

●张玉峰 部先永 晏创业

动态竞争情报及其采集基础*

摘 要 动态竞争情报是网络时代企业取得成功的关键因素,但实时、分布、异构的业务数据阻碍了人们直接获取它。数据集成是采集动态竞争情报的前提。奠定动态竞争情报采集基础的解决方案有:应用数据仓库方案,数据迁移方案,数据联邦方案。图1。参考文献7。

关键词 竞争情报 动态竞争情报 数据集成 数据采集

分类号 G350

ABSTRACT Dynamic competitive intelligence is a key factor for the success of an enterprise in the networked environment. However, real-time, distributed and heterogeneous business data prevent people from acquiring it directly. The authors think that data integration is a premise for the acquisition of dynamic competitive intelligence, and propose some solutions, such as data warehouse, data transfer and data federation. 1 fig. 7 refs.

KEY WORDS Competitive intelligence. Dynamic competitive intelligence. Data federation. Data warehouse.

CLASS NUMBER G350

如果说10年前最让竞争情报从业者感到恐慌的是互联网引发的“信息爆炸”,那么,今天他们所面对的更为尖锐的问题是如何掌控实时、非连续性变化的业务信息,以便为商业决策提供及时的情报支持。其原因在于,作为公共信息平台的互联网加快了信息新陈代谢的速度,使商业局势在瞬间发生变化,竞争情报的时效性也由此而凸显。于是,动态竞争情报(Dynamic Competitive Intelligence)便受到人们关注。

1 动态竞争情报释义

在本质上,动态竞争情报继承了竞争情报的属性,即在竞争活动中它为管理者提供了可以采取行动的情报(Intelligence)^[1]。但加了“动态”这一限定词后,它被赋予了特殊的含义。张念萍和卢伟从系统建设的角度,认为它是企业营销情报系统中与竞争关联密切的情报部分,其主要作用是对竞争态势进行监控^[2]。冯维扬从竞争情报组织结构的角度,认为它的存在是为了突破情报部门的有形界限和延伸情报部门的功能,以充分利用情报部门的外部资源,而不扩大情报部门的实际规模^[3]。而马克·博斯罗格则从情报生产过程的角度,认为它是克服数据的非可测量性、分布性和动态性等属性的一种手段^[4]。

综合以上几种看法,本文所要讨论的动态竞争情报比较接近于马克·博斯罗格的定义,即认为动态竞

争情报的来源是业务活动中产生的非可测量的、分布的和处于动态变化中的数据,它能为竞争活动中的管理者提供行动指导。

动态竞争情报保持了竞争情报的基本特性,如准确性、时效性和隐秘性等特征,但出于对竞争局势的监控,并为竞争主体提供立竿见影的行动方针,动态竞争情报具有一些独特的属性。

(1)实时性。比一般意义上的时效性更进一步,动态竞争情报对时效的要求达到近乎苛刻的地步,它要求对每一天,甚至每一小时、每一分钟所产生的业务数据进行分析,得出关于竞争局势的结论,及时调整竞争主体的行动策略。

(2)广源性。为了保证情报的准确性,动态竞争情报所要监测的不是某一个业务系统、一个职能部门,而是企业的整个业务流程,还要能及时调用历史业务数据以做对比分析用。动态竞争情报的来源非常广泛,包括所有业务系统实时产生的业务数据,以及企业的各种文档资料和历史数据。

(3)精辟性。情报在本质上是对信息进行提炼、浓缩的结果。它不是各种业务数据的堆积,而是对这些数据进行分析和提炼后得出的精辟见解。

2 动态竞争情报的来源

普赖斯科特认为,动态竞争情报重在关注竞争动态

* 本文是国家自然科学基金项目“基于数据挖掘的企业竞争情报智能采集机制研究”(批准号70573082)的研究论文。

的基础上,提炼可以采取行动的“远见”(Foresight)^[5]。他同时界定了竞争动态的范畴——关于竞争对手、供应商、客户、合作伙伴和潜在竞争对手行动和对抗行动的数据。显而易见,不同于普通竞争情报工作的面面俱到,动态竞争情报关注的是实时竞争态势、可以提炼出对短期内的行动提供指导原则的原始信息。

因此,用于生产动态竞争情报的原始信息主要分布在这样几个方面。

(1)实时业务数据(Real-time Business Data)。信息化将企业带入IT系统的应用时代,在一个信息化成熟的企业,同时运行着供应链管理系统、客户关系管理系统、计算机辅助设计系统、产品数据管理系统、办公自动化系统和竞争情报系统等多个不同类型的系统。它们每时每刻都在吞吐关于供应商、客户、自身产品的设计和生、合作伙伴,乃至竞争对手的数据,这些都是生产动态竞争情报最直接、必不可少的原材料。

