

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06Q 10/00 (2006.01)

G06Q 30/00 (2006.01)



# [12] 发明专利申请公布说明书

[21] 申请号 200780011332.8

[43] 公开日 2009年4月15日

[11] 公开号 CN 101410864A

[22] 申请日 2007.3.29

[21] 申请号 200780011332.8

[30] 优先权

[32] 2006.3.29 [33] US [31] 11/394,343

[86] 国际申请 PCT/US2007/007701 2007.3.29

[87] 国际公布 WO2007/126903 英 2007.11.8

[85] 进入国家阶段日期 2008.9.27

[71] 申请人 雅虎公司

地址 美国加利福尼亚州

[72] 发明人 约舒亚·M·克兰

克里斯蒂娜·业·钟

阿比纳维·谷坡塔 乔治·H·约翰

隆-基·林 洪峰·尹

理查德·弗兰克尔

[74] 专利代理机构 北京东方亿思知识产权代理有  
限责任公司

代理人 宋鹤南 霆

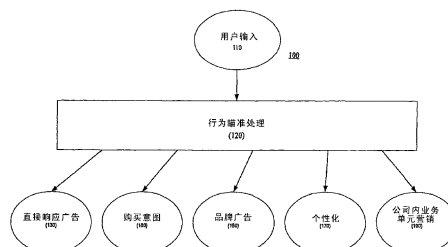
权利要求书 3 页 说明书 29 页 附图 15 页

[54] 发明名称

行为瞄准系统

[57] 摘要

一种行为瞄准系统从在线活动确定用户简档。该系统包括多个定义用于确定用户简档得分的参数的模型。在一实体处接收包括用户的在线活动的事件信息。为了生成用户简档得分，选择一模型。模型包括新近度、强度和频率维度参数。行为瞄准系统为诸如品牌广告或直接响应广告之类的瞄准目标生成用户简档得分。来自模型的参数被应用来生成一类别中的用户简档得分。行为瞄准系统可以应用来向在线用户提供广告。



1. 一种用于从在线活动确定用户简档的方法，所述方法包括：  
在实体处存储多个模型，其中模型定义用于确定用户简档得分的多个参数；  
在所述实体处接收至少一个用户事件信息，其中所述用户事件信息包括所述用户和所述实体之间的在线活动；  
选择模型来生成用户简档得分；以及  
通过将来自所选择的所述模型的所述参数应用到所述用户事件信息来生成一类别中的至少一个用户简档得分。
2. 如权利要求 1 所述的方法，还包括：  
添加包括新参数和规则的至少一个新模型，以便提供可扩展性；  
在所述实体处接收事件信息；  
选择所述新模型来生成用户简档得分；以及  
通过将来自所选择的所述新模型的所述新参数和规则应用到所述用户事件信息来生成一类别中的至少一个用户简档得分。
3. 如权利要求 1 所述的方法，还包括基于所述用户兴趣得分向所述用户提供广告。
4. 如权利要求 1 所述的方法，其中所述参数包括新近度维度参数，该新近度维度参数基于所述用户事件信息发生的新近度来向所述用户兴趣得分提供输入。
5. 如权利要求 1 所述的方法，其中所述参数包括强度维度参数，该强度维度参数基于所述用户事件信息预测用户对所述类别的感兴趣的有效度来向所述用户兴趣得分提供输入。
6. 如权利要求 1 所述的方法，其中所述参数包括频率维度参数，该频率维度参数基于所述用户事件信息发生的频率来向所述用户兴趣得分提供输入。
7. 如权利要求 1 所述的方法，其中  
选择模型以生成用户得分包括基于瞄准目标来选择模型；并且

生成至少一个用户兴趣得分包括为所述瞄准目标生成用户兴趣得分。

8. 如权利要求 7 所述的方法，其中所述瞄准目标包括品牌广告。

9. 如权利要求 7 所述的方法，其中所述瞄准目标包括直接响应广告。

10. 一种用于从在线活动确定用户简档的系统，所述系统包括：

用于存储多个模型的存储装置，其中模型定义用于确定用户简档得分的多个参数；

至少一个服务器，用于：接收至少一个用户事件信息，其中所述用户事件信息包括所述用户的在线活动；选择模型来生成用户简档得分；以及通过将来自所选择的所述模型的所述参数应用到所述用户事件信息来生成一类别中的至少一个用户简档得分。

