

利用数据解决业务问题，你需要的四项修炼

在展开介绍四项修炼之前，先谈谈修炼的背景，业务与数据之间的鸿沟问题。其实这也是我写数源思维这本书的初衷，或者说缘由。因为在这些年的工作中发现大家在意识上都已经知道数据很重要，但在行动上将数据应用到自己的业务中去还存在不少障碍。这个障碍具体表现在哪些地方呢？

之前的文章简介中写了一个数据段子。其实我们的生活、工作中这样的例子是不少的。归根到底，很多人对数据不敏感。那什么叫数据敏感？要说清这个问题，就要先明白数据是什么？

我们大脑思考需要载体的，重要的载体包括影像、声音和自然语言。语言本身不是数据，但被以文字方式记录下来后就成为了数据；影像、声音也一样，被录影、录音器材记录下来后形成了数据。数据，简单说就是人造机制/机器对客观世界观察的最原始成果。当代社会的数据形式和体量急剧膨胀，这个时代背景中，识读数据的能力差距也将成为认知差距的来源。识读数据的能力也就是数据分析的能力。数据分析处理能力的大小，就是代表了数据敏感性的大小。那怎么锻炼这种能力呢？那需要先来看看数据分析是干嘛的？

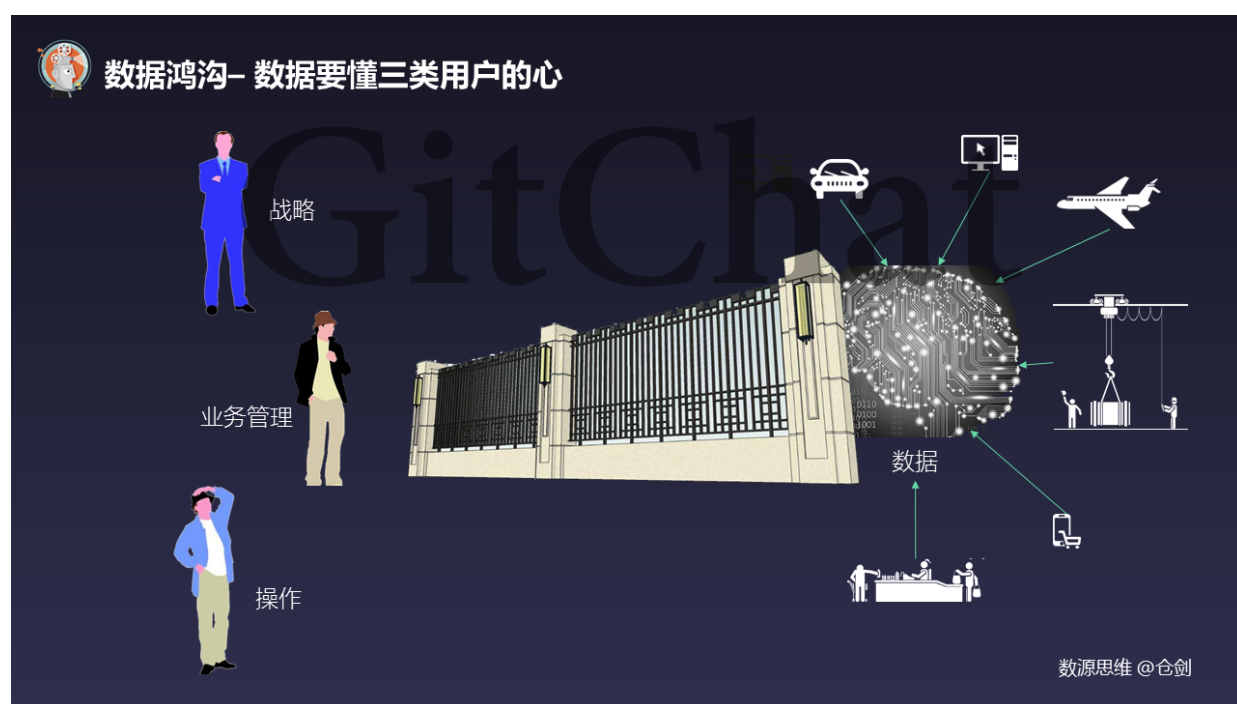
现在我们看到越来越多的岗位，尤其是中层以上的岗位的职能要求里都会有一条对数据敏感，具备数据分析能力。甚至有些职位要求更过分，还要求数据挖掘能力。数据分析的直观含义就是通过数据去认知、推理和处理事物。



