

S U M M E R 2 0 1 7



深圳市大数据研究院



Shenzhen Research Institute of Big Data

NEWSLETTER 第 2 期

二〇一七年六月

目录

- 研究院新成员 3
- 访问学者 4
- 新闻动态 7
- 讲座报告 11
- 研究院项目 13
- 人工智能 15



研究院新成员

汪德亮教授分别于1983年和1986年获北京大学学士和硕士学位，并于1991年获得美国南加州大学博士学位，全部来自计算机专业。自1991年，他任职于俄亥俄州立大学，现为该校计算机科学与工程系及认知科学中心教授。



汪德亮教授

汪教授主要致力于机器感知和信号处理领域的研究，在听觉处理的神经计算研究方面取得了重大成果。他开创性地提出了计算听觉场景分析理论与算法，为解决听觉领域中著名的“鸡尾酒会问题”提供了全新的研究思路 and 方向。汪教授在Neural Computation、IEEE Transaction on Audio、Speech and Language Processing等国际顶级期刊发表学术论文100余篇，在顶级国际学术会议发表学术论文近140篇。

汪教授于1992年获得NSF研究启动奖，1996年获得ONR青年科学家奖，1996年，2000年，2005年，2010年获得OSU工程学院Lumley研究奖。2005年他的论文“The time dimension for scene analysis”获IEEE Transaction on Neural Network优秀论文奖。他还获得了国际神经网络协会颁发2008届亥姆霍兹奖，并于2014被评为大学杰出学者。他是IEEE杰出讲师（2010-2012），是IEEE Fellow。另外汪教授还担任多个国际顶级学术期刊的主编、编委会委员、国际学术会议的大会及程序委员会主席。汪教授现被聘为深圳市大数据研究院的高级研究科学家。

陆彦辉教授分别于1994和1997年获得了西安交通大学的学士和硕士学位，2006年获得北京邮电大学博士学位。博士毕业一直任教于郑州大学，并于2013年获得教授职称。

2010年4月至2011年5月期间，陆教授作为访问学者在美国加州大学戴维斯分校研究学习；2014年3月至8月，她在澳大亚伍伦贡大学从事研究。陆教授的研究兴趣是优化理论与技术在线通信领域的研究和应用。作为项目主持人或主要参与者，她已经完成国家自然科学基金项目4项，其他项目8项。目前已发表论文60多篇，出版译著1部，申请发明专利6项。陆教授是2011年河南高等学校优秀青年骨干教师，2012年教育部新世纪优秀人才称号获得者。陆教授现被聘为深圳市大数据研究院的访问研究员。



陆彦辉教授

访问学者



Jong-Shi Pang 教授

Jong-Shi Pang教授分别于1975年和1976年获美国斯坦福大学硕士和博士学位。他于2013年8月加入美国南加州大学成为工业与系统工程Epstein Family教授。Pang教授在2003年获得George B. Dantzig奖项，该奖项由美国数学规划学会与美国工业与应用数学学会共同颁发，以表彰Pang教授在有限维变分不等式方面的突出工作。1994年，Pang教授获得由运筹学与管理学研究协会颁发的Frederick W. Lanchester奖。Pang教授的多篇论文在各个工程学科领域都获得了最佳论文奖，包括信号处理、能源与自然资源、计算管理科学、机器人与自动化等。

1980年至1999年期间，Pang教授是美国信息科学研究所论文引用最多的教授，他发表了3篇专题论文被广泛引用，并且有100多篇学术期刊发表于顶尖期刊上。Pang教授于2009年成为美国工业与应用数学学会的会士。他的主要研究领域包括数学建模、基于运筹学的复杂工程与经济系统分析、最优化、平衡规划以及约束动力系统。



沈超 教授

沈超教授于2003年和2012年分别获得北京交通大学通信工程、信号与信息处理的学士和博士学位。2013年进入轨道交通控制与安全国家重点实验室（北京交通大学）进行博士后研究，2014年2月至2015年2月在美国马里兰大学帕克分校Sennur Ulukus教授团队进行访问学者研究。2015年4月至今，沈超博士任轨道交通控制与安全国家重点实验室（北京交通大学）副教授。2017年3月至今在香港中文大学（深圳）张纵辉教授的团队进行学术访问。

沈超教授长期致力于无线通信物理层的系统优化设计，在IEEE Transactions on Signal Processing、IEEE Transactions on Wireless Communications、IEEE Globecom/ICC等期刊、会议发表论文多篇。目前，沈超教授的研究集中于无线通信系统的优化设计，主要涉及超可靠低时延通信、基于能量采集的无线通信、面向高速铁路的无线通信、基于大数据的无线通信接入网设计等研究方向。

