

大数据概览综述

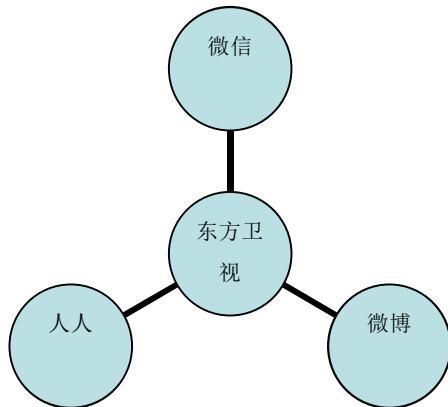
陈海栋 12307110006 材料化学

大数据无疑是这个时代的潮流和热点,经过这一学期程世安教授的课程学习和我校强大助教阵容的精心指导和与同学激烈讨论之后,我从对大数据毫无概念到拨云见日。

大数据按照维基百科的定义就是 4V, Volume Velocity Variety and Verticality, 而并非简简单单的数据容量庞大,数据从新时代而言,最大的区别就是数据的新陈代谢速度加快,原来采集的数据可以经过几天甚至几个月的处理再出结果,可是现在大数据的来领面对的挑战是实时数据反馈和结构调整。根据用户的关注点满意度,商家要调整战略实现发展快速扩张如同病毒滋生。

要掌握大数据的 big picture 就需要先了解互联网世界的基本结构,第一讲中课程的思考题是马航的消息来源,我得出如下结论。

马航失联的案例中,我的第一消息源来自电视新闻,随着东方卫视第一次播出这条重大新闻之后,微博微信人人等社交平台的评论开始进入我的视野。当今世界传统媒体依旧对大部分人产生影响,毕竟每日六点的东方新闻是几代人坚守的习惯和生物钟,总感觉如果吃饭时候没有看到东方卫视的标志就是缺了什么。获得第一手新闻之后就需要深度挖掘的迅速的信息接入,相比之下微博微信有得天独厚的优势。然而传统媒体依然占有强大的媒体资源,其整合能力和特邀嘉宾论坛一直是拳头产品。



正如程教授在课程中描述的小悦悦案例中,南方周末作为传统媒介在微博上依旧成为意见领袖,这绝不是偶然,当今大数据时代,传统媒体纸质媒介依旧占有一席之地,小悦悦案例不是特殊现象。

继而了解了基本结构之后我们需要将目标客户确定才能投放宣传广告,这就是第二讲中的社会化媒体中的人群识别。

人群识别也就是所谓的目标客户挖掘是广告投放有效性的风向标,一个成功的 market campaign 需要做的第一步就是发现定位好自己的目标客户,也就是所谓的客户挖掘。传统行业,纸质媒体中盛行的原则是目标客户群越小集中度越高,广告投放越有针对性,效果越好。相反如果目标客户群模糊,客户是全世界人类,那么这样的广告必然牵涉因素过多过于冗长,不同群体的兴趣点,激发点差异将会葬送这场旷日持久的广告战。

人群识别在网络媒体上有很重大的意义,例如新东方愿意在人人,也就是校内网大大出手不计成本地投入就是因为使用人人的群体大多数都是学生需要英语语言的培训和考察。继而新东方转战微博,也就是看中职场人士需要英语培训来加速自己的升职之路。人群识别在

广告从业者看来就是尚方宝剑，上可斩昏庸策划，下可灭无效设计。人群识别真如同生物体内的靶细胞，有目的有指向地将生物体内的刺激激素识别出来加以处理释放信号。

电子商务更加追求这个点，为了将产品推送到正确的客户手中，许多商家会挖掘你的搜索记录，浏览器 cookies，并且推送出相关产品。其实这样的行为就是一个典型人去识别案例。正是由于这样的物以类聚人以群分的特点，网络的客户群总是以一个群体的对象出现，面对这个群体，研究他们内部的规律和特征是把握住这个客户群的关键所在。这就是第三讲的网络社会的人群聚合规律。