(2)企业内容(Enterprise Content)。企业在长期的生产经营活动中积累了大量文档、信件和多媒体资料,这些虽然不是动态的,但能为动态竞争情报的产生提供依据。比如,关于行业政策的文件能为解释实时的生产数据所处的水平层次提供依据。

(3)历史业务数据(Historical Business Data)。实时产生的业务数据经过积淀都会成为历史数据,它们被备份在磁带、磁盘等存储工具中。与企业内容一样,历史业务数据也是相对稳定的一类资源,但在对实时业务数据做对比分析时,它们将起到标杆作用,因而也属于动态竞争情报的重要来源。

动态竞争情报的来源信息呈量大、分布和异构3个特点。虽然动态竞争情报的原始信息在绝对数量上不及一般意义上的竞争情报,但对时效性近乎苛刻的要求,使得动态竞争情报工作在单位时间内处理的信息量是一般情报工作所远远不及的。组织结构在地域、应用系统上的分布,致使来自各个业务部门(甚至同一部门)的数据都呈分布状态。异构是分步式完成企业信息化建设不能回避的问题,业务数据产生于不同业务部门的不同应用系统,它们在数据类型、字段名、存储格式等多个方面都存在差异。

3 动态竞争情报的采集障碍

如果将竞争情报循环(Intelligence Cycle)的5个环节(规划定向、信息搜集、信息加工、信息分析及信息扩散)用于动态竞争情报的生产过程,我们会发现其中的3

个环节处于相对次要地位:在规划定向环节,动态竞争情报已有明确的目标——对实时竞争态势进行监控,提供立竿见影的行动指导方针;在信息搜集环节,原始信息主要来自企业的业务信息系统、磁带和磁盘等存储设备,这些都是可以直接获取的信息资源,不会耗费人们太多的精力;在信息扩散环节,浓缩、定量化的情报成果便于通过企业的网络平台加以传播和应用。

这样,惟有信息加工和信息分析这两个环节带有不确定因素。事实上,企业复杂、异构的数据环境确实给这两个环节的工作造成了障碍。

3.1 原始信息的加工障碍

众所周知,分析是情报的灵魂,而作为分析环节的上一个环节,信息加工的质量在很大程度上会对分析的效果产生重要影响。虽然在各个业务系统内部,业务数据具有完整的结构,可以直接用于情报分析,但情报分析绝对不能局限于单个业务数据库,而是要将视野拓展到全局业务数据中去。因此,信息加工的主要任务就是来自不同业务系统和数据平台中的数据融为一体,并使之相互关联。比如,将供应链管理系统中某一原材料的数据与生产管理系统中这同一种原材料的消耗数据关联起来,以便在分析环节得出关于生产过程中原材料供需状况的情报。但企业分布、异构的数据环境给完成这项工作造成巨大障碍。

数据环境的分布体现在物理位置和逻辑结构两个方面。在物理位置方面,像银行、电信、零售等行业的营业网点遍布全国,而每一个营业网点都在实时地产生业务数据,这些数据在物理位置上处于极度分布状态;而在逻辑结构上,任何一种数据都不是孤立地静止在某一个业务系统中,它要在业务系统间流动变化,也就是同一种类型的数据可能会分布在不同的业务系统中,比如采购系统、库存系统和生产系统中都分布着关于同一种原材料的数据。数据的这种分布性所造成的障碍在于,如何将不同区域、不同业务系统中的数据关联起来?因为它们不是存在于一个静止的数据库中。

IT总是朝着开放、多元化的方向发展,这是造成数据环境异构的根本原因。一方面,不同的业务系统会选择不同类型的数据平台(如IBM DB2, Oracle, Sybase, MS SQL Server等),甚至同一业务系统会在不同的时期选择不同的数据平台,而不同的数据平台在存储格式、操作方式上都存在差异,致使数据互不兼容。另一方面,即使不同的业务系统采用同一种类型的数据平台,但在不同的系统中,对同一种事物的描述口径也不尽相同。比如在一个零售行业中,同一商品在采购系统和销

售系统中被描述的字段名、明细程度等都不一致,人们很难将这种商品在两个系统中的数据关联起来。

无论是数据平台的异构性,还是描述口径的差异性,都使得人们无法直接将各种具有内在联系的数据融合成一个整体。

3.2 分析方法的应用障碍

与一般竞争情报工作中使用各种定性的分析模型不同,动态竞争情报面对的是大量异构的量化数据,人工手段难以在短时间内完成对海量的业务数据的分析工作。因此必须采用诸如数据挖掘这样的人工智能手段。

