

## 点击流：一种研究网上消费者的新范式

李双双<sup>1,2</sup> 陈毅文<sup>1</sup>

(<sup>1</sup>中国科学院心理研究所, 北京 100101) (<sup>2</sup>中国科学院研究生院, 北京 100039)

**摘要** 点击流, 顾名思义, 就是访问者在网上的持续的访问轨迹。点击流研究带动了一种新的研究潮流, 不仅在多个学术领域, 而且在多个实践领域都是大家关注的热点。点击流数据揭示了网上消费者在网上的冲浪行为, 现有的点击流研究大多致力于建立消费者网上购物行为的模型, 从而更深刻地理解网上消费者的行为, 同时, 点击流数据分析也有利于进行数据挖掘, 改善网页设计和提供决策支持。该文将已有的文献分成三种: 网站内点击流研究, 网站间点击流研究, 以及点击流数据的应用研究, 并对点击流研究作为一种新的研究范式的局限性和研究发展趋势作了预测与展望。

**关键词** 点击流, 消费者, 网上购物。

**分类号** B849:C93

正如同网上购物出现的历史并不长却蓬勃发展一样, 伴随着互联网的出现而出现的点击流 (clickstream) 研究带动了一种新的研究潮流。不仅在心理学、管理学、市场营销、计算机科学等学术领域可见其踪影, 而且在网站设计、数据仓库、决策支持系统等实际应用中也是大家关注的热点。

研究消费者网上购物行为, 首先需要回答一个问题: 消费者的网上购物行为是否与传统购物行为不同? 如果答案是否定的, 那么可以直接将前人对消费者传统购物行为的研究结果套用于网上购物行为, 毕竟前人对消费者传统购物行为的研究已经相对完善了。这个问题已经有一些研究者做出了回答: Bucklin和Lattin等人(2002)在前人研究的基础上, 总结出网上购物与传统购物在很多方面存在着差异: 首先, 从购买意图上来看, 消费者在进行网上冲浪时, 一开始其购买意图并不明确, 可能是浏览、搜索或者购买, 而消费者进入传统的商店就揭示了他的购物意图; 其次, 从购买行为的交互性来看, 网上购物消费者可以参与创造决策环境, 而传统购物消费者则是被动的; 再次, 从购物环境来看, 网上消费者的决策环境是个性化的, 而传统购物消费者的决策情境则是相对一般化的, 固定不变的; 最后, 从交易价格上来看, 在网上购物环境中, 营销者可以通过极低的成本进行动态干预, 而传统购物环境下, 交易价格则一般是静态不变的<sup>[1]</sup>。而Dellaert

等人(1998)则认为, 网上购物与传统购物的一个最显著的区别在于交通成本 (transportation costs), 包括实物成本和心理成本两个方面。由于网上购物交通成本相对较低, 网上购物消费者即使没有购买意图也可能进入购物网站, 或者延迟购买<sup>[2]</sup>。

鉴于上述网上购物行为与传统购物行为的差异, 研究网上购物行为是必要的。与传统的通过 UPC (产品上的条码) 数据来研究消费者购物行为相比, 由于使用研究工具的不同, 我们可以更加广泛和深入的研究消费者的网上购物行为: 那就是通过网页服务器的日志文件获得点击流数据 (clickstream data)。

网页服务器日志文件记录了网站上每个访问者的HTTP事务的执行记录, 它包含很多有用的信息, 比如访问时间, 服务器的动作 (事件) 等, 从而为我们带来了新的、大规模的数据资源, 即为点击流数据 (clickstream data)。如今, Web站点经常是企业和管理层需要了解更多的用户行为而不仅仅是管理站点所需要的统计数据。他们想知道是谁正在访问网站, 在站点上的访问者正在干什么, 以及正在进行什么样的交易 (如果有的话)<sup>[3]</sup>。

点击流数据揭示了网上消费者在网上的冲浪行为: 访问者的每一个点击就像他的每一个脚步, 揭示着他在网站内和网站间的选择。其中, 网站内的选择包括他浏览该网站的哪一个网页, 在某个网页停留多长时间, 特别的, 针对购物网站, 还包括他

