

58%

过去半个多世纪,人类一共解析了5万多个蛋白质结构,人类蛋白质组里大约17%的氨基酸已有结构信息,而AlphaFold2预测的结构将这一数字从17%提高到58%。它带来的在生命科学各分支领域的革命,将在今后几年到十几年中逐渐显现出来。

预测蛋白质结构只是开始 AI或为生命科学领域带来巨变

◎本报记者 崔爽

蛋白质结构预测是生物学的重要“圣杯”,也是人工智能落子生命科学领域最炙手可热的研究之一。

近日,我国自主研发深度学习蛋白质折叠预测平台TRFold传来好消息,其基于2020年第14届国际蛋白质结构预测竞赛(CASP14)蛋白质测试集的成绩仅次于“阿尔法折叠的迭代版”(Alpha-

Fold2),排名全球第二,这是国内目前所有公开蛋白质结构预测模型中的最好成绩,我国计算生物学领域的表现跻身全球第一梯队。

从2018年AlphaFold第一次代表人工智能“参战”,到AlphaFold2用机器学习方法取得媲美结构生物学实验的精确精度,计算生物学给蛋白质预测这一世纪难题带来了颠覆性的解法。人工智能会给生命科学领域带来怎样的巨变?蛋白质结构预测这个生物学里悬而未决的终极难题之一,会被人工智能彻底解决吗?

深度学习可在计算生物学领域广泛应用

蛋白质结构预测是生命科学领域一个由来已久、令人着迷的问题,同时又以难度大、成本高、进展有限著称。但这个人们本以为需要一个世纪慢慢探索的问题近年来却取得了重大突破:2020年CASP14竞赛中,谷歌公司旗下的DeepMind公司研发的AlphaFold2取得了总分(GDT)92.4/100的成绩,也就是说,计算生物学几乎获得了与实验室方法精确度相当的蛋白质结构预测结果。

这个里程碑事件令结构生物学家们感慨,自己用价值1000万美元的电镜努力了好几年得出的结果,AlphaFold2竟然一下就算出来了。“依我之见,这是人工智能对科学领域最大的一次贡献,也是人类在21世纪取得的最重要的科学突破之一。”生物物理学家、西湖大学校长施一公不吝赞美。

为什么预测蛋白质结构?
天壤蛋白质折叠项目负责人苗洪江对科技日报记者解释,“研究蛋白质结构,有助于了解蛋

白质的作用,理解蛋白质如何行使其生物功能,认识蛋白质与非蛋白质之间的相互作用,对于生物学、医学和药学等都非常重要”。

传统观测蛋白质结构的方法主要有3种,即核磁共振、X射线、冷冻电镜,但这些方法往往依赖大量试剂和昂贵的设备,每种结构的研究都要花费数年时间。而人工智能应用于蛋白质结构预测的最新成果,即AlphaFold2,能在几天甚至几分钟预测出以前要花费数十年才能得到的具有高精度度的蛋白质结构。

“刚开始大家还在开玩笑,说DeepMind是不是通过什么方法盗取了真实实验结果,直到大家看到文章和开源代码才敢相信这件事情真的发生了。”苗洪江笑言,这侧面证明AlphaFold2预测结果之震撼,“这开启了人工智能在计算生物学广泛应用的大门,让整个领域的人看到了深度学习在这个领域可以广泛应用,这实打实的双盲实验结果就是证明。”

AI预测结果和实验室水平相当

1994年,美国科学家约翰·莫尔特(John-Moult)发起国际蛋白质结构预测竞赛,每两年举办一届,竞赛的举办正是为了吸引计算机科学、生物物理学等不同领域的专家参与到蛋白质三维结构预测这一极具挑战性的生物信息学问题中来。2018年,人工智能正式参与蛋白质三维结构的预测,AlphaFold首次大显身手,在98名参赛队伍中排名第一。两年后,AlphaFold2带来真正的突破,它用机器学习方法对几乎所有的蛋白

质都预测出了正确的结构,其中有大约2/3的蛋白质预测精度达到了结构生物学实验的测量精度。

事实上,过去半个多世纪,人类一共解析了5万多个蛋白质结构,人类蛋白质组里大约17%的氨基酸已有结构信息,而AlphaFold2预测的结构将这一数字从17%提高到58%,因为无固定结构的氨基酸比例很大,58%的结构预测已经接近极限了。它带来的在生命科学各

仅凭聊天频率 这个算法就能辨别谁是潜在网络违法者

◎孙琳钰

随着网络的普及,网络违法事件也成为了不容忽视的社会问题。互联网消弭了潜在违法者和受害者之间的时空距离,使得人人都有客观条件违法,而人人也有被伤害的危险。中国司法大数据研究院发布的《网络犯罪特点和趋势(2016.1—2018.12)》报告指出,社交类平台,尤其是QQ、微信等已经成为虚拟犯罪的主要工具,不法分子通过它们在网络上策划、实施犯罪行为。这种过程不需要现实接触,因此非常难捕捉,给执法带来了许多困难。

