

← (上接10版)

管是其中的哪一条都不适合放在大学排名模型中。

《美国新闻》的编辑只是挑选了一些和评估目标看似相关的变量。他们研究了高中生的SAT(学业能力倾向测验)成绩、学校的学生教师比和录取率。他们统计了顺利进入大二的学生占总数的百分比和顺利毕业的学生占总数的百分比。他们计算仍在世的已毕业校友为母校捐款的人数占总数的百分比,依据是他们给母校捐款很可能表明他们喜欢母校的教育。排名结果中占3/4权重的分数都来自一种算法,这种算法就包含以上那些变量;另外占1/4权重的分数则来自全美各地的大学校长的主观评价。

### 当这一排名发展成全美标准时,恶性循环出现了

《美国新闻》第一次依据数据确定的大学排名于1988年出炉,当这一排名发展成全美标准时,恶性循环出现了。排名会自行巩固,如果一所大学在《美国新闻》所发布的排名中名次靠后,它的声誉就会下降,生源情况就会恶化。优秀的学生会避开这所大学,优秀的教授也一样。已毕业的校友将减少捐款。由此,这所学校的排名就会继续下跌。简单来说,排名决定了大学的命运。

以前,大学有各种方法可以宣扬学校教育的成功,许多是靠传闻逸事。例如,某些教授得到了众多学生的一致好评;一些学生在毕业后走上了杰出的职业之路,成为外交官或者成功的企业家;还有一些学生出版了一流的小说。这些正面事迹经由口口相传广为人知,学校的声誉也由此提升。但是,麦卡利斯特学院就比里德学院好吗?或者艾奥瓦大学就比伊利诺伊大学好吗?这很难说。不同的大学就像不同类型的音乐或者不同的饮食习惯,对于某所大学的评价众说纷纭,好坏两方面都可以列出充分的理由。而现在,大学的整体声誉被一组数字蒙上了阴影。

如果你站在大学校长的角度思考这件事情,你会发现大学排名其实是很糟糕的。毫无疑问,绝大多数校长珍惜自己的大学经历,因为从某种程度上来说,正是大学经历激励他们攀登学术阶梯,成为一所

大学的校长。但是现在,这些正处在事业高峰期的校长需要投入巨大的精力提高与学校教育优秀度有关的15个考核项的分数,而这15个考核项是由一个二流杂志社的一组编辑定义的。他们就好像又回到了学生时代,每天都在祈求老师给高分。如果《美国新闻》发表的大学排名只在小范围内流行,倒也不会造成什么麻烦。但是,这个排名的影响力发展迅速,很快成为一个全美标准。教育界一下子紧张起来,迅速给大学校长和学生都设定了严格的任务清单。《美国新闻》的大学排名模型规模巨大,造成了大范围的损害,导致了几乎是无尽的恶性循环。

一些大学的校领导想尽一切办法提高排名。贝勒大学设立奖金激励大一新生再次参加SAT考试,希望再考一次能提高他们的成绩以及贝勒大学的排名。有些名校,包括宾夕法尼亚州的巴克内尔大学和加利福尼亚州的麦肯纳学院,则给《美国新闻》反馈了假数据,夸大了其学校新生的入学分数。2011年,位于纽约的爱纳学院承认其学校教师几乎捏造了所有的数据:考试成绩、录取率和毕业率、新生保留率、师生比和校友捐赠额。但谎言起效了,至少在一段时间之内。据《美国新闻》估算,假数据将爱纳学院从东北地区大学排名第50名提升至第30名。

更多的校领导则试图寻找一种更常规的方式来提高他们的学校排名——努力提升学校在影响最终分数的每一个变量上的表现。他们可能会认为这是效率最高的资源利用方式。毕竟,只要他们努力去迎合《美国新闻》的模型,得到更高的排名,他们就能筹集到更多的资金,吸引来更优秀的学生和教授,然后进一步提高排名。除此之外,还有别的选择吗?

### 大学为了排名展开“军备竞赛”,但排名模型把成本排除在算法外

罗伯特·莫尔斯从1976年起就在《美国新闻》杂志社工作,他是这个大学排名项目的组织者,他在采访中称进行大学排名有利于推动大学制定更有意义的目标。如果他们能因此致力于提高毕业率或者把学生分成更小的班级上课以提高教学效果,那就说明

排名是件好事情。他承认杂志社拿不到与大学教育优秀度最相关的数据,即每个学校学生的学习内容。但是,基于替代变量建立的《美国新闻》大学排名模型也足够反映问题了。

但是,当你基于替代变量建立模型时,钻模型的漏洞会变得容易很多。这是因为替代变量比起它所代表的复杂事实更容易操控。举个例子,假设有一个网站要聘用一个社交媒体专家,人事经理决定选定一个变量——重点考虑推特粉丝数排名靠前的应聘者。推特粉丝数是社交媒体参与度的标志之一,没错吧?但是想象一下,应聘者很快就会无所不用其极地增加推特粉丝。有人会花费19.95美元直接“购买”大量由机器操控的粉丝。因为人们钻了招聘系统的漏洞,替代变量失去了效力。

在《美国新闻》大学排名事件中,从高中毕业生到大学校友再到公司的人力资源部,人们很快接受了该排名是大学教育质量的一个体现。因此,各个大学只能选择配合,他们不得不努力提高排名所涉及的每一个考核项的分数。其实,许多学校最焦虑的是那不能控制的占排名结果1/4权重的因素,即声誉分数,来自各个大学、学院的校领导给出的问卷调查反馈。