洪明毅教授于2005年在浙江大学取得通信工程工学学士，2007在纽约州立大学石溪分校获得电气工程硕士并于2011年在弗吉尼亚大学取得系统工程博士学位。从2011年到2014年，他在明尼苏达大学电子和计算机工程系进行博士后工作，很快晋升为研究员，之后升为研究助理教授。他现担任爱荷华州立大学的Black&Veatch教员，担任工业和制造系统工程系和电子与计算机工程系（兼）的助理教授。他主要的研究领域在大规模优化理论、统计信号处理、新一代无线网络以及在大数据领域应用的相关问题。



洪明毅教授



孙若愚教授

孙若愚教授于2009年在北京大学数学系获学士学位，于2015年在明尼苏达大学电子系获博士学位。从2015年到2016年，他在斯坦福大学进行博士后工作。2016年他任职于FAIR（Facebook人工智能研究院）做访问研究员。2017年孙教授加入伊利诺伊大学香槟分校，现为工业与企业系统工程系助理教授。他的研究兴趣主要在于大规模优化及其在机器学习、数据分析和信号处理中的应用，以及无线网络的信号处理和信息理论。

访问学生



张荣升
 复旦大学
 硕士研究生
 通信信息与系统
 研究方向：机器学习
 与深度学习



雷沧粟
 中国科学技术大学
 本科在读
 信息安全
 研究方向：自然语言
 处理



梁豪

华南理工大学
本科在读
信息工程
研究方向：机器学习



张煜

北京交通大学
硕士研究生
通信与信息系统
研究方向：无线通信



范铎

北京交通大学
硕士研究生
电子与通信工程
研究方向：超密度网络



徐彦卿

北京交通大学
博士研究生
通信与信息系统
研究方向：面向 5G 的高可靠低时延通信系统



谭靖虹

新加坡科技与设计大学
博士研究生
通信工程
研究方向：5G 系统的优化，包括资源调度，云计算和雾计算



新闻动态

蔡小强教授入选广东省 “珠江人才计划”领军人才



2016年广东省“珠江人才计划”领军人才评选结果近日揭晓，深圳市大数据研究院副院长蔡小强教授榜上有名。

蔡教授于1988年获清华大学系统工程博士学位。1989-1991年在英国剑桥大学及贝尔法斯特女皇大学从事博士后研究。1991-1993年受聘任教西澳大利亚大学。1993年应聘加入香港中文大学系统工程及工程管理学系，2000年起任教授。自2016年3月起，蔡教授兼任深圳市大数据研究院副院长。

蔡教授是国家杰出青年科学基金（海外类）获得者、国家“千人计划”特聘专家。蔡教授的研究主要集中在工业与系统工程、最优化、供应链与物流管理等领域，已在调度理论和应用、供应链管理、投资组合优化等方面取得了许多世界领先的成果，出版了4本学术专著，发表学术论文200多篇，其中包括主流学术刊物论文100多篇。蔡教授作为项目负责人或主要参与者承担研究课题20多项，多项研究成果已经应用于实业界，创造了可观的经济效益。与此同时，蔡教授在管理科学与运筹学协会（INFORMS）、中国运筹学会调度专业委员会、中国系统工程学会、中国决策科学学会等多个学术组织中担任委员或常务理事。

广东省领军人才计划自2008年起实施，重点引进中国科学院院士、中国工程院院士和同等等级担任省级重大科技项目首席科学家、重大工程项目首席工程技术专家、管理专家。此前，我研究院的罗智泉院长入选2015年度广东省领军人才。

深圳市大数据研究院研究项目获深圳市科技创新委员会资助

深圳市大数据研究院杨柳青博士申请的深圳市知识创新计划基础研究项目（学科布局）《分布式能源智能电网关键技术研究》，新近获得深圳市科技创新委员会批复并予以资助。资助金额300万元人民币，项目执行期3年。该项目旨在针对不同复杂度设备，结合传统物理建模以及基于数据分析建模进行不同程度的建模；建立分布式能源产出与天气因素的关系模型，找到主要和次要因素以及相应的动态关系；基于能源产出预测模型，采用鲁棒优化算法，提出一套智能电力调度方案以最大化利润。

深圳市大数据研究院李文烨博士申请的深圳市知识创新计划基础研究项目（自由探索）《面向医疗与健康服务的人机自然语言对话系统研究及实现》，新近获得深圳市科技创新委员会批复并予以资助。资助金额30万元人民币，项目执行期2年。该项目旨在实现一个面向医疗与健康服务的人机对话系统，使得用户可以通过跟计算机自然语言对话的形式，获得所需的医疗与健康信息。

