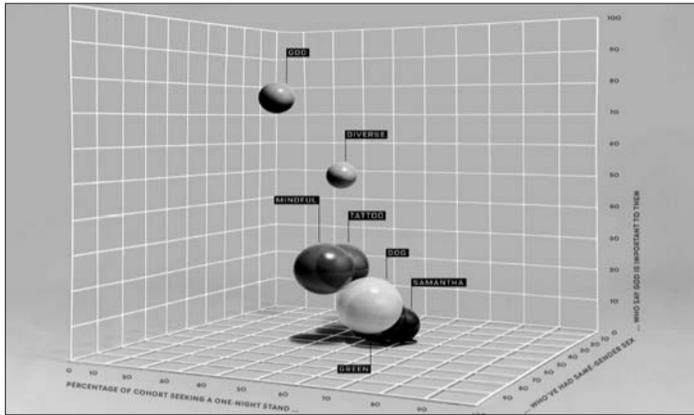


相亲?用大数据吧

谁说数学学霸没有爱的春天?美国数学博士克里斯·麦克金雷就成功运用他的数学天才,在相亲网站上找到了自己生命中的另一半。

□本报特约撰稿 曾宪妹/编译



麦克金雷(左图)在采集海量信息后,将相亲网站上的女性用算法分成了7类。

>> 数学家相亲

凌晨三点,在美国加州大学洛杉矶分校数学研究中心大楼五楼,除了一盏孤灯,就是调制解调器闪烁的灯光,这简直就是克里斯·麦克金雷准备博士论文的绝佳时间和地点。不过,电脑突然卡住了,麦克金雷趁机点开了另一个窗口,看看有没有OkCupid相亲网发来的邮件。

谁是麦克金雷?他是一个35岁、头发蓬乱的竹竿男;他还是4000万OkCupid相亲网注册用户中的一员。

他可不是个一般的数学家。麦克金雷在波士顿的郊区长大,于2001年毕业于明德大

学,拿的是中文学位。在当年8月,他在纽约找了份中译英的兼职,工作地点在世贸大厦的91楼。5周以后,世贸大厦被飞机撞毁。只能说他命大,在第一架飞机于8:46撞入北大楼的时候,他还在家呼呼睡大觉,直到下午2点才跑去上班。“自那次大难不死之后,我就问我自己,我真正想干的是什么。”

他对应用数学的兴趣,激励他考取了应用数学领域的硕士学位,继而考取博士学位。“数学家可以在各种各样的情形下运用数学,”他说,“数学家看到一个新游戏,比如牌九,他就可以回家写一串编码,想出一个策

略打败对手。”

自从上一次分手之后,他在相亲网上足足寻觅了9个月,还是一无所获。只要是OkCupid给他配出的相亲对象,他都会给她们发送俏皮风趣的自我介绍,不过信息大多都石沉大海,只有6个给了回复,同意和他约会。

麦克金雷盯着电脑屏幕,一个窗口在刷新机器编码,另一个窗口闲置的是他的相亲资料,他突然灵光一现,发现自己用错了方法。他怎么可以和一般用户用一样的方式在网上相亲?这是不对的!网上相亲,他应该运用数学家思维。

链接

用计算机算法寻找另一半

OkCupid是一家由哈佛数学专业学生于2004年成立的相亲网站。它一成立就吸引了众人的眼球,因为它有一套独特的计算机相亲算法。

这一套算法要求用户回答一系列多选题,问题涉及政治、宗教、家庭、爱情、性以及智能手机等领域。

用户需要勾选每个问题的答案,标明哪些回答是他们可以接受的,并为问题的权重打分,满分5分,1分意味着“无所谓”,5分意味着“非选不可”。然后网站用答案的匹配程度,来计算相亲者能否成功配对,计算结果越接近100%,就越可能成为相亲者的灵魂伴侣。

一般来说,相亲者要回答从成千上万的问题库中抽选出来的350个问题,比如“你最想看以下哪部电影?”“宗教、上帝在你的生活中有多么重要?”