11. 如权利要求 10 所述的系统，其中所述系统包括可扩展系统，使得可以在无需改变现有系统的情况下添加至少一个模型。

12. 如权利要求 10 所述的系统，还包括广告服务器，用于基于所述用户兴趣得分向所述用户提供广告。

13. 如权利要求 10 所述的系统，其中所述参数包括新近度维度参数，该新近度维度参数基于所述用户事件信息发生的新近度来向所述用户兴趣得分提供输入。

14. 如权利要求 10 所述的系统，其中所述参数包括强度维度参数，该强度维度参数基于所述用户事件信息预测用户对所述类别的感兴趣的有效度来向所述用户兴趣得分提供输入。

15. 如权利要求 10 所述的系统，其中所述参数包括频率维度参数，该频率维度参数基于所述用户事件信息发生的频率来向所述用户兴趣得分提供输入。

16. 如权利要求 10 所述的系统，其中所述服务器还用于为瞄准目标生成用户兴趣得分。

17. 如权利要求 16 所述的系统，其中所述瞄准目标包括品牌广告。

18. 如权利要求 16 所述的系统，其中所述瞄准目标包括直接响应广告。

19. 一种包括一组指令的计算机可读介质，该组指令在被计算机执行时使得该计算机从在线活动确定用户简档的计算机可读介质，所述指令用于：

在实体处存储多个模型，其中模型定义用于确定用户简档得分的多个参数；

在所述实体处接收至少一个用户事件信息，其中所述用户事件信息包括所述用户和所述实体之间的在线活动；

选择模型来生成用户简档得分；以及

通过将来自所选择的所述模型的所述参数应用到所述用户事件信息来生成一类别中的至少一个用户简档得分。

20. 如权利要求 19 所述的计算机可读介质，还包括基于所述用户简档得分向所述用户提供广告。

21. 如权利要求 15 所述的计算机可读介质，其中所述参数包括新近度维度参数，该新近度维度参数基于所述用户事件信息发生的新近度来向所述用户兴趣得分提供输入。

22. 如权利要求 19 所述的计算机可读介质，其中所述参数包括强度维度参数，该强度维度参数基于所述用户事件信息预测用户对所述类别的兴趣的有效度来向所述用户兴趣得分提供输入。

23. 如权利要求 19 所述的计算机可读介质，其中所述参数包括频率维度参数，该频率维度参数基于所述用户事件信息发生的频率来向所述用户兴趣得分提供输入。

24. 如权利要求 19 所述的计算机可读介质，其中

选择模型以生成用户得分包括基于瞄准目标来选择模型；并且

生成至少一个用户简档得分包括为所述瞄准目标生成用户简档得分。

## 行为瞄准系统

### 技术领域

本发明涉及行为瞄准（behavioral targeting）领域，更具体而言涉及在线行为瞄准系统。

### 背景技术

因特网为商家提供了向消费者提供大量产品和服务的机制。因特网门户向用户提供了到大量因特网资源的入口和引导。通常，因特网门户提供一系列搜索、电子邮件、新闻、购物、聊天、地图、金融、娱乐和其他因特网服务和内容。本发明的受让人 Yahoo 是这种因特网门户的示例。

当用户访问包括因特网门户在内的因特网上的某个位置（例如网站）时，用户以在线活动的形式输入信息。该信息可以被记录和分析以确定用户的行为模式和兴趣。进而，这些行为模式和兴趣可被用于瞄准用户以提供在因特网上（例如因特网门户站点）上的更有意义和丰富的体验。例如，如果确定了用户对某些产品和服务的兴趣，则可以向用户提供与这些产品和服务有关的广告。提供广告的行为瞄准系统既有益于向目标受众提供其消息的广告主，也有益于接收用户感兴趣的领域中的广告的用户。