收稿日期: 2006-09-21

通讯作者: 陈毅文, E-mail: chenyw@psych.ac.cn

是否发生购物行为等等；而网站间的选择则说明访问者选择访问哪个网站。同时，服务器日志文件记录了访问者的IP地址，它可以在一定程度上聚焦在该IP地址对应的特定访问者个体或者群体身上；此外，访问者也可以通过特定的用户名和密码登录网站，从而获得更加个性化的服务，也使研究者能够将特定的访问者和特定的点击流数据联系起来，进行更具体的分析，从而帮助改善网站设计和提供决策支持。

利用点击流数据来研究消费者的网上购物行为，根据研究者关注的重点，可以将已有的研究分成三种：网站内点击流研究，网站间点击流研究，以及点击流数据的应用研究。这些研究的目的，大多在于建立消费者网上购物行为的模型，从而更深刻的理解网上消费者的行为。同时，点击流数据分析也有利于进行数据挖掘，改善网页设计和提供决策支持，所以，本文也将介绍点击流研究在这三个方面的应用。

### 1 网站内点击流研究

网站内点击流研究，根据研究者获取点击流数据所涉及的观察时间，可以分为单一时间段（single session）研究和多时间段（sessions）研究；按照研究者所获取的点击流数据的来源，可以分为实验室模拟点击流研究和购物网站真实点击流数据研究；同时，不同的研究者关注的重点也不同，主要集中在点击流数据所体现的浏览行为、搜索行为和购买转换行为上。

由于现有的一些研究基本上结合了上述分类的一些特征，所以接下来按照点击流数据所涉及的观察时间分类来介绍国内外已有的网站内点击流研究的现状。其中，界定一个用户时间段的默认标准是，当一个访问用户在20分钟内没有新的访问行为时，就认为当前访问者的访问时间已经结束。当然，由于服务器日志文件记录IP数据的模糊性，一个IP并不一定对应着一个特定的消费者和特定的访问，因此，这个分类具有相对性。

#### 1.1 单一时间段点击流研究

在传统的购物环境中，售货员或者促销员能够根据顾客在店内的行为来区分他们，例如，一些顾客表现出非常集中于寻找某种特定的商品，对这样的情况，售货员或者促销员就可以上前帮助他们找到他们所要寻找的东西。而对于其他一些情况，进入商店的顾客仅仅只是在“浏览商店橱窗”（Window

Shopping），有经验的售货员或者促销员就可以辨识出这些顾客，或者忽视他们，让他们继续浏览，或者介入并尝试用合适的方式进行促销，从而使他们产生购买行为。但是在虚拟的网上购物环境中，并没有售货员或者促销员去扮演这样一个角色。因此，通过访问者的点击流数据特征实时的对这些访问者进行分类，就可以使电子商务营销者区分可能的购买者，或者同步在线采取相应的类似售货员或者促销员角色的促销行为，并使营销者设计更多有效的、定制的促销信息成为可能。

单一时间段的点击流研究较少，Moe（2003）认为，在每一个访问时间范围内，访问者被不同的动机所驱动，因而对不同的营销信息表现出不同的反应。使用来自一个已知的网上店铺的页面一页面的点击流数据，将访问者的店铺访问分为购买、搜索和浏览三种类型<sup>[4]</sup>。研究者收集了该网站7个星期的点击流数据，获取了5730个访问者7143次访问时间范围的数据（相对而言，这是一个单一时间段的研究），并将数据随机平均分成两部分，分别用作判断性样本和预测性样本。对判断性样本，使用聚类分析的方法区分出五种类型的网上消费者：知识建立者（Knowledge Building），享乐主义浏览者（Hedonic Browsing），直接购买者（Directed Buying），搜索者（Search/Deliberation）和浅尝辄止者（Shallow）；对预测性样本，用判别分析的方法检验了上述五种类型的分类结果，并得到了很好的拟合。