>> 爱情大数据

首先,他需要数据。于是他创建了12个伪账号,并写了一段脚本来运行这些账号。这个脚本会自行搜寻年龄在25岁到45岁间的异性,访问她们的主页,然后根据种族、身高、是否抽烟等要素把她们的用户资料筛选出来。

他还必须通过另外的甄选来获取更多问题答案。OkCupid上的相亲问题,只有在用户自己回答过后,才能看到别人的答案。所以麦克金雷设计了一个程序,随机挑选各种问题,并简单回答。这样一来,麦克金雷就能“偷窥”很多女性的回答,并把其收入数据库中。

第一步成功之后,他收集了上千份异性用户资料,他很开心。但没开心多久,他就看到一块拦路石。OkCupid内设了一个系统,专门防范这种数据采集,所以他伪造的账号开始一个接一个地被列入黑名单。

没办法,他得让自己的黑

客软件模拟人类行为。于是他向一位神经学科学家朋友托利斯求助。托利斯在自己的电脑上安装一个“间谍”软件,模拟点击鼠标的频率和打字的速度,路障就这么被清除了。

三周之后,他采集了600万个问题,以及全国两万名女性的回答。这段时间里,他大部分时间都睡在办公室,后来,他干脆不回宿舍了,整天蜗居在这个昏暗的小房间里,困了就找个垫子,趴在桌子上睡。

要成功找到相亲对象,他必须在海量数据中找到规律,大致通过她们的相似点进行归类。他设计出了一种算法,可以轻松把调整调研结果分门别类。最终,两万名女性被自然分为了7类。“我欣喜若狂!”他表示。

现在,麦克金雷必须决定哪一类女性最适合他。他心仪的一类主要是25岁左右,长得有点像印度人的音乐家和艺术家。所以他建立了两个最优用户资料,一

个设为A组,另一个为B组。

为此,他创建了两个版本的用户资料,一个头像是他攀岩的照片,另外一个是他弹吉他的照片。对于“不考虑未来计划,现在你更想要性还是更享受爱”这样的问题,他在两个用户组上的回答都是“爱更重要”,但是针对更为年轻的A组,这个问题的权重为“非常重要”,对B组,则为“非选不可”。

当他回答完最后一个问题,以配对指数搜寻了OkCupid上的相亲对象,第一页整页的配对指数都是99%。

他还需要再走一步,让自己有关关注度。当有人访问自己主页的时候,OkCupid用户会得到提醒,所以他就编写了一个程序去访问配对指数最高的对象的主页。然后,源源不断的信息就向他涌来。

万事俱备,只欠东风。他必须得离开办公室,进行些实地调研了!是时候出来约会了。

“不过,我喜欢。”

这是他的88号相亲对象。不久他们就约出来再见面。两周之后,他们注销了自己的OkCupid账号。

后来,王田回卡塔尔了,这对情侣每天都会用Skype通话。有一天,麦克金雷拿出了一枚钻戒,把它放在摄像机前,向王田求婚。王田答道:“我愿意”。他们还不确定具体什么时候结婚。要找个完美婚期,还有一大堆调研要做,一大堆数字要运算呢!

“我利用数学知识相亲,其实就是巧用了算法,借用了计算机,进行了一场规模更大的网上寻亲,本质上和大家所做的没什么差别。”麦克金雷说,“每个人都想创建属于自己的完美档案,我只是利用数据,让电脑生成了一个。”

还是大同小异。很快,他的约会次数超过了55次。

对一般人来说,多次失败的约会会让人怀疑自己的魅力。但是对麦克金雷来说,打击还不止于此,他还要质疑自己的计算是否有纰漏。

就在这时,有个叫王田(音译)的女士给他发了条信息,她28岁,是个艺术家。她寻找身高1.8米以上、蓝色眼睛、住在加州大学洛杉矶分校附近的单身男子。她正好在那儿进修艺术硕士。他们的配对指数是91%。