广义的数据分析分两个层面，定量业务分析和统计分析挖掘，狭义仅指前者。如果业务是商业领域的，那就是商业分析。所以在我们普遍的商业语境里，业务分析和商业分析是通用的。商业分析包括定性分析和定量分析，这两者是互补融合的。但随着数据采集能力的提升，扩展，定量分析的作用越来越大，业务问题的解决因此可以越来越的依靠数据及其方法、技术。

数据，是对各种现象、事件、问题的观察成果，也可以说是一种抽象的表示。从业务运作到数学推导就是一个不断抽象的过程，也就是一个不断去除掉业务性质的过程。业务分析靠近业务端，统计分析挖掘，靠近统计、数学、计算机。数据分析工作的核心内容就是三个抽象。对象抽象，关系抽象和变化抽象。数据分析能力的训练从某种角度说就是抽象及再还原能力的训练。

而最开始说到的业务与数据之间的鸿沟，就是因为抽象能力断层造成的。如上面那张图中所示，在现实中那条绿色趋势线实际是断开的，并不是从业务那头一直通往数学那头的。实践中是分别两拨人在往中间走。从数学、技术那头走的快，声音大，让人感觉数据分析好像就是一项技术活；业务那头走的慢，还有点懵，没有完全接上技术那端的发展。所以出现断层，造成业务那端的人知道数据很重要，但怎么去重用，很难下手。那对于业务人员来说如何破解这个难题呢？



四项修炼

我们业务人员，不管处在哪个层面都必须要思维升级！必须要数源思维的修炼！思维集中代表你的工作模式。如果公司上下左右思维模式不统一，高层制定的战略怎么落实？一线业务怎么让高层理解市场的变化？

数源思维，就是旨在提供一种让你与数据接轨的思维升级方式。



填平数据鸿沟的数源思维，归纳起来就是“问拆解谋”这四步。

问，提出问题

将业务问题分析定义为数据问题，这其中包括解决问题的目标，定义问题的需求者，或问题解决的相关者。这些考虑的因素都是与最后一步“谋”有对应关系的。这个步骤中有四项落实方法，分别是：

提出原始问题

比如有人会提出这种问题“怎么扩大产品销量？”，但这是一个缺胳膊少腿的问题，非常不具体，构不成原始问题。主语、定语、状语都没有。

如果改成“我们部门怎么扩大面向学生用户的产品，在暑假期间线上的销售量？”这样才算原始问题。

分离问题载体

分离出分析载体——既是数据的载体，也是分析任务的载体。比如“注册多长时间后的用户变成老用户？”