### 发明内容

一种行为瞄准系统从在线活动确定用户简档（user profile）。该系统包括多个模型。模型定义用于确定用户简档得分的参数。在一实体处接收事件信息。事件信息包括用户的在线活动。为了生成用户简档得分，选择一模型。该系统通过将来自该模型参数应用到用户事件信息来生成一类别中的用户简档得分。该行为瞄准系统可用在对在线用户的广告服务中。

该行为瞄准系统的体系结构是可扩展的。例如，可以添加包含不同的规则和权重参数的额外模型。基于新模型的数据和元数据，行为瞄准系统

可以在相同的整体框架内执行不同的行为模型和规则。

在一个实施例中，模型包括新近度（recency）、强度（intensity）和频率（frequency）维度（dimension）参数。新近度维度参数基于用户事件信息发生的新近度指定对简档得分的输入。强度维度参数基于用户事件信息预测类别中的瞄准目标（target objective）的有效度来向用户简档得分提供输入。另外，频率维度参数基于用户事件信息发生的频率来向用户简档得分提供输入。

在一个实施例中，行为瞄准系统为瞄准目标生成用户简档得分。例如，瞄准目标可以包括品牌广告和直接响应广告。

### 附图说明

图 1 是示出通用行为瞄准系统的框图。

图 2 示出识别营销目标（marketing objective）的营销漏斗（marketing funnel）。

图 3 是示出通用行为瞄准系统的一个实施例的框图。

图 4 是示出本发明的行为瞄准系统的一个实施例的框图。

图 5 是示出本发明的通用行为瞄准系统的一个实施例的流程图。

图 6 是示出行为瞄准系统的事件分类的一个实施例的框图。

图 7 是示出根据本发明一个实施例的用于促进网络中的自动事件分类的方法的流程图。

图 8 是示出用于利用行为瞄准系统来提供广告的一个实施例的框图。

图 9 是示出用于行为瞄准系统中的维度处理的一个实施例的框图。

图 10 是示出用于生成长期直接响应用户兴趣得分的一个实施例的框图。

图 11 是示出用于生成短期直接响应用户兴趣得分的一个实施例的框图。

图 12 是示出生成长和短期用户兴趣得分的行为瞄准系统的一个实施例的框图。

图 13 示出了一曲线图，该曲线图基于从多个用户聚集的统计数据示

出了作为直接响应得分的函数的 CTR 得分。

图 14 示出了用于本发明的行为瞄准系统的操作的网络环境的一个实施例。

图 15 示出了行为瞄准系统的概念图。

### 具体实施方式

行为瞄准系统可以应用来针对一个或多个瞄准目标识别在线用户的兴趣和行为。图 1 是示出通用行为瞄准系统 100 的框图。一般地，行为瞄准系统 100 基于用户的在线活动来剖析因特网用户的兴趣和行为。如图 1 所示，通过行为瞄准处理 (120) 来捕捉用户输入 (110)。在一个实施例中，用户输入包括一个或多个“事件”。如下文更充分描述的，事件是一类由用户发起的动作（例如，用户点击横幅广告）。如图 1 所示，行为瞄准处理 (120) 生成对某一瞄准目标特定的多个用户简档。一般地，用户行为简档预测特定瞄准目标预测的用户成功程度。例如，用户行为简档可以预测用户对直接营销广告活动作出响应的倾向。在一个实施例中，用户行为简档是在每类别的基础上表达的。例如，用户简档输出可以指示出用户是关于主题“金融”的目标的良好候选，并且是关于主题“音乐”的目标的不好候选。

行为瞄准处理 120 为各种瞄准目标生成用户行为简档。在图 1 所示的一个实施例中，行为瞄准系统 100 为直接响应广告 (130)、品牌知名度广告 (150)、购买意图活动 (180) 和公司内业务单元营销输出用户行为简档。在一个实施例中，行为瞄准处理 120 为营销目标生成用户兴趣简档。