深圳市大数据研究院陈怿博士申请的深圳市知识创新计划基础研究项目（自由探索）《面向SDN数据中心网络的流量调度和路由算法》，新近获得深圳市科技创新委员会批复并予以资助。资助金额30万元人民币，项目执行期2年。该项目旨在为基于SDN的数据中心网络设计复杂度低的流量调度算法和路由算法，使得数据中心网络能够快速高效地运转，满足各种业务的需求。

深圳市科技创新委员会领导访问大数据研究院

2017年5月9日，深圳市科技创新委员会智能装备制造处冯石林处长一行三人访问深圳市大数据研究院。深圳市大数据研究院院长罗智泉教授、副院长蔡小强教授，研究员李文烨博士、尹峰博士、陈怿博士、周健君博士、张纵辉博士参加了座谈。

罗院长对客人的来访表示热烈欢迎，并介绍了深圳市大数据研究院的历史、团队成员及目前已取得的成果；李文烨博士介绍了研究院在分布式能源管理领域的最新研究进展和规划。其后，双方还就研究院建设、经费申请、科研设施等话题进行了深入的交流和探讨。



福建省食品药品监督管理局到访 深圳市大数据研究院



2017年6月2日，福建省食品药品监督管理局副局长黄玲女士等一行到访深圳市大数据研究院，调研发学习大数据在食品药品安全相关工作中的运用。双方就食品药品安全相关的检验检测、违法行为、行政处罚等数据的整合方案及合作意向进行了商

讨，旨在加强数据开放，保障社会公众对食品药品安全信息的知情权，运用大数据对食品药品安全进行监测，及时发布风险预警研判，助政府不断提高食品药品风险预警的防控能力。出席本次会议的还有福建省食品药品监督管理局信息中心主任马林军、科技处副处长林颖怡、科技处副调研员李海明、信息中心业务科长林旭、深圳市大数据研究院院长罗智泉、深圳市市场和质量管理委员会相关工作人员及深圳市大数据研究院相关科研、行政人员。

大数据研究院研究员应邀 在LocWorld'2017发表主题演讲

应大会组织者和中国翻译协会共同邀请，2017年3月2日深圳市大数据研究院项目负责人李文烨博士在第33届本地化世界大会发表题为：《AI, Big Data and Language Collaborative Innovation（人工智能、大数据和语言的协同创新）》的主题演讲。在演讲中，李博士重点介绍了人工智能和大数据在语言处理，特别是机器翻译，领域内的应用现状和前景展望，以及如何促进语言、人工智能和大数据技术的共同发展和协同创新。会后，逾百位来自世界各地的翻译界和本地化工作人员，以及来自IBM、Google、阿里巴巴、华为等科技企业的技术人员就相关技术问题开展了热烈的讨论，多家从事语言处理的专业公司纷纷表示要与深圳市大数据研究院建立合作关系，在语言领域利用人工智能和大数据技术寻求新的应用。



深圳市大数据研究院新型计算服务器投入使用

经过一个月的试运行，大数据研究院的一台新型计算服务器于2017年5月投入正式使用。该服务器基于CPU/GPU计算构架，配备了了了两颗最新型的intel E5-2697v4中央处理器（每颗处理器具有18个计算核心）和两颗nvidia P100图形处理器（每颗处理器具有3,584个计算单元），其速度峰值超过每秒20万亿次浮点运算。该服务器同时具有256GB内部存储器和8,000GB外部存储器，可以有效地满足研究院所需的大数据计算和分析任务。



大数据研究院

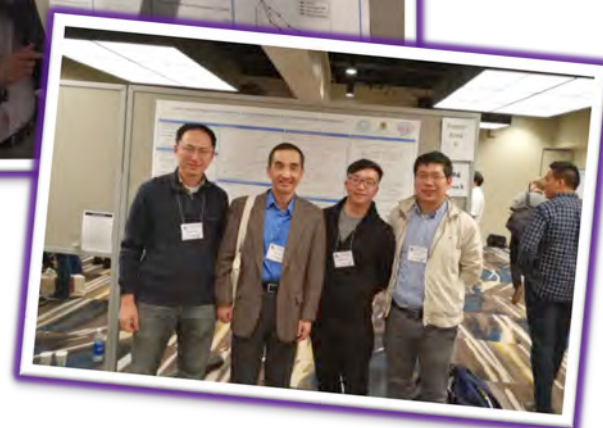
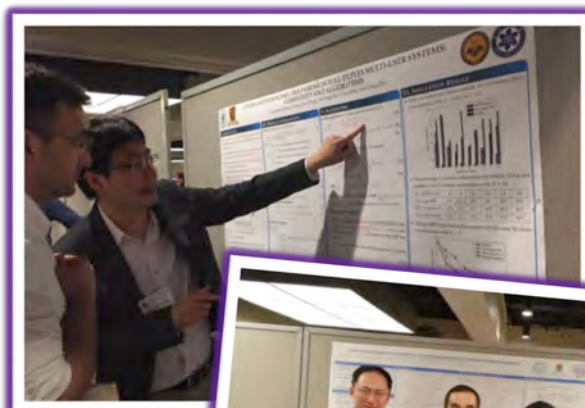
研究员出任“挑战杯”评审委员