新注册的用户，老用户，变化，多长时间都是任务载体，后续分析就是围着这些个载体，再去明确对象，收集数据，进行分析的。

明确决策点

谓语是问题里的动作，从这个载体，我们就能发现“多长时间”变“老”是整个决策的核心，而变老中的老，又是这个核心的关键。这个核心“老”怎么定义呢？

老用户的老，从业务源头看，就等于产品使用习惯，问题的研究就是看用户什么时候形成了习惯了。这样原始问题中的决策点：多长时间变成老用户，就可以转换为：养成习

惯需要多长时间？或者说习惯养成的时点在哪里？

这个新的问题就是我们问题真正可以着手解决的决策问题了。

提取核心因素

新用户习惯养成的时点在哪里？要做出这个决策，其中要解决的核心因素是什么？也就是习惯的核心是什么？这里我们就提取出了习惯的核心因素——行为的稳定性。

到此，我们就把原始问题新老用户的转换时点，明确为了决策点——新用户习惯养成时点，再进一步提取出了核心因素——对新用户行为稳定性的考察。

问题按照结构化的方式呈现：

决策对象：新用户习惯养成的时点。

约束条件：无。

核心因素：新用户行为。

数据决策：统计衡量行为稳定性。

数源思维对这个问题的定义：如何统计衡量新用户行为稳定性来确定其习惯养成的时点。

这个就是数源思维的第一步，修炼的第一步——问，的中心内容。分析原始问题，抓载体，定对象，提取核心因素，结构化问题，最终就落实问题到数据衡量方法上。

拆，拆分问题

第一步中定义好的问题往往因为比较宏观而无法直接着手解决，我们就需要将其拆分为更微观的细节问题，并进而再去寻找细分问题的解答方法。拆分问题要记住两点原则：

- 拆分原则，也就是拆的时候照什么规则或标准来执行——逐层不漏不重。
- 终止条件，也就是拆到哪儿算完——对象方法单一。

“逐层”的意思是学会提问要一步一步进行，每一步只分出同一层的子问题。比如将用户划分为男、女，男、女就是同一个层次的概念。如果把用户划分成了男、少女就错层了。少女是从女性中分出的一个子类，处在男、女的下一层。要保证同层，就是学会提问时一层只能使用一个维度或一个标准。

“不漏不重”从集合的概念来讲，就是拆分出的所有子问题必须是父问题的一个剖分，既彼此互斥，合并起来又是全集。也就是麦肯锡金字塔理论中提到的MECE（Mutually Exclusive Collectively Exhaustive，相互独立，完全穷尽）原则。要做到这个原则，关键在于理清楚问题的所有构成元素。但这一点并没有像图示那样的显而易见。需要仔细分析，而且拆分的结果也没有唯一性。每往下细分一级，你对问题的认识就深一层，细分出来的问题解决的难度就降低一层。

- 拆分修炼的第二个要点：终止条件——对象方法单一

比如“考察哪些新用户行为？”这个问题对象方法是否单一了呢？显然不单一，因为考察的对象“新用户行为”你并不知道有哪些，所以这本身也是一个问题。所以这个问题还可以分为两个子问题：“新用户行为有哪些？”“新用户行为中哪些符合考察对象的标准？”

第一个子问题：对象是新用户行为，方法就是罗列或叫穷举。所以都单一了。不用再往下分就能给答案。

第二个，对象是新用户行为中的一部分，是哪部分不知道。要从新用户行为中选。选的办法呢？不知道。所以这个子问题对象方法不单一，需要继续拆分。

解，解答问题

当问题已经足够明确和细分，能着手开始解决后，我们就进入了“解”的步骤。在解答问题的整个过程中也有四项方法，分别是：

识别难题类型

问题有难易之分，那这个难易主要就是取决于解决问题方法的难易程度。举个最简单的例子，比如你问幼儿园小朋友，你几岁了？一般都能立刻回答你，因为问的人太多了，他们早就记住自己的年纪，直接提取记忆就行。但如果你问年纪大点的人，比如我，几岁了。我反应就会慢一点。因为我不记得自己几岁，要通过出生年月与当前的年月来计算一下。所以同样的问题，用了不同方法，难易程度就不一样了。

回到商业问题，一般情况下我们会碰到五类问题，它们的解决方法是需要仔细去选择或设计的。具体哪五类？我们一个个看下。

第一个定性问题，以心理问题为典型。比如说用户喜不喜欢我们的产品啊，有多喜欢啊？一般喜欢，你只能卖1999元，喜欢的要死，就能卖5000块。但是你怎么知道有多喜欢呢？这个就是定性问题的难点。