图 2 示出识别营销目标的营销漏斗。在漏斗的顶部，广告主可能希望获取广告主的品牌的品牌知名度。通常，对于这类营销，广告主的目的在于通过将一个或多个正面形象与产品的品牌相关联来推广该品牌。该营销目标在图 1 中被示为品牌广告 (150)。在漏斗的第二阶段中，用户可能希望为产品考虑而收集信息。为了应对买方的这个周期，广告主可以使用直接响应广告。可能有许多不同的目标与直接响应广告相关联，其中包括

获取、保持、吸引和货币化。获取的目的在于使得消费者成为产品/服务的顾客或访问者。保持的目的在于维持现有的顾客或访问者（例如网站的访问者）。与吸引相关联的目的是在现有顾客中诱发更多的活动。货币化的目的在于经由诸如交叉销售之类的主动购买活动以及诸如消耗横幅广告之类的被动活动来增大顾客的获利能力。

在直接响应广告中，总体目的在于诱发来自顾客的动作或响应。本发明的行为瞄准系统为直接响应广告（130）提供了用户简档数据。例如，在网页上显示的、包括供用户“点击”的链接的广告是直接响应广告的示例。漏斗的最后一个也是最受关注的部分是顾客的购买意图周期。在这一阶段中，用户在积极地购物，并且意图进行购买。本发明的行为瞄准系统可以应用来为购买意图应用（180）提供用户简档数据。

图 1 还示出了作为营销或瞄准目标的公司内业务单元营销应用。例如，因特网公司可能希望将用户吸引到网站或门户。行为瞄准处理可以生成用户简档以将用户获取到因特网门户。例如，可以瞄准来自门户的一个业务单元的用户以基于用户的简档在一个不同的业务单元领域中获取用户。类似地，可以为吸引用户更频繁地访问因特网门户或网站的瞄准目标生成用户简档。另外，行为瞄准处理可以生成用户简档以保持先前已经访问过因特网门户或网站的用户。

营销目的和目标诱发用户简档数据以帮助广告主和营销者。本发明的行为瞄准系统还可以应用来提供用户简档信息以促进用户的在线体验。一般地，目的是在正确的时间将正确的内容匹配到正确的用户（即，刺激）。例如，行为瞄准系统可以为个性化应用生成用户简档。一般地，个性化应用使用行为瞄准信息来为用户定制网页或网站体验。例如，因特网门户可以显示可用的产品和服务的类别。因特网门户可以基于用户简档信息来定制一个或多个兴趣视图。在另一个实施例中，行为瞄准信息可用于确定供用户消耗的内容主题。例如，在确定用户希望的内容之后，去往这种内容的链接可被显示在因特网门户上，以允许用户访问该内容。

在用户的计算机上显示的边栏（sidebar）应用提供了去往用户的关键目的地的链接，而不要求用户启动浏览器。在个性化应用的另一个实施例



中，行为瞄准信息被用于选择链接以便定制边栏应用。个性化应用的瞄准目标可以是使用户体验最大化（即，以用户为中心的瞄准目标）或者使营销者向用户递送内容的目标最大化。

本发明的行为瞄准系统的体系结构提供了作为通用行为瞄准系统工作的框架。这些模型包含用于定制对事件信息的处理的元数据和规则。从而，通过选择不同的模型和规则，实现了由元数据指定的不同的程序流程或执行。本发明的行为瞄准系统的体系结构是可扩展的。如果新的模型和规则（例如，新的数据和元数据）被创建，则新的模型和规则被插入到一般体系结构中以向行为瞄准系统提供新的功能和流程。从而，系统可以被修改，以支持新的模型并且处理新的事件类型。

一般框架可被配置为从相同输入或输入子集生成一个或多个输出。例如，行为瞄准系统可以被配置为为不同的用户或瞄准（例如营销）目标生成多个输出。本发明的行为瞄准系统依据系统的目标而使用不同的模型。从而，可以实现单个行为瞄准系统来支持各种目标，包括以上指明并在图1中示出的那些应用。