应深圳市信息职业技术学院邀请，大数据研究院尹峰博士、李文焯博士出任《第四届“挑战杯”大学生课外学术科技作品竞赛》评审委员会委员。在2017年3月22日举行的终审决赛中，尹博士、李博士对同学们提交的十余项科技作品进行了专业细致的评议，并介绍了相关领域的发展前景，获得了参赛团队老师和同学们的热烈欢迎。



大数据研究院成员出席国际信号处理年会

2017年3月5日至9日，深圳市大数据研究院的成员出席了2017年国际信号处理年会（ICASSP）。该会在爵士音乐之都（新奥尔良，美国）举行。ICASSP是世界上最大和最全面的技术会议，专注于信号处理及其应用。



讲座报告

机器学习技术在肿瘤转移预测中的应用

时间：2017年3月20号

嘉宾：邱育珊教授（深圳大学助理教授）

嘉宾介绍：

邱育珊教授本科毕业于华南师范大学。她于2015年从香港大学获得博士学位。从2015年9月至2016年5月，她任美国西北大学博士后研究员。从2016年6月至2016年8月，她在美国贝勒医学院任职博士后研究员。邱教授现为深圳大学助理教授。她的研究兴趣包括生物信息学，机器学习与优化。



内容摘要：

肿瘤转移涉及癌症细胞从最初的位置到远处器官的扩散。尽管人们在癌症治疗方面已经取得了进步，肿瘤扩散依旧是癌症患者死亡的最大原因。因此，了解瘤转移的途径对于发展有效治疗这种致死疾病的方法至关重要，在转移过程中，肿瘤细胞经常会掠夺上皮细胞间质转型（EMT）的发育程序，以此入侵周边组织及远处器官。识别EMT的标志对于抑制肿瘤转移十分重要。为了达到这个目的，我们将会将已有的RNA序列导入器械的学习功能中，以此达到识别EMT标识以及进一步将病患情况分类的目的。在这次的讲座中，我们会集中介绍应用机器学习技术预测肿瘤转移。另外，就支持向量机SVM而言，我们提出了一个新的核心构架方式来改善分类的功效。

Wasserstein度量标准在机器学中的福音

时间：2017年3月21日

主讲人：GAO Rui 博士研究生（乔治亚理工学院）

嘉宾介绍：

Gao Rui于2013年获得西安交通大学的学士学位。目前是美国乔治亚理工学院的博士生。他的研究兴趣是在具有不确定性情况下的数据驱动决策，应用背景包括机器学习，系统设计和电力系统。



演讲摘要:

散度的测量，比如说K-L散度，在机器学习中发挥了极其重大的作用。然而，他们同时也暴露出了许多会阻碍发展的问题。为了清除这些障碍，人们引入了Wasserstein度量标准作为其中一种待选的解决方案。演讲的第一部分提及了在统计学习中的分布式鲁棒框架，该框架为所有在特定的概率度量空间意义上趋近于实证分布的分布做出了很好的概括。传统沿用的Phi-散度也许不再表现优异。Gao Rui向我们展示了Wasserstein度量标准应用于常规设置，实现易处理和可升级的再形成，以及产生更少传统而糟糕分布的功能。这是与Anton Kleywegt共同工作的成果。

在演讲的第二部分，Gao Rui与我们分享了一个最新的研究成果，这个是Arjovsky等人在Wasserstein生成对抗网方面的研究成果。考虑到Wasserstein度量标准，新的演算法有更多的训练空间，在图像生成方面展现了无限的前途。