2008年,沃思堡市的得克萨斯基督教大学(TCU)排名猛降。三年前,该校的名次是97,之后三年名次递降为105、108和113。该校的校友和支持者为此感到很愤怒,校长维克多·博西尼也因此陷入尴尬境地。他坚称得克萨斯基督教大学在每个指标上的表现都在进步,“我们的新生保留率在提高,我们的筹款等所有方面都在改善”。

博西尼的申辩有两个问题。首先,《美国新闻》排名模型并不是对各个大学进行孤立的判断。即使是各指标分数均有所提升的学校在排名中也会落后于其他分数提升得更快的学校。用学术术语来说,《美国新闻》的评估模型是一种分布模型。这导致了一场学校间的“军备竞赛”。另一个问题是,得克萨斯基督教大学无法控制占1/4权重的声誉分数。招生主任雷蒙德·布朗指出,声誉是模型中权重最大的变量,“这很荒谬,因为它完全是主观的”。新生招生主管威斯·瓦戈纳则指出,为了提高声誉分数,各大学都在纷纷为自己打广告。

尽管如此,得克萨斯基督

教大学仍然决定着手提升那可控制的占3/4权重的分数。这所大学发起了一个2.5亿美元的筹款活动。到2009年,学校已募集到4.34亿美元,远远超过目标额度。由于筹款额是排名的指标之一,仅此一项成绩就提升了排名。得克萨斯基督教大学花费了其中的大部分资金用于校园设施改善,其中1亿美元用于兴建中央商场和学生活动中心,努力让得克萨斯基督教大学的校园看上去更具吸引力。这些做法本身没有什么不对,但其初衷是迎合《美国新闻》的排名模型。

也许更重要的是,得克萨斯基督教大学兴建了一个其时最高水准的体育训练场馆,并将大量的资源投入到足球项目之中。在接下来的几年里,得克萨斯基督教大学的角蛙足球队成为国家强队。2010年,他们在玫瑰杯足球赛中打败了老牌强队威斯康星队,取得了全美总冠军。

这次胜利为得克萨斯基督教大学带来了所谓的“弗洛特尔效应”(the Flutie effect)。1984年,在一场极为精彩的大学橄榄球比赛上,波士顿大学队的四分卫道格·弗洛特尔在最后一秒完成了一个扭转败局的超长距传球,打败了迈阿密大学队。弗洛特尔由此成为一个传奇。这场比赛结束后的两年内,波士顿大学的大学申请率上涨了30%。乔治城大学也曾拥有带来过同样的宣传效果的传奇。该校由帕特里克·尤因带领的篮球队三次打进全美锦标赛。看来,赢得体育比赛是吸引学生申请某所大学的关键因素。当大批体校的高三学生在电视上观看大学体育比赛时,球队实力强劲的学校对他们形成了极大的吸引力。这些学生会为自己是该校的学生、身着写着该校校名的队服而感到骄傲。这些大学接到的入学申请因此暴涨。随着更多的学生申请入学,招生处就可以提高入学门槛,以提高大学新生的SAT平均分,而这有助于提高大学排名。另外,学校拒绝的申请学生越多,其录取率就越低,对排名就越有利。

得克萨斯基督教大学的策略奏效了。到2013年,该大学已成为得克萨斯州学生选择度排名第二的大学,排在第一的是著名的休斯敦莱斯大学。这一年,得克萨斯基督教大学的新生高考和入学考试平均成绩均达到史上最高水平,其在全美的排名也因此大幅上升。2015年,该校全美排名76,也就是说,仅用

了7年时间,该校就上升了37个名次。

现在,我们终于发现《美国新闻》大学排名模型最大的缺陷是什么了。我们不能说《美国新闻》的编辑为评判“教育优秀度”选择的替代变量是无效的,但他们犯下的更大的错误来自他们没有纳入考虑的变量:学杂费、学生助学金。这些变量被该排名模型遗漏了。

这引出了我们将会频繁讨论的一个关键问题:建模者的目标是什么?在大学排名这个案例里,你需要站在1988年《美国新闻》编辑们的角度来考虑。当他们在建立第一个统计模型的时候,他们怎么知道这一模型是否有效?首先,如果模型能反映一些已有定论的大学排名,这就表明其有一定的可信度。比如,如果哈佛大学、斯坦福大学、普林斯顿大学和耶鲁大学在大学排名模型中位居前列,这就在一定程度上证实了《美国新闻》编辑设计出的大学排名模型是有效的。而要建立这样一个模型,他们只需要去研究那些一流高校,思考这些大学的特殊之处是什么就可以了。优秀大学的共同点是什么?这些学校与其隔壁镇的保底学校差距何在?他们发现:优秀大学的新生SAT成绩都很高,而且绝大部分都能顺利毕业;已毕业的校友都很有钱,会不断给学校捐款;等等。就这样,《美国新闻》的大学排名项目组通过分析名牌大学的优势,建立了一个测量教育优秀度的评估指标体系。

现在,如果该项目组将教育成本纳入算法,则其模型输出也许会发生奇怪的变化——学费便宜的大学很可能因此闯入优秀大学之列,而这一结果将遭到广泛的质疑。由于公众可能会把《美国新闻》最终公布的大学排名看得特别重要,因此采取保守、常规的算法,保证一流大学位于排名输出结果的前列,是一种更安全的做法。当然,高成本也许正是优秀的代价,这也不是没道理。

《美国新闻》的排名模型把成本排除在算法外,这就好像是给大学校长们递了一本镀金支票簿。后者要遵循的唯一指令,就是最大限度地提高15个考核指标的评分,而降低成本则不在其列。事实上,提高学费反而能让他们有更多的资源用于提升考核项目的表现。

从此,学费一路飙升。从

(下转12版) →