图3是示出通用行为瞄准系统的一个实施例的框图。通用行为瞄准系统为一个或多个目标生成用户得分。为此，模型和规则340为不同的营销和用户目标存储多个模型。在操作中，用户行为信息被输入到行为处理310。另外，一个或多个用户/营销目标被输入到模型和规则340，以选择一个或多个模型。进而，来自模型和规则340的一个或多个模型被输入到行为处理310。行为处理310使用所选择的模型和规则来生成用户得分。具体而言，为一个或多个用户/营销目标生成用户得分。

在一个实施例中，用户行为得分或简档可用于报告。对于此实施例，系统将输入处理到用户简档中。然后，营销者确定如何最有效地使用简档：瞄准内容/搜索结果、个性化内容/搜索结果、或者通过将简档输入到各种报告和分析系统中来利用简档了解消费者的态度。

在一个实施例中，行为瞄准系统还向营销者输出信息，以允许它们了解每得分范围每类别的清单（即，在特定类别中具有给定得分范围的人/cookie的数目，以及/或者与这些人/cookie相关联的事件的预报）。该信

息向营销者标识出大致的受众数量，以便有效地使用行为简档，而不管是以营销者为中心的使用（即，瞄准）还是以访问者为中心的使用（即，个性化）。

在一个实施例中，本发明的行为瞄准系统按类别生成简档得分，以为瞄准目标的类别中的主题预测用户倾向。例如，兴趣得分可用来模拟用户购买某一类别内的产品或服务的兴趣。图 4 是示出本发明的行为瞄准系统的一个实施例的框图。对于此实施例，以这里被称为“事件”的输入事件被输入到事件分类模块 410。事件分类模块 410 结合分类体系 420 工作。具体而言，事件分类模块 410 将每个输入事件分类到分类体系 420 中的一个类别。一般地，分类体系 420 定义用于对用户兴趣分类的多个类别。分类体系 420 内的类别可以被分级地排列。例如，分类体系 420 可以包括针对“音乐”的高级类别，以及位于“音乐”类别之下的若干个子类别，用于不同风格的音乐。用于对主题分类的类别的任何分类体系都可以结合行为瞄准系统使用，而不脱离本发明的精神或范围。

如图 4 所示，经过分类的事件被输入到维度处理 450 并被输入到模型和规则 440。另外，用户/营销目标被输入到模型和规则 440。模型和规则 440 包含用于不同目标或目的（例如，直接响应广告（442）、品牌广告（444）、购买意图（446）和个性化（448））的多个不同模型。每个模型具有与该模型相关联的多个规则。规则被用于在维度处理（450）中对事件进行处理，以为不同的目标或目的生成用户得分。在一个实施例中，与模型相对应的规则包含对该事件特定的参数以及为该事件分类的类别。例如，如果事件类型是在“金融”中分类的“页面查看”，则模型和规则 440 基于事件“页面查看”和类别“金融”来选择参数。

在模型和规则 440 中选择的参数被输入到维度处理 450。维度处理 450 还接收经过分类的事件作为输入。一般地，维度处理 450 利用参数为类别生成用户得分。如下文中更充分描述的，维度处理 450 在指定的一段时间内累积事件以便处理。例如，在一个实施例中，维度处理 450 累积事件以生成短期用户兴趣得分。在另一个实施例中，维度处理 450 在较长的一段时间内（例如，在一个月内）累积事件以编制长期用户兴趣得分。

如图 4 所示，用户兴趣得分被输入到映射模块 460。一般地，映射模块将原始用户得分映射到一个或多个输出度量。例如，一个输出度量可以确定用户点击被分类在该类别中的广告的倾向。以百分比表达的该度量被称为“点通率”（click through rate, CTR）。对于此示例，映射模块 460 可提供用于将原始用户得分映射到 CTR 值的函数。

图 5 是示出本发明的通用行为瞄准系统的一个实施例的流程图。当用户接收到用户事件或者在用于累积事件的预定量的时间之后，系统将事件分类在分类体系的一个类别中（图 5，块 510 和 520）。基于用户/营销目标选择模型（图 5，块 530）。在块 540、550 和 560 中，行为瞄准系统对一个或多个事件进行处理以为某一类别生成原始用户兴趣得分。对于此实施例，行为瞄准系统基于所选择的模型应用新近度维度（图 5，块 540）。一般地，新近度维度依据事件发生的新近度来对一个或多个事件加权。为了应用第二维度，行为瞄准系统基于所选择的模型应用强度维度（图 5，块 550）。强度维度应用一权重或因子，以赋予一指示出事件在衡量用户对所涉类别的兴趣方面的有效度的值。对于此实施例，行为瞄准系统还应用第三维度：频率（图 5，块 560）。一般地，频率维度基于事件发生的频率来为所涉类别修改用户得分。