数据挖掘是一个完整的过程(内含一个对原始数据抽取、转换和装载的过程)。该过程从大型数据库中挖掘先前未知的、有效的、可实用的信息,并使用这些信息做出决策和制定可行性方案^[6]。数据挖掘的应用对象是大型数据库,这显然与动态竞争情报关注各个应用系统实时产生的业务数据和企业内部大量的非结构化数据的初衷是不一致的。至少,在采集动态竞争情报的过程中应用数据挖掘技术,必须克服实时业务数据和非结构化数据带来的困扰。

(1)分析实时数据的障碍。实时业务数据的最大特点在于其随机性和分布性,因为与各个网点的业务状况有关,人们无法准确预测挖掘对象将在什么时候来自哪个系统的哪个数据平台,因而人们总是很难及时地将它们装载到一个特定环境中,以待采用数据挖掘的方法对其进行分析。

(2)分析非结构化数据的障碍。在企业的信息资源中,结构化的数据只占很少的一部分,80%以上的是被统称为企业内容资源(Enterprise Content Resource)的非结构化数据。这些数据的结构多样化,包括电子邮件、即时消息、传真、图片、音频、视频等多种类型,人们很难将它们统筹到一个包容一切的数据平台中去,而它们又是动态竞争情报不容忽视的对象,这显然是应用数据挖掘技术的一个两难问题。

4 解决方案分析

对于以量化数据为对象的动态竞争情报,套用IT的量化分析技术能达到事半功倍的效果。通过以上分析可以看到在这种应用之前还存在若干障碍,因此需要一套可行性方法来扫除障碍,以便为动态竞争情报的采集奠定基础。

4.1 数据集成——动态竞争情报的采集基础

换个思路来看问题,如果企业的所有业务工作都

基于同一个数据平台,所有的非结构化信息也能通过一种特殊的处理方法被存储到这个平台中来,那么只用配置一套内嵌数据挖掘功能模块的软件,就能轻松获取动态竞争情报。

从思路来看,奠定动态竞争情报的采集基础就是要构造这样一个超级的数据平台,尽管这种数据平台也许只能存在于逻辑层面。

最初,面对不同业务系统中的数据平台形成的一个个信息孤岛,人们期望通过建立一本企业词典,作为沟通不同数据平台间的信息交流和共享的桥梁。从理论上说这是可行的,但当有企业花3~5年时间、投资逾百万元在这上面,最终却对源源不断出现的新词汇无能为力时,注定了这并非一种值得肯定的方法。特别是企业词典只能用于描述关系型数据库间的字段对应关系,而对非结构化数据却束手无策。

那么,获取这样一个超级数据平台的手段究竟是什么呢?其实,数据挖掘所关心的只是被挖掘对象在逻辑上的整体性和相关性,而并不关心数据所处的物理位置。在奠定动态竞争情报的采集基础时所要做的,就是将用于提炼动态竞争情报的所有原始数据集中起来,这就是数据集成(有时也称数据整合)。

4.2 几种方案

针对用于提炼动态竞争情报的原始信息——实时业务数据、企业内容和历史业务数据,数据集成所要实现的就是沟通不同来源的数据,使存在内在联系的数据相互关联,消除数据流通和共享的障碍,以便能直接应用量化的IT分析方法来提炼情报。

4.2.1 应用数据仓库(Data Warehouse)

数据仓库是面向主题的、整合的、稳定的,并且实时地收集数据以支持管理决策的一种数据决策形式^[7]。数据仓库可用于动态竞争情报的信息集成工作,在于它立足于关系型数据库、并行处理和分布式技术,并将功能扩展到互联网,强化了对非结构化数据的管理,特别是它能实时地收集数据,保证了对动态业务数据的分析利用。

一般而言,数据仓库在大数据量处理的高效性及全局业务数据的整合能力方面具有较大优势,因此通常是大企业开展动态竞争情报工作的主要选择。但对于普通的企业用户而言,数据仓库的高成本投入使用范围受到一定程度的限制。

4.2.2 数据迁移(Data Transfer)方案

像数据仓库这样将所有数据归集到一块,形成一个可以直接用于分析的整体,更符合人们对数据集

成的理解,但当数据仓库的应用受到限制时,是否有可替代措施?数据迁移就是这样一种替代性方案。

在动态竞争情报工作中,数据迁移的工作就是将分布在各种数据源中的数据精华部分萃取出来,并迁移到一个针对动态竞争情报工作的数据平台上来。

与数据仓库相比,数据迁移更注重集成的效率。一方面,在集成前,它要将那些关系太复杂、质量太差,且无法进行联合访问的源数据剔除,只保留精华;另一方面,它注重对应用率比较高的数据的集成,比如对于一些可预测的、并频繁重复的复杂查询,为便于在本地访问数据,便将其一次性加载到目标数据平台中去。

