



## 大数据成BAT“新战场”

**Baidu 百度**

牢牢把住用户  
上网入口的百度加  
速储备大数据领  
域人才。

**阿里巴巴**  
alibaba.com.cn

马云的思路  
很明确,阿里未来  
要做的是数据公  
司。

**腾讯**

腾讯启动与人  
大社会管理大数据  
中心首批博士后共  
同培养项目。

百度用大数据成功预测德国夺冠 马云直言阿里最值钱的是数据

# 大数据颠覆未来?

“我们每个人乘飞机时,都是自己选择航线,这是人的智慧,但当这反映到具体的一些航程中来,就会有大量的数据记录下来。我们从这些原始的航程记录中,就可获取一些航程的最优设计方案。这就是大数据的方法。”中国人民大学信息学院院长杜小勇这样解释什么是“大数据”。

银监会日前宣布,正式批准三家民营银行的筹建申请,其最大亮点就是互联网公司腾讯的入选。而互联网公司做银行当然要有自己的看家本事,大数据或被作为其最强有力的武器。同时,今年“大数据”还首次被写入政府工作报告:要设立新型产业创业创新平台,在大数据等方面赶超先进,引领未来产业发展。

其实,大数据的应用早已无处不在,利用网购数据授信买车;通过博彩和球队比赛数据预测世界杯;依据观众数据来打造的互联网电影;就连今年全国高考语文,多省高考作文题,被“百度大数据”预测命中……你能想到的和想不到的都在被数据包围,未来大数据对人们生活的颠覆性影响将堪比互联网带来的猛烈冲击。

### ■释疑

#### 大数据助互联网公司做银行

银监会上周五宣布,已正式批准三家民营银行的筹建申请,其最大亮点就是互联网公司腾讯的入选,而首批落选的阿里也在摩拳擦掌。很多人可能比较疑惑,互联网公司凭什么做银行?

记者了解到,深圳前海微众银行将结合互联网,提供高效和差异化的金融服务。以“普惠金融”为概念,主要面对个人或企

业的小微贷款需求。未来会利用互联网平台开展业务,依托平台,与其他金融机构合作开展业务。

生硬而有繁琐的“表述”背后,其实就一句话,要用“互联网平台”满足“小微贷款需求”。这不禁让人联想到此前被央行叫停的腾讯和阿里“网络信用卡”,该业务是一种运用大数据技术

机选并即时调整授信额度的网络数字信用卡,而发放的对象就是从传统银行拿不到贷款的“屌丝”。这让我们看到了“互联网”银行产品的“雏形”。

而上周阿里宣布,和中行、招行、建行等7家银行深度合作,也不失为“互联网银行”服务小微企业的一次有益探索,其为中小企业提供基于网商信用的

无抵押贷款,最高授信额度1000万元。电商起家的阿里,将平台拥有的数万家企业交易数据开放给银行,降低银行贷款风控成本的同时,亦为苦于无法自证信用的中小企业,提供了第三方担保。

这次再问互联网公司凭什么做银行?就再清楚不过了,没错,就是大数据。

### ■案例

#### 百度大数据预测世界杯淘汰赛结果全对

那么,什么是大数据?枯燥的名字解释可能会让很多“技术盲”望而却步,中国人民大学信息学院院长杜小勇在腾讯互联网与社会研究院主办的“大数据连接的未来”高峰论坛上给出了一个非常鲜活的解释,“我们每个人乘飞机时,都是自己选择航线,这是人的智慧,但当这反映到具体的一些航程中来,就会有大量的数据记录下来。我们从这些原始的航程记录中,就可获

取一些航程的最优设计方案。这就是大数据的方法。”

大数据又如何连接未来?举个例子,百度在世界杯期间准确预测德国夺冠,也是唯一一家通过大数据准确地预测了比赛结果。百度是如何通过大数据进行预测的呢?

百度大数据研究院特别派遣了资深数据科学家团队,利用百度大数据全面搜索过去5年内全世界987支球队的3.7万场

比赛数据,并与国内著名彩票网站乐彩网、欧洲必发指数独家数据供应商 Spdex 等公司建立数据战略合作伙伴关系,将博彩市场数据融入预测模型中,构建了本次“世界杯预测”产品的足球赛事预测模型。

该模型共涉及19972名球员和1.12亿条相关数据,所参考的数据包括百度搜索数据、球队基础数据、球员基础数据、赔率市场数据等,所分析的球队不仅

包括207支国家队,还囊括了欧洲、南美、亚洲等联赛俱乐部及低级别球队信息。在国家队胜负场预测上,准确率接近80%。

百度使用这个模型对2006年和2010年世界杯的淘汰赛进行了验证,准确度接近75%。从此次巴西世界杯的比赛结果来看,这套模型的准确率要高于微软、谷歌和高盛,其小组赛阶段的预测成功率为58.33%,淘汰赛阶段全部预测准确。

### ■竞争

#### BAT 加紧布局大数据

“大数据连接未来”,这让互联网行业看到了新的机遇,BAT(百度、阿里、腾讯)三大巨头正在加紧布局,而未来的生死存亡或要在大数据一战线上见分晓。

很多人都比较疑惑,阿里到处撒金,动辄上十亿级的大手笔投资让人惊叹,更重要的是“不务正业”,一会儿买足球,一会儿又搞文化,一个电商企业如此扩张让外界觉得阿里创始人马云退休后“很不着调”。事实上,那不过是“烟雾弹”,阿里的内部中层在采访中曾告诉记者:“马云的思路很明确,阿里未来要做