如图 5 所示，该过程应用循环以处理多个事件。但是，如下文中更充分描述的，行为瞄准系统可以在一段时间中累积事件，并且可以在有延迟的情况下对事件进行分批处理，以为多个类别获得用户兴趣得分。在另一个实施例中，行为瞄准系统可以实时操作。对于此实施例，行为瞄准系统可以在一小段时间（例如一小时）中累积事件，然后对这些事件进行处理以给出短期用户兴趣得分。

在一个实施例中，为输出度量转换原始用户兴趣得分（图 5，580）。但是，原始得分不需要转换来将内容匹配到用户。为了确定适当的权重和衰减参数以生成得分，使用单个营销目标（例如，对于直接响应是点通率，对于购买意图是购买事件）。来自这些行为瞄准应用的原始得分将与瞄准目标相关联起来。对原始得分的映射帮助了提高相关度。但是，一旦原始得分已被映射，就丧失了跨类别为用户的相对兴趣排名和排序的能

力。

### 事件处理和分类

在一个实施例中，事件包括广告点击、搜索查询、搜索点击、赞助列表（sponsored listing）点击、页面查看和广告查看。但是，这里使用的事件可包括任何类型的在线导航性交互或者与搜索有关的事件。一般地，页面查看发生在用户查看网页时。例如，用户可在因特网门户内通过点击针对音乐类别页面的链接来进入网页。对于此示例，为用户对音乐类别页面的查看记录页面查看事件。

广告查看事件发生在用户查看用于广告的网页时。例如，因特网门户可以在门户的主页上显示横幅广告。如果用户点击该横幅广告，则门户将用户重定向到针对相应广告主的链接。响应于点击而对网页的显示构成了广告点击事件。用户可以通过在广告主的网站处访问多个网页而生成多个页面查看事件。

广告点击事件发生在用户点击广告时。例如，网页可以显示横幅广告。当用户点击该横幅广告时，发生广告点击事件。

搜索查询事件发生在用户在基于 web 的搜索引擎上提交一个或多个搜索项（search term）时。例如，用户可以提交查询“Deep Sea Fishing”，并且相应的搜索查询事件与搜索项“Deep Sea Fishing”一起被记录。响应于用户查询，基于 web 的搜索引擎返回去与搜索查询项有关的网页的多个链接。如果用户点击这些链接之一，则发生搜索点击事件。

赞助列表广告指的是响应于用户的搜索标准显示的广告。赞助列表点击事件发生在用户点击为该用户显示的赞助列表广告时。

图 6 是示出行为瞄准系统的事件分类的一个实施例的框图。如图 6 所示，系统 608 包括事件分类平台 600，该事件分类平台 600 耦合到多个数据库，例如令牌数据库 610、兴趣数据库 620 和规则数据库 630。

在一个实施例中，令牌数据库 610 存储从各种服务器、从编辑者和/或从其他的第三方实体自动或者手动收集的单个单词或多个单词的关键字的列表，这些关键字也被称为令牌。一般来说，令牌表示单个概念，并且即

使其包含多个单词单元也被视为单个关键字。基于令牌与其各自的起源事件的关联，将令牌进一步组织成数据库 610 内的分级分类体系。在一个实施例中，令牌数据库 610 中存储的分级令牌分类体系被手动映射到进一步存储在兴趣数据库 620 内的经分类的令牌的分级分类体系。分级分类体系被事件分类平台 600 自动地审查、编辑和更新，或者被编辑者或者其他的第三方实体手动审查、编辑和更新。