近日,日本德岛大学的计算机研究者联合日本大型网络公司代理商Cyber Agent在《人类行为计算》上发表论文,他们用机器学习的方法,分析了Cyber Agent旗下一款社交类游戏的使用数据,并且在不用监测聊天内容的情况下,仅基于聊天次数、聊天对象、聊天时间等基本信息,就能较为精准地识别出潜在网络违法者,并预测出违法行为的大概时间。

“疑犯追踪”的理论基础

这不是一个异想天开的想法。尽管在游戏

中大家只是依靠网线交流,但我们在网上的行为也留下了海量数据,为预测网络违法行为提供了丰富的材料。

研究者基于两种传统犯罪学理论开发了这套算法:日常活动理论和社会传染理论。

日常活动理论提出,许多犯罪行为并不是随机发生的,犯罪者和被害人往往在日常活动中有交集。例如,在现实生活中,小偷在盗窃前会去目标地点踩点,并观察目标人物的行为规律;同样的,网络上的犯罪者更需要提前与“猎物”取得联系,套取信任。因此,玩家的社交活动数据中或许就藏着“犯罪预告”。

另外,社会传染理论还补充了重要的一点:违法倾向或违法行为也会传染。最常见的例子就是网络暴力。网络暴力往往来源于某种过激情绪的广泛传播:在群体的裹挟下,有的人不知不觉就失去了独立判断能力,无意间成为了网上的施暴者。有研究指出,在“目睹”群体内其他人的网络骚扰行为后,旁观者也很容易对同一名受害者发起攻击,或者尝试骚扰其他人。这样的传染行为也为预测网络违法事件提供了重要的对象和时间线索。

在这两种理论的基础上,研究者选择了一款名叫Pigg Party的手机游戏。它主打社交功能,用户登录账号后,可以装扮虚拟的房间



分支领域的革命,将在今后几年到十几年中逐渐显现出来。

施一公在接受媒体采访时曾谈到,人类蛋白质组里能够被预测的以单个蛋白为单位的空间三维结构,已经基本都被AlphaFold2预测了。总体而言,预测结果可信,也比较准确。对结构生物学来说,这是一个颠覆性突破。此前人类尚未被解析的一些结构,现在基本上都已经被预测。

对生物化学、细胞生物学、遗传发育、神经生物学、微生物学、病理药理等一大批生命学科和研究领域来说,这会大大改进人们对于生命过程

单个蛋白质结构预测只是起点

今年7月,DeepMind公开了AlphaFold2的源代码,并在《自然》上发表论文阐述了AlphaFold2的技术细节。

“这次开源在生物学界掀起了巨大的波澜,意味着生物学家终于摆脱了先进设备的掣肘——此前这些昂贵的先进设备只有经费充足的大学或研究机构才有条件配置,而此后,小型团队或者个人研究者也有了参与蛋白质研究的可能。”天壤创始人、上海交通大学计算机系副教授薛贵荣说。

苗洪江认为,目前的单个蛋白质结构预测只是一个起点,更加精准的侧链优化、蛋白质的动态分析、蛋白质与其配体(如小分子、DNA、RNA、多肽、蛋白质等)的相互作用等一系列的问题还没有解决,接下来的工作重点将是利用目前的全蛋白质组协同进化分析,建立起蛋白质与蛋白质之间相互作用的精准链路。

有了算法模型只是开始,向前走依然困难很多,薛贵荣坦言:“算力是很大的制约因素,比如AlphaFold2做了大量的数据蒸馏工作,他们的算

法模型是基于30%的真实数据和70%的蒸馏数据一起训练的,背后是巨大的算力支持。”

充足的算力可以让蛋白质结构预测从单一结构向相互作用、从两两研究向规模化、从微观结构向宏观系统前进,“生物界有很多蛋白结构,比如基因组测序大概已经测了几十亿条序列。但我们只知道序列,不知道结构,这就是很大的信息缺失问题。”薛贵荣说,“蛋白质通常以复合物的形式成对或成组地承担生命所需的种种功能。然而许多蛋白质复合物的结构仍然是谜,蛋白质之间的相互作用也尚未被识别。我们需要有充足的算力去支撑整个体系,进行蛋白质结构预测、蛋白质设计,研究蛋白质相互作用,药物研发等漫长而富有挑战的工作,寻找精准的疾病治疗新方法。”

同时,在数据来源和应用方面,也需要药厂、医院等进行协同和联动。“未来更多的医药企业、机构,以及人工智能公司,要一起把这个行业做大,现在只是个开始。”薛贵荣说。

研究者认为,借助AI,系统管理员能更好地维护网络安全和用户权益。尽管目前的AI不能进一步预测违法事件的具体类型,但它或许能抓住网络上“隐形”的违法者,更好地保护我们的安全。

