机信息	20
43	信息化建设	19
44	科技资讯	19
45	信息与电脑(理论版)	19

9 发表机构地区统计

人工智能与数据-地区统计		
序号	名称	数量
1	江苏省	962
2	湖北省	689
3	上海市	900
4	陕西省	684
5	四川省	634
6	浙江省	621
7	辽宁省	586
8	广东省	580
9	安徽省	461
10	吉林省	451
11	黑龙江省	440
12	山东省	419
13	重庆市	355
14	湖南省	323
15	天津市	296
16	河北省	235
17	福建省	213
18	山西省	188
19	河南省	177
20	江西省	167
21	云南省	134
22	广西壮族自治区	121
23	甘肃省	107
24	贵州省	53
25	内蒙古自治区	48
26	新疆维吾尔自治区	37
27	宁夏回族自治区	12
28	青海省	6
29	海南省	4
30	西藏自治区	1

10 核心期刊统计

人工智能与数据-核心期刊统计		
序号	名称	数量
1	中文核心期刊(北大)	2472
2	CSCD 中国科学引文库(中科院)	2243
3	统计源期刊(中信所)	2214
4	EI 工程索引(美)	766
5	CSSCI 中文社科引文索引(南大)	150
6	SCI 科学引文索引(美)	6

11 扶持基金统计

人工智能与数据-扶持基金		
序号	名称	数量
1	国家自然科学基金项目	1047
2	省市基金项目	575
3	科技部国家科技计划项目	244
4	国家教育部基金	172
5	国防部国防科技基金项目	32
6	国家社会科学基金项目	28
7	其他基金项目	13
8	中国科学院基金项目	1

12 结论

统计发现, 关键词前 5 名依次人工智能(3163)、数据挖掘(1652)、专家系统(727)、神经网络(722)、故障诊断(462); 作者机构前 5 名依次是浙江大学(423)、华中科技大学(372)、吉林大学(295)、中国科学院(257)、电子科技大学(228); 作者前 5 名依次是张天序(56)、董金祥(38)、李德华(37)、田金文(37)、潘云鹤(33); 学科前 5 名依次是工业技术(9506)、经济(1314)、文化科学教育体育(568)、交通运输(449)、医药卫生(414); 刊种前 5 名依次是模式识别与人工智能(682)、计算机科学(112)、计算机工程与应用(78)、计算机研究与发展(64)、吉林大学学报(工学版)(59); 发表机构所在地区前 5 名依次是江苏省(962)、湖北省(689)、上海市(900)、陕西省(684)、四川省(634); 核心期刊类前 5 名依次是中文核心期刊(北大)(2472)、CSCD 中国科学引文库(中科院)(2243)、统计源期刊(中信所)(2214)、EI 工程索引(美)(766)、CSSCI 中文社科引文索引(南大)(150); 扶持基金前 5 名依次是国家自然科学基金项目(1047)、省市基金项目(575)、科技部国家科技计划项目(244)、国家教育部基金(172)、国防部国防科技基金项目(32)。

这份人工智能与数据学术研究的分析报告，包含了目前人工智能与数据中文类学术研究几乎可查到的所有元数据的大集合，代表了人工智能与数据中文类学术研究情报的最前沿，希望能对产业界或者相关学者有一定借鉴意义。

参考文献：

- 1 超星发现系统[EB/OL].<http://www.chaoxing.com/>
- 2 百度百科[EB/OL].<https://baike.baidu.com/item/人工智能/9180?fr=aladdin>