映射为每个存储的令牌指派一个或多个类别，所指派的类别随后被存储在兴趣数据库 620 内与每个相应令牌相关联的各节点处。在替换实施例中，类别可以不被映射到分级分类体系中，而是可以作为类别的集合被存储在兴趣数据库 620 内。

事件分类平台 600 从前端 web 服务器接收各种事件，例如由用户通过网页发送的搜索查询、网页查看、搜索结果点击、广告点击以及其他类型的交互性事件，并且使得能够基于存储在相关联的数据库 610、620 和 630 中的数据来对接收到的事件进行自动分类，这将在下文中更详细描述。

在一个实施例中，事件分类平台 600 还包括解析器模块 602，该解析器模块 602 被配置为接收诸如搜索查询之类的事件并且解析事件以生成多个事件单元，例如查询项。事件分类平台 600 还包括令牌分析模块 604，该令牌分析模块 604 耦合到解析器模块 602 并且被配置为接收来自解析器模块 602 的事件单元并且基于事件单元和存储在相关联的数据库 610 和 620 中的数据来对事件分类，这将在下文中更详细描述。最后，事件分类平台 600 包括多义性 (ambiguity) 处理模块 606，该多义性处理模块 606 耦合到令牌分析模块 604 并且被配置为生成与每个经分类的事件相对应的多义性值，这将在下文中更详细描述。

图 7 是示出根据本发明一个实施例的用于促进网络中的自动事件分类的方法的流程图。如图 7 所示，在处理块 701 处，从用户接收事件，例如搜索查询。在一个实施例中，用户访问在客户端计算机 820 (图 8) 的客户端程序中显示的网页，并且经由客户端计算机 820 和网络 830 向 web 服务器发送搜索查询。前端 web 服务器 840 接收搜索查询，并且将查询转发到事件分类平台 600。

在处理块 702 处，解析事件以生成一个或多个事件单元。在一个实施例中，解析器模块 602 解析事件（例如搜索查询）以获得一个或多个单元（例如查询项），并将这些单元发送到令牌分析模块 604。

在处理块 703 处，从数据库取回与所解析的事件单元相匹配的令牌。在一个实施例中，令牌分析模块 604 访问兴趣数据库 620，并且取回与所解析的事件单元相匹配的一个或多个经分类的令牌。或者，令牌分析模块 604 可以访问一般的令牌数据库 610，以取回一个或多个匹配令牌。

在一个实施例中，令牌分析模块 604 将每个事件单元与存储在数据库 620 或者数据库 610 中的令牌相比较，并且选择尽可能最长的令牌（即，具有最多的单词或者最长的长度的令牌）。或者，令牌分析模块 604 选择出现在登记的事件内的可能性最高的令牌。选择是基于与每个令牌相关联的单元频率参数的，该参数指定每个特定令牌有多少次被包含在事件内。

在处理块 704 处，识别与取回的令牌相关联的一个或多个类别。在一个实施例中，令牌分析模块 604 分析所取回的经分类的令牌并且识别与所取回的令牌相关联的一个或多个类别。或者，如果令牌是从一般令牌数据库 610 取回的，令牌分析模块 604 则以编辑或算法的方式向每个取回的令牌指派一个或多个类别，所指派的类别形成相应的分级分类体系，或者，令牌分析模块 604 可以在没有相关联的类别的情况下丢弃令牌。

在处理块 705 处，为每个取回的令牌指派多义性参数。在一个实施例中，多义性处理模块 606 从令牌分析模块 604 接收令牌，并且例如为每个令牌计算相应的多义性值，作为在给定所分析的事件内特定令牌的存在的情况下令牌类别为整体事件类别的条件概率的因子。

在处理块 706 处，基于相关联的令牌类别、每个令牌被指派的多义性参数值以及存储在规则数据库 630 内的一组事件处理规则，从所取回的令牌中选择主导令牌。在一个实施例中，令牌分析模块 604 应用预定的处理规则来选择主导令牌，例如规定去除包含一个或多个停止单词的令牌、规定在存储的事件内令牌的最小出现频率的规则，以及被设计来为所取回的令牌排名的其他规则。