深度学习读书会

深度强化学习

日期：2017年5月3日

下午15:00-17:00

演讲者：王帅先生，CUHKSZ的博士生

表征学习

日期：2017年5月10号

下午15:00-17:00

演讲者：李然先生，电子科技大学本科生

深度残差网

日期：2017年5月17日

下午15:00-17:00

讲者：郭玮先生，电子科技大学本科生

变分推断

日期：2017年5月24日

下午15:00-17:00

讲者：蒲文强先生，西安电子科技大学博士生

机器翻译

日期：2017年6月21日

下午19:00-22:00

讲者：黄天健先生 & 张荣升先生
复旦大学硕士生

优化算法

日期：2017年6月23日

下午15:00-18:00

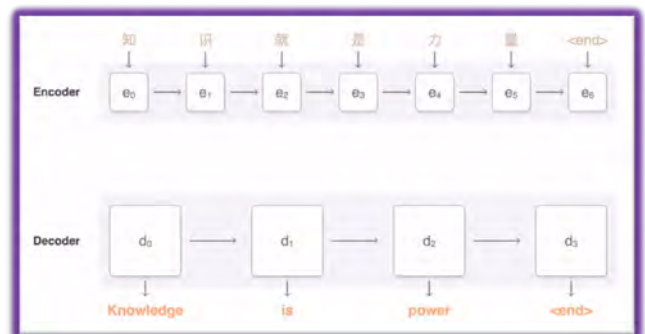
讲者：X. Y. Chen，爱荷华州立大学博士生

研究院项目

基于神经网络的机器翻译系统

基于神经网络和深度学习技术，深圳市大数据研究院李文焯博士研究组开发出一个中英文机器翻译系统。该系统由3-4层递归神经网络（Recurrent Neural Network）和卷积神经网络（Convolutional Neural Network）构成，通过大量中英文例句的训练，该系统可自动发现例句之间的对应关系并做出高质量的中英翻译。

研究人员对系统参数做了多次优化，使得其运行速度大大提高。在最新配置的CPU+GPU计算平台上，该系统可以在48小时内完成1000万中英文例句的训练。目前，研究人员正在进一步整合该系统，并引入客观评价指标评价翻译质量。



```

I tensorflow/core/common_runtime/gpu/gpu_device.cc:975] Creating TensorFlow device (/gpu:1) -> (device: 1, name: Tesla P100-PCIE-16GB, pci bus id: 0000:82:00:0)
Reading model parameters from train/translate.ckpt-203000
Model loaded succeed
> 世界人民爱好和平。
The world 's people are peaceful .
> 在过去三十年，中国经济取得了高速发展。
China has made a steady economy in the past 00 years .
> 中国提出了和平共处五项原则。
China has proposed five peaceful principles of association .
> 中国和日本建立外交关系，对两国人民而言是一件大事。
China and Japan have established diplomatic relations , which is a vital one for the peoples of the two countries .
> 最近几年，世界的粮食产量保持稳定。
In recent years , the world food production has been stable .
> 随着信息技术的高速发展，人们的生活水平得以改善。
As the information technology has been developed , the standard of living in the lives of the population is improved .
> 我们反对一切形式的恐怖主义。
We are against terrorism in all its forms .
> 世界杯足球赛每隔四年举行一次。
The UNK has been held every four years .
> 美国总统选举每隔四年举行一次。
The United States elections are held every four years .
> 贫富差距的加大，成为很多地区政治不稳定的根源。
The gap between the rich and the poor is a source of political instability in many regions .
    
```

移动互联网用户行为分析

移动互联网用户行为分析是移动大数据一个重要的应用，可为运营商个人化营销，网络资源优化，智慧运营提供重要的数据依托。深圳市大数据研究院和中国移动（深圳）有限公司全面合作研究多源异构移动大数据的融合，分析移动用户行为，检测监控异常用户。团队将在崔曙光教授和罗智泉教授的带领下在项目实施期内完成移动大数据计算平台的硬件和软件建设并将和中国移动（深圳）公司一起，把研究成果转化为实实在在的工业或商业应用，在完成目标产值的同时也产生显著的社会影响力和服务效应。



自然语言志愿者工作

为了对自然语言研究项目做准备，大数据研究院发起了自然语言研究志愿者工作。从四月初开始，截止至五月底，已有312人次参与了志愿者的工作，工作时间长达156小时。

志愿者的具体工作为：听一段包含医生与病人对话的音频、找到其中的问诊对话内容、判断人物及其明显的情绪、用文字的形式记录在软件上，并标注对应的音频时间。这些音频结合文本记录可以用于进行语音、人物和情绪的识别，并用来做人与机器对话的研究等。

人工智能

Artificial Intelligence

• 前言

为了满足人们日益增长的医疗需求，目前全球的医生、护士和其他医疗工作人员的资源很是紧缺，且这样一种局势正在不断加剧，对于服务水平低下的地区来说情况更是不容乐观。同时，培训医生和其他医疗工作者是一个较为艰难的过程，医护人员需要多年的教育的铺垫和经验的累积。但幸运的是，人工智能（Artificial Intelligence (AI)）飞速发展使得其可以帮助医疗行业克服当前和未来的挑战。本篇文章将以此为切入点，简要总结人工智能算法和软件如何改善医疗保健服务的质量以及其可实施性。

• 人工智能作为健康助手的应用



目前来看，人工智能最基础但最有效的用例之一便是应用于优化临床过程。传统上，当患者生病时，他们会选择去看医生、检查他们的生命体征。而医生也会相应地问一些与治疗相关的问