4.2.3 数据联邦(Data Federation)方案

并非将先前的数据结构完全打散,然后重新组合起来才叫信息集成。数据联邦就是这样一种不需大动干戈的柔性集成方式,它在不需要事先对所有数据进行物理迁移的情况下,提供统一的数据资源利用方式。

数据联邦方案主要是应对两种情况:其一,向数据仓库中不断存入新的实时数据的成本很高,对大多数企业来说很不划算;其二,有些数据由于它们的使用、大小或格式不适合于数据仓库或用户查询,因而不能或不需保存在数据仓库中。

的数据。这样,无论数据是来自 DB2、Sybase、Oracle,还是其他,无论数据的形式是文本、XML、Excel,还是即时消息,在竞争情报从业者看来,它们都好像是来自同一数据源。当然,数据集成器在这里做了真正的集成工作,它在平台的底层设置高速缓存,然后做相应的映射和转换工作,从而将数据格式、位置、接入协议及功能之间的差别等隐藏在动态竞争情报生成模块的后面,不为人知。

5 总结

毫无疑问,随着市场竞争的加剧,企业对竞争情报的要求越来越严格,人们期望所获取的情报能直观地反映实时的竞争态势。届时,我们今天所提的动态竞争情报在准确性、实时性上也必将被提高一个层次。高质量的情报来源于高可用性的原始数据,集成分布、异构的数据是提高数据的可用性、保障高质量动态竞争情报正常生产的基础。

尽管业务数据依然呈几何级数增长,数据的类型也日趋多样化,但人们也在不断探索集成数据的技术和方法。有了数据集成做保障,动态竞争情报也将在企业的竞争活动中发挥更大的能动性。

参考文献

- 1 Prescott, J. and G. Bhardway. 1995. A survey of competitive intelligence practices. *Competitive Intelligence Review*, 6. 2: 4-14
- 2 张念萍, 卢伟. 市场动态竞争情报系统及其运用. *经济与社会发展*, 2004(5)
- 3 冯维扬. 面向任务的动态竞争情报组织结构模型分析. *情报学报*, 2002(4)
- 4 Mark Boslough. 2005. A Distributed Dynamic Intelligence Aggregation Method. *International Conference on Intelligence Analysis*, McLean, VA, May 2-6
- 5 John Prescott, J. 1999. The Evolution of Competitive Intelligence. *Proposal Management*, 37-52
- 6 范亚芹, 刘颖, 李兴男. Web 数据挖掘原理及实现. *吉林大学学报(信息科学版)*, 2003(11)
- 7 Inmon W H. *Building the Data Warehouse*. Canada: John Wiley & Sons Inc, 1992

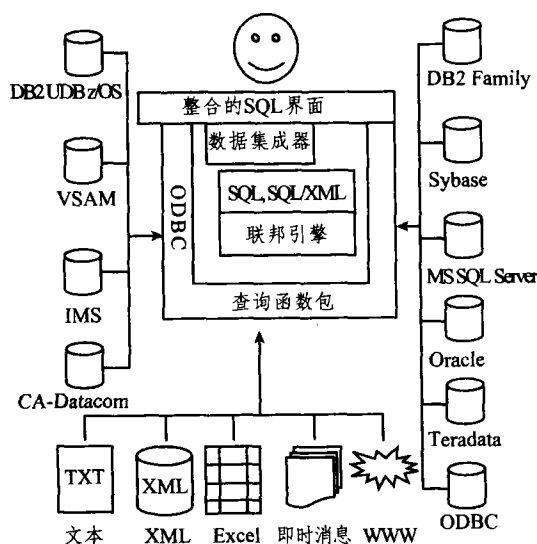


图1 数据联邦的原理示意

数据联邦为竞争情报从业者提供了一个访问、分析各种分布、异构数据的平台,这个平台通过整合的SQL界面作为统一访问口径(图1)。整合的SQL界面对查询需求进行分解,让它们分别作用于不同来源

张玉峰 武汉大学信息管理学院教授,博士生导师。通信地址:武汉。邮编430072。

部先永 武汉大学信息管理学院04级博士生。通信地址同上。

晏创业 通信地址:北京市金宝街58号华丽大厦9层。邮编100005。

(来稿时间:2006-03-28)