#### 1.2 时间段点击流研究

多时间段的点击流研究相对稍多，一般采用真实的购物网站的长时间的点击流数据，拟合出相应的访问行为模型，并把关注的重点，集中在站内浏览行为和购买转换率上。

##### 1.2.1 站内浏览行为

Bucklin和Sismeiro（2003）从访问者决定继续浏览（通过递交另一个页面请求）或者结束对该网站的浏览，和浏览每个网页所花费的时间这两个方面来考察访问者浏览行为的基本特征，提出了网站浏览行为的Type II tobit模型。该模型非常适合对持续性日志文件的审查，并且允许页面浏览和页面请求这两个浏览行为成分具有正相关或者负相关。作者主要关注访问深度和重复访问次数对访问者是否继续留在该网站以及在该网站的停留时间的影响，并控制了一些可能对其产生影响的因素，如访问页

面的内容，特定任务的完成，以及其他网站特定性因素。研究者用这个模型去拟合一个网上汽车公司的5000个访问者的随机样本的个体水平的点击流数据<sup>[5]</sup>。结果发现，访问者继续浏览网站的倾向随着对该网站访问的深度和对该网站重复访问的次数的改变而改变。具体来说，重复访问减少了网页浏览的倾向但并没有减少网页浏览的持续时间，这个结果揭示的浏览类型也许反映了访问者的节省时间策略。因此，作者认为在考虑网站水平的浏览行为时，需要去控制跨时间段的异质性。

Montgomery, Li和Srinivasan等人(2004)整合了页面类型、重复访问的次数、访问者人口统计特征等变量来建立网上浏览行为的预测模型<sup>[6]</sup>。并使用来自一个大型网上书店的数据来评估这个模型。结果发现，模型的记忆成分对精确的预测网上浏览行为是至关重要的。这与Bucklin和Sismeiro(2003)的结果相一致。同时，研究者还认为，该模型对预测购买转换率也很有效，研究结果发现，在6次浏览之后，预测购买的准确性增加了40%，这比没有路径信息的基础购买转换预测率7%更好。

### 1.2.2 购买转换率

事实上，许多网站，包括零售商和大型购物商城，以及产品目录提供者，都会例行公事一样去监控网站访问者的交易情况，从而从这些监控数据中获益，提升网站的价值。但是，这些例行公事般的监控，如每个月的总访问量等，只应用了点击流数据所能提供的信息中的一小部分。点击流数据提供了非常丰富的信息，哪些信息是有用的，如何获得，尤其是访问者的访问行为如何转换成为购买行为，都是研究者和实际应用者关注的焦点。

Moe和Fader(2004)基于网站点击流数据，提出个体水平的访问和转换行为模型<sup>[7,8]</sup>。该模型使用参数表征了网站访问行为跨时间段的变异性，以及访问者随着该网站获得经验而随时间产生的一些变化：访问效应和购买阈限效应。访问效应与访问者的动机相联系，购买阈限说明了对网上购买的心理抗拒，它会随着顾客在该购物网站通过购买过程获得更多的经验而增加或者减少。该模型根据这两个关键成分的复杂性不同而有一定的变化。根据对两个大型电子商务网站——Amazon和CDNOW的8个月的点击流数据的分析结果，发现访问者对一个零售商店访问得越频繁，其购买的倾向越大。

一般的，消费者在网上买东西时，必须完成一

些特定任务(Nominal User Tasks, NUTs)，如在购物篮里添加商品，提供个人信息，确认购买，结账等。因此，Sismeiro和Bucklin(2004)假设，访问者完成这些任务就表明他们具有购买意图，并用一个多水平的二元选择模型来预测访问者对这些任务的完成，从而来预测购买行为<sup>[9]</sup>。也就使说，作者将购买过程分解成一系列有顺序的需要完成的任务，同时也用相关参数来表征用户的异质性。为了评估该模型，作者用来自一个卖汽车的网站的点击流数据来拟合，结果发现，访问者的浏览经验和导航行为预测了所有水平的任务完成。而重复访问的数量本身并不是购买倾向的诊断性因素，而且网站提供的决策帮助系统对增加购买转换率并没有显著影响。同时，用该模型去拟合一个坚持不进行网上购物的样本，发现该任务完成方法比单策略基准模型提供了更好的预测并且更好的辨识了可能的购买者，尤其是在任务序列的早期。