的是数据公司。”

据了解,马云很早就意识到大数据的价值,曾经说过阿里巴巴集团价值最高的不是淘宝,不是天猫,也不是支付宝,而是阿里平台上所产生的大量数据。阿里巴巴、天猫、淘宝、支付宝、阿里金融产生的数据构成了个人、企业、商品和金融之间的完全数据链,这些数据能够让阿里巴巴提供更精准、高效的服务,而这些数据同时也能够给阿里巴巴带来不菲的收入。据悉,很多天猫和淘宝的卖家都购买了数据魔方、量子衡道等基于大数据的增值服务。

如果说阿里仍争分夺秒地抢“数据”,那么作为搜索行业老大,牢牢把住用户上网入口的百度则加速储备“人才”。在大数据这一前沿领域,全球的高精尖研究人才数据极为有限。李彦宏掌舵的百度也正在加大对大数据的布局,其推动成立百度深度学习研究院,并亲任院长,吸引了一大批世界级科技精英的加盟,比如前Facebook资深科学家徐伟、美国新泽西州立大学统计系教授张潼等,最近还挖来了“谷歌大脑之父”吴恩达。

拥有国内丰富的海量数据

的腾讯,也在人才培养和引进上加大力度,上周五成立腾讯互联网与社会研究院,启动与人大社会管理大数据中心首批博士后共同培养项目,双方将结合企业数据优势和高校研究力量,共同培养产学研相结合的高端人才,并聘请牛津大学互联网研究院主任 Luciano Floridi 教授等为名誉顾问。

可以预见的是,随着以BAT三大巨头为首的互联网公司在大数据方面的布局加快,未来大数据的应用场景将更加丰富,用户也期待尽快看到由大数据连接的“未来”。

### ■观察

#### 别让用户隐私“裸奔”

随着大数据应用越来越广泛,如何保障用户数据的安全隐私,成为大数据应用的最大挑战。法律界人士师晓丹在其题为“大数据时代的法律应对”一文中甚至直言,大数据时代人人“被裸奔”。

在大数据时代,每个人都是数据的贡献者,当你浏览网页、网购、扫描二维码、微博、微信以及安装手机APP时,你的个人信息、消费习惯、偏好,甚至你的社交圈子,就已经被大数据分析工具捕获。大数据分析工具使智能、高效地处理庞大数据成为现实,但同时它也能嗅探到你的所有信息,我们的城市在变得越来越智慧的同时,似乎也越来越危险了。

腾讯公司云平台部总经理陈磊日前在“大数据连接的未来”高峰论坛上也表示,腾讯曾经对90个要求用户用信用卡或银行卡支付的电商网站做过安全扫描分析检查,发现超过60个都或多或少存在安全问题,其中20多个的问题非常严重,存在包括盗取用户的身份、恶意去替用户消费等行为。

“今天我住酒店的时候,如果酒店的工作人员让我把信用卡留下,我是非常焦虑的,因为今天我们面临的互联网产品里的安全问题非常多。”陈磊如此表达自己的担心。

陈磊强调,“要做好大数据的服务,我们首先要解决的就是信息安全的问题。特别是对腾讯而言,首先发生的挑战就是安全的挑战。”

师晓丹也提出,大数据时代的来临,使人类历史仿佛突然进入了一个崭新的世界。在大数据面前,传统的保护手段显得苍白无力。传统的保护个人信息的法律手段“告知与许可”基本失效,因为大数据的价值不单纯来源于数据的基本用途,更多的源于数据的二次利用,很多数据在收集时并无意用作其他用途,而最终却产生了很多创新性的用途,这些都是无法事先告知的,也就没有所谓的事先同意了。传统的保护个人信息的技术手段“匿名化”基本失灵。

师晓丹建议,在传统手段无力的情况下,大数据时代个人信息保护需要新的治理思维,“告知与许可”的基本法律手段依然可发挥作用,但只适用于数据收集阶段,如浏览网页时普遍存在的cookie。此时应由用户选择是否接受数据的收集与分析以获得更好的用户体验,如果用户选择“否”,其任何数据不得被捕获。在数据的“二次利用”阶段,可考虑设置数据使用时效机制、大数据使用者惩罚机制、新技术强制适用机制。将数据使用限制在一定时效范围内,意味着大数据收集者不再可以永久地保留和利用数据。大数据的价值决定了个人信息保护不可能单纯依赖企业自律,大数据使用者的责任只有在强制力规范下才能确保履行到位,只有严格的罚则才能防止企业为了利润罔顾大众安全。

新的时代,法律始终要有技术支撑,“匿名化”技术可更新为“差别隐私”技术。企业真正需要的是有价值的信息,而不是窥探个人隐私。“差别隐私”技术通过故意的数据模糊处理,可以实现大数据的查询只显示近似结果,而不是精确结果,挖出特定个人与特定数据点的联系将难以实现且耗费巨大,强制推行该技术,在现阶段不失为良策。(据北京青年报)