网络人群聚合主要是依靠一个引起共鸣引起注意力的过程，也就是日本电气通模式的 AISAS 中的 attention 也是 SIPS 中的 sympathize 模式。个人认为网络人群聚合的过程核心就是课程所说的人群聚合驱动力（Driving Forces of Group Aggregation），举例来说就是在微博上面我比较受社会热点问题商业时间的影响，由于大多数微博好友关注的都是陌生人所以对微博而言驱动力就是社会热点区问题。然而相比微信，其驱动力就完全不同，对微信而言其驱动力就是人，不同群体之间，群体内部总有一个人依靠为支点，大多数讨论大多数议论都是依靠这个人也就是 opinion leader，所以也印证了我的观点就是一个好的聚合案例必须要有一个好的人群聚合驱动力。当然相比其他的社交软件，比如当时的易信，在推广期和微信的抢用户战打得轰轰烈烈，从送流量到送资金都一一做到，但是其核心问题在于用户基础的薄弱环节。微信当时被腾讯收购，接下来的扩张如火如荼，但是腾讯在 qq 的宣传力度远大于易信的威逼利诱。其核心原因就在于：腾讯 qq 作为一个社交软件也是以人作为关联对象，所以众多的 opinion leader 成功地将微信客户群培养了起来而且为自己所用。如果要总结易信的失败原因就是败在一个电信移动的高贵冷艳不亲民上。其实微信的成功很大一部分可以用马科维兹的理论来解释，人与人的关联是层层推进的，正是 qq 由于其拥有的基础和雄厚客户群造就了他们先发之人完全垄断的局面。

网络人群以各种形式聚合之后，其产生的对话也各不相同，因为不同平台的不同背景导致网络人群的关注点大相径庭，这就是第四讲网络人群“互动、对话”的缘故。

我经常使用微信人人围脖等，他们都有不同的连接属性，我将一一解析。首先我想介绍下人际关系价键理论，人与人的人际关系产生和亲密度与化学成键规律类似，化学当中有离子键共价键配位键，对于人际关系中也是这样，像弱连接一类就是类似离子键，一旦有强有力的吸电子基团，就会完全置之不理，比如微博。平时使用微博也就是娱乐下打发时间，一旦考试或者科研任务降临，微博第一个被冷落，几乎不再接触。人人则有所不同，其连接可以比成是共价键，其互动频率远高于微博，由于都是校友和熟人，情感力量精密，最好的体现就是在考试前夕大家分享复习资料共同学习。互惠互赢，一键分享一一超乎所在。就像共价键，一样，原子之间互相依存，互相帮助，一旦发现电子倾斜转移立即启动自我调节机制，利用共轭体系一传十传百，达到相对强连接的目的。最后讨论微信，其实微信已经成为了断行的替代品，其具有短信一样的强制性特点深受人民喜爱，在对家人亲戚的问候祝福突然从有成本变成了几乎没有成本的流量费用，所有人一夜之间倾巢出动。于是将手机通讯录的所有联系人加了遍，其互动频率，感情力量，亲密程度，互惠交换已经达到顶峰，最重要的是其廉价和强大服务性能的特点。

当群体的聚合群体的特点被一一剖析之后，我们的关注点更加细节到讨论的内容对社会的影响，也就是所谓的口碑，（word of mouth），口碑自古早已有之，并不是网络时代的产物，但是当今大数据时代，我们面对的是网络口碑（iwom）。相比传统口碑有相似之处也有不同之处。

口碑营销的通用模式是鼓动价值回报三大步骤，无论是传统还是网络都不例外。网络口碑再大的特点就是参与度高流动速度快，相比传统口碑，人们需要在物理上地理上接触才可以进行口碑的交换。网络口碑破除了这一障碍，用于英特网的科技含量提供给我们一个足不

出户就可以交流的平台，口碑营销变得极其重要。Iwom 中好的营销项目，核心的炒作热点就像是生物学中的 DNA 模板，在半保留复制，转录逆转录的过程中不断壮大增强。营销的终极秘诀就是制造稀缺和打造富有价值的产品形象，针对这一点网络营销有一些难以控制的因素，由于网络大多是匿名评论，故诋毁产品者不在少数。然而不对这些反对的声音加以控制，后果不堪设想。