他俩约在学校附近的雕塑公园见面,他们一起散了很久的步,他感到缘分到了,他们谈了很多。她说,在发信息之前,她改了些自己的资料,听完她的坦白,麦克金雷就把自己如何为爱做黑客的全过程统统说了出来。

“很黑暗很讽刺,”她说,

>> 难得化学反应

这天,麦克金雷健完身,洗好澡,开着他帅气拉风的日产车横穿市区,去和他首个“黑客”出的对象约会。她的名字叫雪拉,网页设计师,属于A类年轻艺术家类型。他们在一家咖啡厅碰面。“我很紧张,”麦克金雷说,“之前都只是纸上谈兵,现在是真正实践了。”

约会到最后,两人都明显感觉到不来电。第二天,他就约了第二个人,来自B组的魅力博客编辑。他本打算浪漫地散散步,却发现对方最近心情颇为低落,一直在吐苦水。“那个步散得其实挺郁闷的。”他说。

相了一个月的亲,他两个账户上约到的对象数目差不多。他发觉,自己漫无目的地寻找,还是太浪费时间。于是,他把自己适用于A组的账户给删了。这样效率是提高了,但结果

一周人物

本报记者 张文 整理

把自拍玩出朵花儿来



巴登和他24年以来的自拍照。

2013年,英国牛津大学出版社的编辑们挑选出来的年度词语是“自拍”。的确,越来越多的人——国家元首、演艺明星、普通民众,高举着智能手机摆出pose咧开嘴。被社交网站上妹子们的自拍刷屏弄得审美疲劳了?来,给你介绍两位自拍大神,他们把自拍这事儿玩出了新花样。

美国波士顿大学摄影的教授卡尔·巴登,制作了一段长约两分钟的视频。在这个视频里,巴登相似角度、相同姿势的自拍照一一闪过。这些照片是他在1987年2月23日到2011年11月3日期间所拍摄的自拍照的一部分。巴登说,在这24年的时间里,他一共自拍了9800次,只有一天因故没拍。

这些数量庞大的自拍照,清晰记录了这些年巴登样子的变化。巴登说,他之所以坚持24年如一日地自拍,是想了解生命的不可持久性如何影响人们的日常生活,以及老化会如何改变人们的外貌。

自拍是对“自恋”的绝佳解释,而巴登很好地地把这种自恋强迫症演绎成了一场行为艺术。巴登说,直到离开这个世界那天,他都会将自拍进行下去。

生活中这点小小的“疯狂”,别有一番趣味。而美国著名脱口秀喜剧演员艾伦·德杰尼勒斯,则把自拍玩成了一场全民狂欢。3月2日,在奥斯卡颁奖典礼上,身为主持人的艾伦招呼了一众大腕明星,拍下一张包括凯文·史派西、梅丽尔·斯特里等数位奥斯卡影帝影后在内的自拍照。这张照片被艾伦上传到推特上后,迅速引起疯狂转发。35分钟后,它就以81万次的转发量成为推特史上转发量最大的照片,还把推特搞宕机了。

此前,推特的转发纪录是由美国总统奥巴马创下的。在2012年连任时,奥巴马和妻子米歇尔深情相拥。这张照片获得了78万次的转发量。而艾伦和众明星的合照,截止到3月4日上午,已经被转发了超过300万次。

如果你以为此举仅仅是明星们一时兴起的自娱自乐,那就太小看名利场的价值了。媒体爆料,这张自拍照实际上是三星公司的广告策略。在颁奖典礼开始前几天,艾伦提出想要在节目中玩自拍的点子,而三星是本届奥斯卡颁奖典礼的赞助商。于是,她被建议用三星手机来自拍。彩排期间,三星高管还亲自指导艾伦使用手机的方法。

事实证明,这是一次极好的植入广告和病毒营销。在电视直播时,观众很明显能看到自拍所用的白色手机来自三星。而在推特上,三星一度在一分钟内被提及900次。

这就是网友在被明星们的娱乐精神取悦之后,给广告主三星做出的最好回报——虽然上不了头条,也能上个话题榜。



艾伦(前排右二)和明星们玩自拍。