题，然后给出处方。现在，AI 助手可以应付很大一部分临床和门诊的案例，从而让医生有时间去看诊更多的关键病例。

Your.MD 是一个移动端上的人工智能应用程序，它可以提供基本的医疗保健服务。它会先向用户询问他们的症状，然后给患者提供易于理解的有关其医疗状况的信息。除此之外，Your.MD 会建议患者采取相应措施来治疗疾病，它也会在患者需要看医生时提醒用户。不仅如此，因为该平台拥有庞大的信息网络，所以它能将各种症状与相对应的病因联系起来，从而可以提供实际的帮助。值得一提的是，医疗人工智能使用自然语言处理和生成来提供丰富流畅的用户体验，还利用机器学习算法创建用户状况的信息库从而提供个性化的用户体验。英国国家卫生服务局（NHS）已经证实了 Your.MD 所提供的信息的可靠性。这意味着，与自我诊断不同，用户不必担心他们通过人工智能所获得的指导是否真实以及可靠。

其他的一些健康助手也在医疗领域表现得很出色。如 Ada 将其技术与 Amazon Alexa 整合，以改善用户体验。因为 Ada 熟悉用户的

病史，它除了能够生成详细的症状评估报告外，还能够为用户提供联系现实生活中的医生的选项。另一个出色的健康伴侣 Babylon Health 通过跟踪用户以及对过去症状的补充，及在用户需要的情况下帮助用户与全科医生进行现场视频咨询，来加强人工智能本身所能提供的帮助。

总的来说，人工智能作为健康助手可以帮助病人解决一些不必亲自到医院咨询的疾病问题，从而可以节省医生的时间让他们去解决更棘手的病例。此外，在医生和诊所供不应求的地区，人工智能还可以节省病人等待排队所耗费的时间。

• 人工智能对于病症早期诊断的精确性

一些罕见和危险疾病能否被治疗通常取决于其能否在正确的时间检测到症状。在许多情况下，若能早期诊断出病症，疾病是可以完全治愈的。但相反，诊断迟误或诊断错误可能会导致潜在的致命结果。目前医生的诊断方法主要是通过观察样本和照片图像发现异常。然而在检查图像和样品以及依此作出可靠的决策时，人们的技能和经验往往有限并且难以达到一定高度。但 AI 算法可以快速学习数百万个样本，并发现有用的模式。不仅如此，与人类不同的是，当他们的使用时间增长时，他们不会像人类变老一样从而失去自己的优势。

目前，有几家机构和公司正在投资开发有关应用人工智能的医疗保健解决方案。斯坦福大学的研究人员创建了一个可以识别皮肤癌的 AI 算法。他们通过 13 万张图像中的痣、疹和病变来训练他们的深度学习算法。从结果来看，通过使

用 AI 算法来诊断皮肤癌的效率与专业医生的诊断效率旗鼓相当。斯坦福大学的研究人员希望将来这一应用程序在移动端上也能被应用。

DeepMind 是 Google 拥有的一家 AI 公司，它正与 NHS 合作希望使用机器学习来治疗失明。该公司的研究人员正在通过匿名的一



百万双眼睛的扫描样本来训练其深度学习算法。这将有助于在早期阶段发现诸如湿性年龄相关性黄斑变性和糖尿病性视网膜病变等眼科疾病。据专家介绍，在某些情况下，他们有可能最终能够预防 98% 的最严重的视力损失。

Morpheo 也是一个 AI 平台，它能够帮助诊断睡眠障碍。分析睡眠模式的传统过程较为复杂，且需要消耗大量的时间。但通过机器学习算法的帮助，Morpheo 通过自动识别睡眠模式从而帮助医生进行更高效地诊断。Morpheo 的创造者认为这将有助于实现预测性和预防性治疗。

• 人工智能用于疾病的动态监护

对于某些疾病而言，确定正确的治疗途径并使其适应患者健康状况的变化是至关重要但同时也是极具有挑战性的。

IBM 正在使用自己的人工智能进行抗癌斗争，他们的 Watson for Oncology 平台已经准备投入生产。该平台将被首先应用于佛罗里达州的一家医院，其目的是帮助治疗癌症患者。Watson 特别擅长分析结构化和非结构化的数据。它吸收了临床试验数据和医学杂志中的数据，然后向癌症治疗团队提供新的治疗模式，以



及一系列有效的疗法和治疗方案。北卡罗来纳大学医学院的专家通过使用 Watson 对一千个癌症病例进行诊断来评估这一人工智能的性能。数据显示，该平台在 99% 的案例中提供了与专业肿瘤学家相同的建议。类似于 Watson for Oncology 平台这一类人工智能的成功应用，对于缺乏人力资源和技术资源的小型医院和医疗中心来说可能是至关重要的。