Poel和Buckinx(2005)在前人研究的基础上，总结了四类变量来预测网上购物行为：(1)在访问水平的一般点击流行为，(2)更细节的点击流信息，(3)消费者人口统计信息，以及(4)历史购物行为。并通过一个卖酒和相关产品的网上零售商的点击流数据，来评估这四类预测变量的预测力，从而期望获得最佳的预测变量集<sup>[10]</sup>。在所有的这四类变量中，包括以前的研究者使用的一些变量以及作者新增的一些变量，作者使用logit模型，并用向前和向后的变量选择技术来预测访问者下一次对该网站的访问是否会产生购买，结果发现，来自所有这四个类别中的预测因素都包含在了预测因素的最佳子集中，在分类消费者上是相对最重要的是细节点击流变量。

## 2 网站间点击流研究

我们所获取的点击流数据，可以分为站点中心(Site-centric)的点击流数据和用户中心(User-centric)的点击流数据。在任何单一的电子商务网站获取的数据(站点中心的数据)都是不完整的，因为它没有获取用户在网站间的浏览行为(用户中心的数据)，而大多数目前的个性化技术都仅仅基于网站中心的数据。Padmanabhan, Zheng和Kimbrough(2001)对来自20 000个用户的浏览行为的用户水平的点击流数据进行收集，发现在预测任何一个时间段的保留者是否会产生购买行为，以及一个特定的用户是否会在任何下一个时间段产生购

买行为时,建立在用户中心数据上的模型,比建立在站点中心数据上的模型,有更好的表现<sup>[11]</sup>。因此,站间的点击流研究对更全面具体的理解用户的点击流行为具有非常重要的意义。

站间的点击流研究主要集中在对站间浏览行为和站间搜索行为的研究上,或者进一步完善和发展了站内浏览行为模型的适用范围,或者从搜索的角度来研究用户对网站的选择,以及搜索行为的深度和动态。

### 2.1 站间浏览行为

Park和Fader(2004)认为,一个消费者在一个或者多个其他网站间的访问类型可能提供了他未来对这个他感兴趣的网站的访问类型的相关信息,如访问时间、频率等<sup>[12]</sup>。基于上述考虑,作者发展了一个跨网站访问行为的随机时间模型,该模型整合了两个来源的变量:对一系列竞争性网站的到访时间以及访问倾向(比率)。作者使用来自Media Metrix公司的近20 000名专项调查小组成员的点击流数据,记录了他们对两种产品类型:书籍和音乐,以及两类相关网站:Amazon.com, Barnesandnoble.com和CDNOW.com, Musicboulevard.com的访问情况,来理解如何根据从一个网站获得的信息来帮助解释消费者在另一个网站的行为。变化这个跨网站访问行为随机时间模型的复杂程度,考察了独立模型(不同网站的访问类型完全独立),只考虑访问倾向的模型,只考虑到访时间的模型,以及全模型的拟合程度,发现其拟合程度依次提高。此外,该随机时间模型使用来自一个竞争性网站的过去访问类型的概要信息(如,时间和频率)能够准确陈述对一个给定网站的“0阶层”(以前的非访问者)的未来行为。

### 2.2 搜索行为

与浏览行为不同,对站间搜索行为的研究主要集中在对不同的搜索工具的使用和搜索的产品数量上。Rowley(2000)对网上产品的搜索行为作了一个综述性的回顾,介绍了两种信息搜索的方法:浏览和直接搜索,并认为,以浏览开始的搜索可能目的非常明确,反之,目标非常明确的搜索也可能导致其他方面的浏览行为<sup>[13]</sup>。同时,网上也存在着两种类型的搜索工具:一般目的的搜索工具(如浏览器,搜索引擎),特定目的的搜索工具,如购物机器(shopping bots),相同的搜索策略,不同的搜索引擎,可能会得到完全不同的结果。同时,由