第二个全局问题，以空间问题为典型。比如说有段时间我们喜欢讲蓝海战略，你要往竞争小的方向去走，这就是对空间的判断，选择市场空间中较空白的地方。

第三个大型问题，不复杂，但解决起来很繁琐。比如跟踪竞争对手的动向，办法不复杂，但执行起来很耗资源。

第四个难测问题，以不规则问题为典型。比如边界不规则，用户市场边界，哪些算你用户，哪些不是，很难说清楚。还有结构不规则，行业结构，谁是你的竞争对手，谁是合作伙伴，也不一定。

第五个没数问题，探索问题。比如某个新产品或新功能，用户认不认可啊。

最后要提醒一点，这个分类不是一个维度下分出来的，类别之间是有交叉的。也就是具体问题可能同时属于上面五类中的2类或更多。这个分类主要是引导你去找方法的。

解法选择

难题的类型，与解决方法类型有对应关系。也就是你如果能将问题归入到特定的类型，那么就能通过这种对应关系快速寻找解决方法。



当然右边的解法，也是指解法的类型，每一类解法里面包括多种具体的方法。

指标算法设计

算法是解法逻辑的规范化

一谈算法可能让人觉得很高深，好像跟计算机技术，甚至跟数学是密不可分的。

算法，顾名思义就是计算方法，计算广义的理解就是解答问题，所以算法就是解答问题的方法。方法嘛，可以用各种表达形式，可以用数学符号，也可以用计算机语言，更可以用自然语言。所以数学不好，不会编程，不妨碍设计算法。

举个例子，问：你们团队有几个人？回答很可能是这样的：张三、李四……总共10个人吧。

这是算法吗？对，这个就是一种算法，叫穷举，只不过回答的形式不太规范。如果规范描述出来就是：

- 在记忆里把团队的人名，有可能还有影像归成一个团队成员集合；这个集合之前就有，现在就是激活这块记忆。
- 从这个集合里取出一个人，计数器+1。
- 到最后一个人取完，集合空了，就输出计数器结果。

这就是规范化的解答问题的方法的含义。这个问题当然很简单，复杂一些的问题其实也一样。

定指标，什么是指标？

刚才讲的数人头的办法，有一个条件，就是要有唯一识别人的标志。通过这个标志，我们能把不同的人区分开。当然你可以直接用他们的头像相片，但图像识别和处理的效率很低，用名字，会重名。用身份证号，用手机号，或者我们自己给每个人编个号码，如员工号，用户号。这些都叫ID，用户号就是UID，这就是一个指标。就是对分析对象的一个抽象表达。数人头就变成数UID数量就行。

复杂一点的指标，不仅要抽象对象，还包括抽象关系。也就是将对象和关系融合在一起进行抽象，相当于做了一层封装或打包。这样做的目的是通过观察这个指标，我们就能对某些问题有快速的了解。比如说互联网产品最常用的指标DAU，日活跃用户指标。这也是数人头的，有多少个活跃用户嘛。但是到底哪些用户能被放入活跃这个集合是需要业务逻辑的，那就是需要算法的。

谋，谋划实施

第三步解，只是回答了问题，或者说解决了提出的数据问题，但业务问题到此还没有被解决。我们必须将前面的结论结合业务实际形成实施方案，并说服相关者获得资源，进而着手去解决业务问题。

数源思维的谋划阶段就是强调你的方案必须与业务工作的认知和能力相匹配。好比航空发动机，这个被誉为工业皇冠上明珠的产品，其科学原理并不复杂？但能独立做出像样产品那就难了？非不知，是不能也。不是你不懂，是你做不到。数据，是客观的，也是主观的。数据要被用户接受，企业要应用好数据，也要做到知行合一，必须要把数据研究成果落实到执行层面，才能有效说服。

所以不要满足解答，要给方案去解决。这是谋划的第一要义。

谋划实施，很重要的另外一点是说服。不光要会做，还要会说。会说的核心，你要说到用户心里去，满足他们的心理需求。