所以网络口碑营销必须让产品方放下身段仔细聆听网络的呼声，加以合适合理的导向，最终可以将自己的产品推到风口浪尖，打造成完美理想的品牌。所以网络口碑相比于传统口碑的最大特点就是流动性大，这既是好事也带来担忧，一则享受信息普及面广成本低二来时刻面临来自各方出于各种目的批评指责。

口碑在移动互联网同等重要，随着手机 app 的流行发展，网页购物大有被越俎代庖的势头，此时 app 终端的口碑在互联网入口将影响千千万万之后的消费行为。

例如有关京东手机客户端的用户体验，手机 app 的精髓所在就是简洁，明朗。纷繁复杂而消耗手机大量资源的应用在这个快节奏的社会已经被完全淘汰。试想如果我急于在淘宝购买礼品，结果淘宝应用闪退，手机卡死。这么差的用户体验难以维持客户群。相比而言京东的界面就通俗易懂，上京东的消费者一般分为两类，一是有明确目标商品的只需键入关键字即可，所以搜索栏做得非常醒目。另一类是闲来无事随意逛逛的，一般手机消费大多数是这类，针对这一情况，京东退出的抢购活动既迎合了闲逛用户的打发时间的心态，又将价格压低以满足闲逛用户不太想高消费的目的。另一点就是用户评论，这是又淘宝首先发起的，所以淘宝有丰富的经验，再将网页平台全部移植到手机应用，效果明显很完美，但是随着评论数量的无限膨胀，评论中的大数据需要由一个用户亲和啊界面呈现出来，也就是所谓的统计数据。我们需要一定的客户反馈来决定是否下订单，是否选择这个商家。所以用户反馈对网络消费行为影响深远。

App 为我们提供的口碑排名，产品推荐很大一部分成为我们购物的第一选择，所以网络入口的 app 很大程度上决定了消费行为和消费理念。但是当我们面对纷繁复杂的广告活动或者商品介绍时候我们却很难分辨网络公关和网络营销。

大数据时代企业公关部 (public relation) 和广告部 (advertising) 存在明显差异。公关部的主要职能是协调社会视角对产品的关系，看法，这一点和市场部有些类似，而相比广告部就完全不同，广告部的目的在于将产品推到舆论热点风口浪尖，可以说只要满大街的人都在议论我的产品，那么这个广告案例就是成功完美的。举一个文章出轨的案例，如果文章是经纪公司的一个产品，那么这场出轨大戏中广告部可圈可点成绩斐然，因为他们成功将文章推到了舆论前言，几乎所有人都在议论这件事，文章这个产品的知名度大大提升。然而对于公关这就是莫大的失职，在公众眼中文章的形象轰然倒塌，苦心经营数十载的好男人形象在一夜之间化为乌有，所以公关部肩负的是将产品做成高贵典雅有内涵，公众可望不可即的神形象。然而广告部则是尽量做到亲民，压低身段作为公众的议论焦点出现，提高讨论的知名度。

在微博和微信上的广告公关依然如此，一个是将成品塑造成神形象，一个压低身段作为谈资。每个产品的成功都离不开公关和广告的倾力合作，高贵冷艳的产品没有广告部无人问津，烂俗大众的产品没有公关成为众矢之的。

大数据时代的网络产品最终还是以产品的形式呈现，所以不可避免的需要对产品的品牌价值和意义做出评价。无论产品的各方要素如何具备，其结果，也就是销售业绩直接决定了这场 campaign 的成功与否。

以这门课程为例，程教授和她的团队耗时无数，从 coursera 平台上线之间就已经设计好课程的目标客户群，就是互联网从业者和广告市场从业者以及相关领域的学生。因此在课程开始的前几节中，自然淘汰选择规律已经为教授删选好了这样的一个群体，然后通过线上讨

论，先下 face to face 交流的方式加深讨论深度。可以说这个产品在开始就完全合乎设计。另外口碑效应也同样明显，在课程结束前两个月，注册课程的人数激增，远超预期。不难分析这些后来者一定是受到网络口碑的影响，问道而来。所以这门大数据课程本身作为一个产品，名副其实的也是一个完美的大数据产品案例。

最后感谢这个一个学期的指导和讨论，这样一门平台课程作为复旦第一将会名垂青史历久弥新。