其他公司也正在使用人工智能来处理疾病的动态监护中一些微小但关键的步骤。其中的一个例子便是 AiCure，它通过使用移动端应用程序的图像分析来控制患者对处方的依从性。它的功能包括确保患者按时服药，以及执行主治医师规定的其他任务。这对于有严重医疗状况的病人和有可能违反医生处方的病人是有帮助的。

• 人工智能预测心脏病发作概率

据估计，每年有二千万人死于心血管疾病。而幸运的是，来自英国诺丁汉大学的一组研究人员已经开发了一种机器学习算法，可以预测用户心脏病发作或中风的可能性。

美国心脏病学会/美国心脏协会（ACC / AHA）基于年龄、胆固醇水平和血压等八个因素，制定了一套评估患者心血管风险的指南。平均而言，英国诺丁汉大学研发出的这一系统，以 72.8% 的精确度正确地估测了一个人患有心血管疾病的风险。从数据上来看，成果是十分乐观的，但是 Stephen Weng 和他的团队打算做得更好。他们建立了四种计算机学习算法，然后收集了来自英国的 378,256 名患者所提供的数据。系统首先使用大约 295,000 条记录来生成其内部预测模型，然后他们使用剩下的记录来测试和改进模型。结果表明，这一算法显著优于 AAA / AHA 指南，其精度也从 74.5 % 提高到了 76.4%。其中，利用神经网络的算法测试精度最高，高过现有指南 7.6%，并且减少了 1.6% 的虚假警报。此外，在另一组有 83,000 个患者的测试记录中，据估计该系统可以额外多拯救 355 个生命。

有趣的是，这一人工智能系统补充了现有指南中未涵盖的一些风险因素和预测因子，如严重的精神疾病和口服皮质类固醇的消费。Weng 说到，“生物系统有很多内在互动，这是人体的真相，而计算机科学能够让我们探索这些内在的关联。”

• 一场利用人工智能检测疾病的竞赛

日前，一场旨在利用人工智能自动检测肺癌的竞赛向大众展示了机器学习如何检查医学成像。该挑战为最准确地识别出低剂量的计算机断层扫描图像中肺癌症状的算法提供了 100 万美元的奖品，虽然获奖的算法不一定会被临床医生采用，但它们可以激发人工智能应用于医学成像上的算法创新。

近年来，低剂量 CT 扫描已经展现出了较早发现肺癌的潜力。它们使用较少的辐射，并且不需要将对比染料注入体内。但这种诊断仍然是十分困难的，存在大量的错误阳性诊断以及太多不必要的医疗程序。

而同时，被称为深度学习的机器学习技术已经被证明在图像的特征识别上十分有效。这种机器学习方法也越来越有希望可以通过自动识别疾病的相应特征来帮助提高医学诊断的标准，其所能识别的病理特征也包括那些对于人眼来说所难以发现的特征。

正如前文所提到的那样，深度学习已经被用于检测皮肤癌，其检测结果的错误率与专业皮肤科医生大致相同。此外，这一类技术已被证明通过分析视网膜图像从而探测失明的常见原因也是有效的。医生和企业家也对于拓展这一类技术的应用层面越来越有兴趣。然而，在这种情况下，研发者们可能需要更加努力使得这一技术被人们所信赖。

提供了此次比赛中所使用的成像数据的国家癌症研究所的项目主任 Keyvan Farahani，

表示说减少由低剂量 CT 扫描造成的肺癌误诊数量将对患者产生真正的影响。根据美国癌症协



会的数据库，美国每年有大约 222,500 例新发肺癌。Farahani 认为现有的识别肺癌症状的软件是不可靠的，他说：“初步的结果表明顶级算法比现有的算法更完善。”不过，Farahani 说他并没有预见到能够代替医学专家的算法。“深入学习将有助于消化大量数据，但我认为他们将不会替代放射科医生的角色。”

这次比赛的主要挑战之一在于组委会给每支参赛队伍只提供了 2000 张图片，然而机器学习通常需要非常大的数据集才能开发出一种有效的算法。但在所提供的数据集中，其他数据比如所使用的设备的细节是被包括在内的。获奖团队采用了神经网络的训练方法，并且通过对图像进行注释从而获得了更多的数据。此外，他们还使用了一个额外的数据集，并将挑战分为两部分：确定小瘤与诊断癌症。来自中国最重要的学术机构之一——清华大学的获奖团队成员 Zhe Li 说，我们认为将这个问题明确地分为两个阶段是至关重要的，因为这似乎正是人类专家所做的。但目前尚不清楚最佳的算法如何符合医生的标准，因为每个算法提供的