于购物机器的出现,消费者可以不用浏览相应的目标网站就可以进行购物行为,因此,作者也建议网上零售商需要把店铺设计得更加有趣,从而吸引消费者的访问,而不是单纯的购物机器。

Johnson, Moe, Fader等人(2004)认为,个体倾向于搜索非常少的站点,并且随着时间而动态发展,服从一种对数的分布,也就是说,个体逐渐学会搜索<sup>[14]</sup>。通过对来自超过10 000个网上家庭的3个日用品(书,CD,航空服务)的专项调查点击流数据,从搜索深度,搜索动态和搜索行为三个方面来进行分析,发现,平均而言,每个家庭在每个月仅仅访问1.2个书籍网站,1.3个CD网站,以及1.8个旅行网站,其中,除了航空服务(逐渐减少),对书籍和CD的搜索倾向并没有随着时间而改变。但通过对消费者特征的分析作者还发现,更积极的网上购物者倾向于搜索更多的网站,说明可能是消费者特征,而非搜索经验,影响了消费者的搜索水平。

## 3 点击流研究的应用

点击流研究的应用,其实是点击流研究在实践领域的一些发展。目前,涉及点击流研究的实践领域主要集中在数据挖掘、决策支持与推荐,以及电子商务网站的设计上。这三个方面并不是完全孤立的,例如,消费者决策支持系统事实上整合在零售电子商务网站中,从而使电子商务网站更加个性化,而数据挖掘的数据来源,则是电子商务网站后台的服务器日志文件,或者是相应的数据库。

### 3.1 数据挖掘

网页服务器日志文件是一般电子商务数据挖掘的来源,它提供了一个成功的数据挖掘所需要的所有因素:大量的数据,众多的特征,干净的数据,实践性的领域,可测量的投资回报(Return On Investment, ROI)等,但是相对的,网页服务器日志并不是无所不包的,毕竟它本身是设计用来排除服务器故障,而不是用来做数据挖掘的,因此存在一些局限性,如,网页服务器日志文件并不能准确辨识时间段或者访问者,缺乏鉴定性事件(如购物篮中包含哪些商品?),缺乏网站类型和该URL(Uniform Resource Locator)的具体信息,信息过于冗长等等<sup>[15]</sup>。

通过数据挖掘我们可以获得什么,如何进行数据挖掘?其实是一个技术性问题。Han和Kamber(2001)认为,数据挖掘方法主要是为了解决两种类型的数据挖掘问题:预测和发现知识。Perner和

Fiss (2002) 补充了这个分类, 并认为为了解决预测问题, 会主要使用分类 (Classification) 和回归 (Regression) 方法; 而为了发现知识, 则主要使用离差检测 (Deviation Detection)、聚类分析 (Clustering)、联结规则 (Association Rules) 以及可视化 (Visualization) [16]。上述的分类并不是说, 一种分类就只使用一种相应的统计方法, 例如, 为了检测联结规则, Blanc和Giudici (2002) 考虑了序列规则 (Sequence Rules)、差额比率和图示线性模型 (Graphical Loglinear Models) 的使用, Blanc和Tarantola (2002) 考虑了贝叶斯网络 (Bayesian Networks) 和从属网络 (Dependency Networks) 的使用, 而Rognoni, Giudici和Polpettini (2002) 则考虑用马尔科夫链 (Markov Chains) 来直接估计从一个页面到另一个页面的转换概率[17]。而IBM的Lee和Podlaseck (2000) 则在Starfield显示器的基础上, 开发出Starfield可视化 (Visualization) 软件来分析网上商店的点击流数据和产品的绩效。这个可视化系统通过分配并行的坐标到点击流的顺序序列中, 提供给用户更大的能力来解释和探索网上商店的点击流数据, 并使这些点击流信息更加形象和生动 [18,19]。