和个人形象,与朋友或陌生人以私聊、群聊、公聊的方式进行交流。研究人员采用擅长从复杂数据中提取特征的算法——多层非线性模型,对55万用户6个月内产生的聊天数据,包括每名用户的聊天频率、聊天时间、消息的接收者等进行了分析。

想做坏事的心逃不过AI的眼

研究者组合多种神经网络模型和算法,搭建

了预测违法事件的人工智能。性能测试结果显示,AI能根据用户数据,较为准确地预测未来的违法者和受害者账号。输入用户两个月内聊天的时间、频率、对象,AI对接下来两个月内违法账户的预测准确率可以达到84.85%,对受害者账户的预测准确度也接近85%。

除了对个体账户违法或受害的风险有较好的预测能力外,只需提供一周内的用户活动数据,AI就能基本精准地预测接下来的一周网络社区里发生违法事件的时间,对小时和日期的预测准确率高达95.83%和85.71%,并且结果与预测受害给出的时间相吻合。更有趣的是,AI分析数据后预警发生违法事件的时间,并不一定处在以往发生违法事件的时间段中,可见它掌握的真并不仅仅是固定的规则,还有违法者言行中的真正“逻辑”。

可进行违法预测的AI将海量、零散分布的用户日常活动记录压缩转化成可以定量分析的数据,并从中提炼、理解规律,最终形成了强大的预测能力。研究者认为,借助AI,系统管理员能更好地维护网络安全和用户权益。尽管目前的AI并不能进一步预测违法事件的具体类型,但它或许能抓住网络上“隐形”的违法者,更好地保护我们的安全。

(据《环球科学》)

情报所

联合国教科文组织通过首份 人工智能伦理问题全球性协议

新华社讯(记者陈晨)总部位于法国巴黎的联合国教科文组织近日举行新闻发布会,介绍该组织正式通过的首份人工智能伦理问题全球性协议。

该协议名为《人工智能伦理问题建议书》(以下简称《建议书》),旨在促进人工智能为人类、社会、环境以及生态系统服务,并预防其潜在风险。《建议书》包含规范人工智能发展应遵循的原则以及在原则指导下人工智能应用的领域。

联合国教科文组织总干事阿祖莱当天在新闻发布会上表示,人工智能除应用于传统科学研究领域外,当前还积极助力新冠疫情防控和疫苗研发等诸多方面,但同时也带来隐私保护问题、性别偏见等挑战。

阿祖莱说,世界需要为人工智能制定规则以造福人类。教科文组织通过该领域首个规范性全球框架协议,是多边主义取得的重要胜利。教科文组织将为各会员国落实《建议书》内容提供支持。

智能抓拍 能识别违规操作,还能解决鼠患

科技日报讯(张绍忠 陈科)“以前检查食堂员工健康,管理员要对员工进行体温检查、手部卫生检测、穿着检测并将结果记录在案,过程繁琐还有可能测不准。而现在,我们用食品安全健康检测机器人进行检查,再通过AI智能自动抓拍功能采集员工体温、伤口等,一旦发现异常马上报警。这就是食品安全物联网智能管理平台带来的便利。”12月8日,四川省妇幼保健院(武侯院区)食堂技术管理员潘海军,指着新开发的“食品安全物联网智能管理平台”说。

作为成都市投入应用的首套“人工智能+物联网+互联网+云平台+大数据”食品安全物联网智能管理平台,该平台由智能食品留样柜、食品安全检测机器人、AI抓拍系统等多项技术构成,首次形成了“数据留痕可追溯、全程隐患先告警”的一体化食品安全监管。

“从前的后厨加工过程中存在抽烟、不戴口罩等违规情况,现在借助平台中的AI抓拍功能,不仅能有效发现违规行为,对其进行及时规范,还能解决鼠患。”潘海军说,AI抓拍能自动识别老鼠行动轨迹,以便工作人员对下水道进行针对性整改。“而关于食材监管,我们从之前只能通过食材的颜色、气味等查看其状态,到现在能快速通过检测仪对瘦肉精、注水肉、农药残留等抽检并上传,把好了食材安全第一关。”他说。

图说智能

智慧警局不“打烊”



在安徽铜陵公安无人智慧警局,智能政务机器人、智能文件柜、综合警务自助亭等自助服务设备为群众提供24小时服务。办事群众携带相关证件,即可自助办理户政、车驾管、出入境等3大类、32项高频公安业务。自2020年12月底启用以来,无人智慧警局已服务群众9000余人次。图为12月7日,市民在铜陵公安无人智慧警局向智能政务机器人咨询。
新华社记者 曹力摄

人工智能纹枰对弈



近日,2021世界人工智能围棋大赛半决赛在福建省福州市举行。图为人工智能“鱼图棋”在比赛中。
视觉中国供图