

· 主编心语 ·

博士生的成功之道(36)

——数据怎样挖掘

先说一个经典的人物和他的故事。

话说红军时期有一位带兵的人,身上总是带个小本子,上面记载着每次战斗的缴获、歼敌数量。每次打完仗,就往上面添加数字。而且,这个习惯并没有因为时间的消逝而消失。1948年辽沈战役开始之后,在东北野战军前线指挥所里面,每天深夜都要进行例行的“每日军情汇报”,由值班参谋读出下属各个纵队、师、团用电台报告的当日战况和缴获情况。这样的要求当然出自同一个人。

那几乎是一件重复着千篇一律的枯燥无味数据的事:每支部队歼敌多少、俘虏多少;缴获的火炮、车辆多少,枪支、物资多少……这位司令员的要求很细,俘虏要分清军官和士兵,缴获的枪支,要统计出机枪、长枪、短枪;击毁和缴获尚能使用的汽车,也要分出大小和类别。经过一天紧张的战斗指挥工作,人们都非常疲劳。整个作战室里面估计只有定下这个规矩的司令员本人,还有那个读电报的“倒霉”参谋在用心留意。

1948年10月14日,东北野战军以迅雷不及掩耳之势,仅用了30小时就攻克了对手原以为可以长期坚守的锦州,并全歼了守敌十余万之后,不顾疲劳,挥师北上与从沈阳出援的敌精锐廖耀湘兵团20余万官兵在辽西相遇,一时间形成了混战。战局瞬息万变,谁胜谁负灾难预料。

在如此紧急的大战中,这位司令员无论有多忙,仍然坚持每晚必作的“功课”。一天深夜,值班参谋正在读着下面某师上报的其下属部队的战报,说他们下面的部队碰到一个不大的遭遇战,歼敌部分、其余逃走。与其他之前所读的战报看上去并无明显异样,值班参谋就这样读着读着,司令员突然叫了一声“停!”他的眼里闪出了光芒,问:“刚才念的在胡家窝棚那个战斗的缴获,你们听到了吗?”

大家带着睡意的脸上出现了茫然,因为如此战斗每天都有几十起,不都是

差不多一模一样的枯燥数字吗?司令员扫视一周,见无人回答,便接连问了3句:“为什么那里缴获的短枪与长枪的比例比其他战斗略高?”“为什么那里缴获和击毁的小车与大车的比例比其他战斗略高?”“为什么在那里俘虏和击毙的军官与士兵的比例比其他战斗略高?”

人们还没有来得及思索,等不及的司令员大步走向挂满军用地图的墙壁,指着地图上的那个点说:“我猜想,不,我断定!敌人的指挥所就在这里!”随后司令员口授命令,追击从胡家窝棚逃走的那部分敌人,并坚决把他们打掉。各部队要采取分割包围的办法,把失去指挥中枢后变得混乱的几十万敌军切成小块,逐一歼灭。司令员的命令随着无线电波发向了参战的各部队。

而此时的廖耀湘,正庆幸自己刚刚从偶然的一场遭遇战中安全脱身并与自己的另外一支部队汇合。他来不及休息就急于指令各部队尽快调整部署,为下一阶段作准备。可是好景不长,紧随而来的解放军迅速把他的新指挥部团团围住,拼命攻击,漫山遍野的解放军战士中,不断有人喊着:“矮胖子,白净脸;金丝眼镜湖南腔,不要放走廖耀湘!”把对方指挥官的细节特征琢磨到如此细微,并变成如此威力巨大的顺口溜,穿着满身油渍伙夫服装的廖耀湘只好从俘虏群中站出来,无奈地说:“我是廖耀湘”,沮丧地举手投降。廖耀湘精心隐蔽的精悍野战司令部就这么快地被发现、打掉了。(见于“清华领导力”所发微信故事,2014-01-14)

值得我们记住的是,取得这场重要战役胜利的其中一个关键因素,居然出于获胜方的统帅夜半时分,对一份普通遭遇战之后的战报的数据分析,来源于他“从红军带兵时起,身上有个小本子,上面记载着每次战斗的缴获、歼敌数量”的军事素养。这位司令员,就是林彪。

对于廖耀湘来说,这是一次不可思议

的失败。而今天,这可以说是一个十分经典的数据挖掘案例。我们当然还记得,在抗日战争中,日本侵华部队利用侦测战斗中密集的无线电通讯,判断出八路军指挥中枢。在严酷的突围战中,八路军牺牲了一位杰出的指挥员——左权将军。

在今天的科研中,数据挖掘值得引起高度重视。数据挖掘的能力,是一位优秀的科研工作者的最重要技能之一,博士生当然不能例外,也许更为紧迫。

首先要关注数据源。数据源不仅仅是指各种各样的电子数据库,尽管这些库规模巨大,检索速度难以置信的快。除了你是一位数据库管理员,还没有单纯靠数据库而成功的情况。最有效的数据库,其实是你每天用的实验数据记录本。今天,科研的核心任务,是获得各种数据。保证你的数据记录准确、可靠,你可能从中发现意想不到的结论和价值。文献调研也不是仅仅为了给学位委员会交一份综述报告,这是你数据挖掘能力的训练平台,这么说来,你手上一篇一篇的文献,就是最接近你科研的数据源。数据库不会属于你的学位论文,倒是你从中发现了哪些新结果、新结论、新数字、新作者、新方法……才是最重要的。关注专著,从而关注专著的“家园”——各类专业图书馆,也是值得你关注的数据库之一。你认真阅读专著,其实就是一次和大师的对话,因为大师们就是专著的作者。目前,互联网帮不了这个忙。能力,特别是数据挖掘的能力,是不会从互联网“跑”到你身上的,尽管网上可以提供千千万万有关知识的答案,让知识成为力量,还得靠你身上锻炼而成的一种能力——理解和分析能力。从本文开始的故事,你可以看到,数据挖掘能力实际上是一种习惯使然的能力,是一种长期积累的能力。你在科研中有什么习惯?是好习惯吗?你在日常生活中有什么积累?是像“战场胜利品记录”那样的积累吗?这些问题,值得你想一想。

文/冯长根