### 3.2 决策支持与推荐

决策支持与推荐是电子商务中通常用到的决策支持系统的功能, 决策支持系统的理论、方法和应用研究比较多[20], 大多数应用研究采用了问卷调查的方法, 或者根据其研究的重点进行实验室实验, 只有很少的决策支持系统的应用研究使用到了点击流数据。Kim, Yum和Song等人 (2005) 使用来自导航和消费者行为类型中获得的点击流数据, 提出并检验了一个新的基于协作过滤 (collaborative filtering, CF) 的推荐系统, 来估计消费者对其点击过但却没有购买的产品的偏好水平, 然后CF根据消费者的偏好水平来进行推荐[21]。作者认为, 该方法能够应用到大多数电子商务网站, 而且能够同时获得导航和消费者行为类型。

### 3.3 电子商务网站的设计

网上个性化是电子商务公司的兴趣之所在, 而所有的个性化技术都是基于储存的消费者阶段性历史数据, 而且越多越好, 这样, 当消费者访问该电子商务网站时, 特性良好, 细节的历史阶段性数据能够基于当前的网上导航行为类型来对消费者提出问题, 从而获得实时的回复。当网上消费者越来越

多时, 用户登录和历史数据的数量也会持续增长, 从而当前几乎所有的个性化技术, 都需要考虑量化相关的问题。VanderMeer, Dutta和Datta (2000) 提出并实证检验了一种实时个性化系统架构[22]。该架构将网页服务器 (Web/E-commerce Server, WES) 与eGlue Server (eGS) 整合起来, 两者可以同时交换信息, eGlue服务器通过对阶段性数据的分析, 进行第三方的数据挖掘, 从而形成能快速抽取消费者剖面图的数据仓库, 进而到达点击流存储器 (Clickstream Cache) 和剖面存储器 (Profile Cache), 此外, 消费者不仅可以与网页服务器进行交互, 而且其阶段性数据也被直接输送到eGS中的阶段性日志文件模块 (Session Log) 进行分析。

## 4 点击流研究的局限性及研究展望

网络技术的发展和电子商务的出现改变了人们对交易是什么以及如何操作等各个方面的看法和实际行为, 使经济发展的前景呈现出网络化、信息化的新局面。网上购物作为区别于传统购物的一种新的购物方式, 对该购物方式相关领域的研究就显得十分有意义。而将本用来排除服务器故障的服务器日志文件, 作为获取点击流数据的来源之一, 来研究网上消费者, 包括网站内的点击流和网站间的点击流行为, 并反过来更加促进数据挖掘、决策支持和推荐以及电子商务网站设计技术的发展, 可以获得事半功倍的效果。

虽然点击流数据存在着一些局限性, 如信息冗长, 不能准确辨识时间段和访问者, 缺乏鉴定性的事件等[15], 给实际操作带来了很大困难, 但是相应的可供数据挖掘的工具也正在发展中, 如IBM的Lee和Podlaseck (2000) 开发的Starfield可视化 (Visualization) 软件[18,19], VanderMeer, Dutta和Datta (2000) 提出的实时个性化系统架构[22]等; 同时, 点击流数据仓库也在蓬勃发展, 虽然其关键技术——抽取、转换和装载机制 (Extract-Transform-Load, ETL) 尚未有能够很好的处理点击流数据的相应软件, 但是目前主要的ETL供应商, 如Informatica, Ardent和Sagent也能够为点击流数据提供初级的支持[3]。可见, 点击流研究是一种跨领域, 技术相关性高的新的研究范式, 未来有志于点击流的研究者, 尤其是心理学研究者, 要么需要具有管理学、市场营销、计算机科学等领域的更加广博的知识; 要么需要与这些领域的研究者和网站设计、数据仓库、决策支持系统等实际应用领域的专家共