是一个概率而不是确定的诊断结果。

• 一家由人工智能驱动的基因组学公司转型成为制药公司

Deep Genomics，一家利用机器学习来追踪潜在的遗传因素的加拿大公司，在周二宣布他们正打算向开发药物发展，成为众多人工智能中的一员，其技术可以允许研发者通过在大量基因组数据中发现微小的信号来帮助生产有效的新型药品。

Deep Genomics 由多伦多大学教授 Brendan Frey 创立，他本人专门从事机器学习和基因组学。Frey 曾是计算机科学家，曾经在多伦多大学就读，并跟随深度学习的领袖人物 Geoffrey Hinton 进行学习。他的公司使用深度学习和极为庞大的神经网络来分析基因组数据。到目前为止，该公司正专注于研究可能与特定疾病有因果关系但难以检测的突变基因组。该公司首先将重点放在开发治疗孟德尔病症的早期药物。

孟德尔一种由单一基因突变而引起的遗传性遗传疾病。据估计，这种疾病对全球 3.5 亿人造成了不同程度的影响。

研发者之所以迫切地将 AI 技术应用于医药和药物开发，部分是由于有更为完善的新算法的出现，另一部分是由于有对人类整个基因组进行排序的更为高效的新方法的出现。Frey 说：“以信息为基础的医学新时代正要到来，你可以获得很多不同类型的数据，而我们处理大量数据的最佳技术便是机器学习和人工智能。”

近年来，深度学习通过训练大量数据成为了描绘抽象图景非常有效的方式，特别是在语音识别和分类问题上成就非凡。这一种方法目前正在迅速拓展包括医学在内的更多领域，比如它提供了一种在医学成像中发现疾病的方法，以及从患者的病历记录中预测疾病。

加州大学旧金山分校医学院的助理教授斯蒂芬·桑德斯专门负责使用基因组学和生物信息学来研究疾病，他说，深度学习可以通过结合稀疏的病理学数据与大型基因组数据来寻找药物开发的新模式。

其他几家公司也正在尝试将机器学习应用



于药物开发。这些公司包括一家英国的 AI 公司 BenevolentAI，以及 Alphabet 的子公司 Calico。BenevolentAI 创始人 Ken Mulvany 说，他的公司利用大部分未使用的研究数据来研究炎症和神经退行性疾病，以及一些罕见的疾病。Mulvany 说：“开发药物始终是一个非常冗长，危险且昂贵的过程，并且具有高消耗率。但是，目前尚未有任何开发计划的制药研发机构中有大量等待开发的数据可以利用。”

• 结语

人工智能在许多领域获得全面牵引之前，仍有一些需要克服的挑战。医疗保健也不例外，特别是在人工智能的应用以及用户隐私保护的协调方面。比如，去年 DeepMind 在与 NHS 的数据共享交易中就与英国当局和隐私权维护团体发生了冲突。医疗信息是极为敏感的，因此，一些公司和机构在处理这些信息时需要着重把关他们对于收集、存储和分享信息的政策实施。

一些公司正在考虑通过使用区块链，一种支持 Bitcoin 和 Ethereum 的分布式分类账的方式来作为解决方案。例如，Morpheo 通过使用区块链来确保患者数据在其平台上的透明度和隐私性。

另一个开放性的问题是人工智能将如何影响医疗行业的工作形势。从现阶段来说，人类对彼此的呵护仍然属于人类的工作范畴，还没有算法能够模拟医生或护士的社会和职业功能。事实上，机器人并没有取代人类，而是鞭策人们更加努力，以提高医疗服务的整体质量和可实施性。基于人工智能的医疗保健工具目前所提供的建议性作用会在将来的某一天转变成决策性作用吗？我们不知道，只有时间能告诉我们答案。但最近人工智能的发展表明，仍有不少惊喜等着人类去收获。

注：本文内容参考 MIT Technology Review 和 Flipboard 的如下文章

<https://flipboard.com/@thenextweb/-how-artificial-intelligence-is-revoluti/f-0616fb05d0%2Fthenextweb.com>

<https://flipboard.com/@Engadget/-ai-can-predict-heart-attacks-more-accur/f-fc68f93d52%2Fengadget.com>

<https://www.technologyreview.com/s/604339/million-dollar-prize-hints-at-how-machine-learning-may-someday-spot-cancer/>

<http://www.technologyreview.com/s/604305/an-ai-driven-genomics-company-is-turning-to-drugs/>



撰稿：李杰

编辑：王文琪、陈怿

指导：罗智泉



深圳市大数据研究院

中国广东省深圳市龙岗区龙翔大道 2001 号道远楼 225 室

邮编：518172

电话：(86)755-84273615

<http://sribd.cn/index.php/cn/>