同合作,从而为这一研究范式的蓬勃发展开拓一片崭新的天地。

### 参考文献

- [1] Bucklin R E, Lattin J M, Ansari A, et al. Choice and the internet: from clickstream to research stream. *Mark Lett* 2002, 13(3): 245~258
- [2] Dellaert, Benedict G C, Arentze T A, et al. Timmermans, Investigating consumers' tendency to combine multiple shopping purposes and destinations. *Marketing Research*, 1998, 25(4): 342~355
- [3] Sweiger M., Madsen M R., Langston J, et al 著. 陆昌辉等译. 点击流数据仓库. 北京: 电子工业出版社, 2004. 28~53
- [4] Moe W W. Buying, searching, or browsing Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. *Journal of Consumer Psychology*, 2003, (1&2): 29~39
- [5] Bucklin R.E., Sismeiro C.. A Model of Web Site Browsing Behavior Estimated on Clickstream Data. *Journal of Marketing Research*, 2003, 40(3): 249~267
- [6] Montgomery A L, Li Shibo, Srinivasan K, et al. Modeling Online Browsing and Path Analysis Using Clickstream Data. *Marketing Science*, 2004, 23(4): 579~595
- [7] Moe W W., Fader P S. Capturing evolving visit behavior in clickstream data. *Journal of Interactive Marketing*, 2004, 18(1): 5~19
- [8] Moe W W., Fader P S. Dynamic Conversion Behavior at E-Commerce Sites. *Management Science*, 2004, 50(3): 326~335
- [9] Sismeiro C, Bucklin R E. Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach. *Journal of Marketing Research*, 2004, 41(3): 306
- [10] Dirk Van den Poel, Buckinx W. Predicting online-purchasing behaviour. *European Journal of Operation Research*, 2005, 166(2): 557~575
- [11] Padmanabhan B, Zheng Zhiqiang, Kimbrough S O. Personalization from Incomplete Data: What You Don't Know Can Hurt. *ACM*, 2001. 154~163
- [12] Park Young-Hoon, Fader P S. Modeling Browsing Behavior at Multiple Websites. *Marketing Science*, 2004, 23(3): 280~303
- [13] Rowley J, Product search in e-shopping: a review and research propositions. *The Journal of Consumer Marketing*, 2000, 17(1): 20~35
- [14] Johnson E J, Moe W W, Fader P S, et al. On the Depth and Dynamics of Online Search Behavior. *Management Science*, 2004, 50(3): 299~308
- [15] Kohavi R. Mining E-Commerce Data: The Good, the Bad, and the Ugly. *ACM*, 2001. 8~13
- [16] Perner P, Fiss G. Intelligent E-marketing with Web Mining, Personalization, and User-Adpated Interfaces. *Advances in Data Mining 2002, LNAI 2394: 37~52*
- [17] Blanc E, Giudici P. Sequence Rules for Web Ckicestream Analysis. *Advances in Data Mining*, 2002, LNAI2394: 1~14
- [18] Lee J, Podlaseck M. Using a starfield visualization for analyzing product performance of online stores. *ACM*, 2000. 168~175
- [19] Lee J, Podlaseck M. Visualization and Analysis of Clickstream Data of Online Stores with a Parallel Coordinate System. *ACM*, 2000. 145~154
- [20] 李双双, 陈毅文. B2C电子商务中的消费者决策支持系统. *心理科学进展*, 2006, 14 (3) : 438~442
- [21] Yong Soo Kim, Bong-Jin Yum, Junehwa Song, et al. Development of a recommender system based on navigational and behavioral patterns of customers in e-commerce sites. *Expert Systems with Applications*, 2005, 28(2): 381~393
- [22] VanderMeer D, Dutta K, Datta A. Enabling scalable online personalization on the web. *ACM*, 2000. 185~196

## Clickstream: A New Paradigm to Study on Online Consumer

Li Shuangshuang<sup>1,2</sup> Chen Yiwen<sup>1</sup>

<sup>1</sup>*Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China*

<sup>2</sup>*Graduate School of Chinese Academy of Sciences, Beijing 100039, China*

**Abstract:** Clickstream researches bring a new research stream, not only in many academic domains, but also in many practical domains. Clickstream data record the track when a consumer clicks on the sites. Most of the former researches concentrate on building consumers' online behavior models, so to get further understanding of consumers' online behavior. The analysis of clickstream data is also benefit for data mining, web designing and decision-making supporting system as well. Therefore, we classify three types of researches in existence: Within-Site clickstream researches, Between-Site clickstream researches, and applied researches of clickstream data. The limitation and the future development of clickstream research as a new paradigm are also discussed.

**Key words:** clickstream, consumer, online shopping.