在一个实施例中，除了多义性参数值之外，令牌分析模块 604 还指派

表示对主导令牌选择和整体事件分类的准确度的评估的置信度得分，并且将置信度得分与相应的整体事件一起存储。

最后，在处理块 707 处，基于与主导令牌相关联的一个或多个令牌类别对整体事件分类，并且更新相应的数据库 610、620 以包括新分类的事件。

在替换实施例中，如果所指派的置信度得分低于预定的阈值得分，表明对以上详细描述的分类过程的准确度的置信度较低，则令牌分析模块 604 可以丢弃整体事件。

在另一个替换实施例中，即使所指派的置信度得分低于预定的阈值得分，令牌分析模块 604 仍可将置信度得分与相应的整体事件一起存储。在此实施例中，即使与存储的事件相关联的置信度得分低于预定的阈值得分，其他外部模块和/或系统，例如被配置为基于用户的在线活动来识别用户兴趣的行为瞄准系统，或者其任何组件，也取回并丢弃存储的事件。

在另一个替换实施例中，可以根据与每个令牌相关联的多义性参数值在与所取回的令牌相对应的多个类别之间细分事件。然后，可以根据所指派的与相应的多义性参数值相等的权重来在每个令牌类别内对事件分类。

在促进自动事件分类的方法的替换实施例中，在取回一个或多个匹配令牌之后，包括与每个所取回的事件相对应的多个统计参数的值的向量可以被组装并输入到已知的神经网络结构（未示出）中。或者，值的向量可以被输入到已知的支持向量机器（未示出）中、输入到已知的非线性回归机制（未示出）中、或者输入到任何已知的接受向量输入的机器学习单元中。

在一个实施例中，值的向量包括与每个所取回的令牌有关的数据。每个令牌的统计参数可以包括在事件日志中令牌出现的频率、在特定事件（例如搜索查询）内令牌出现的频率、令牌的多义性值、令牌主导整体事件的概率、令牌主导另一令牌的概率、与令牌相关联的类别主导整体事件的概率、该类别主导与另一令牌相关联的类别的概率，以及使得能够确定事件内的主导令牌的其他已知统计参数。

在一个实施例中，诸如神经网络结构或者支持向量机器或非线性回归

机制之类的机器学习单元被配置为接收向量输入并且确定与每个令牌相关联的输出值，该输出值指示出相应令牌是事件的主导令牌的概率。然后，机器学习单元对计算出的输出值排序、选择排名最高的输出值并且将所选择的输出值发送到令牌分析模块 604。

在一个实施例中，除了确定每个输出值来作为相应令牌是事件的主导令牌的概率之外，机器学习单元还计算置信度得分，该置信度得分表示对主导令牌确定和整体事件分类的准确度的评估。置信度得分随后被发送到令牌分析模块 604。在替换实施例中，令牌分析模块 604 可以利用从机器学习单元接收的信息来计算置信度得分。

令牌分析模块 604 识别与所选择的最高输出值相对应的令牌并且取回与令牌相关联的一个或多个类别。最后，令牌分析模块 604 基于一个或多个所取回的类别对事件进行分类并且更新相应的数据库以包括新分类的事件。

在替换实施例中，如果所指派的置信度得分低于预定的阈值得分，表明对以上详细描述的分过程的准确度的置信度较低，则令牌分析模块 604 可以丢弃整体事件。

在另一个替换实施例中，即使所指派的置信度得分低于预定的阈值得分，令牌分析模块 604 仍可将置信度得分与相应的整体事件一起存储。在此实施例中，即使与存储的事件相关联的置信度得分低于预定的阈值得分，其他外部模块和/或系统（例如行为瞄准系统）或者其任何组件也可以取回并丢弃存储的事件。

在另一个替换实施例中，值的向量包括与一对令牌有关的数据。如以上详细描述的，机器学习单元接收输入向量并且选择主导令牌。然后，机器学习单元接收与额外的令牌有关的数据，并且将额外的令牌与所选择的主导令牌相比较，并且进一步选择新的主导令牌。该过程对于剩余令牌迭代性地继续，直到用尽所有数据并且选择了最终主导令牌为止。

### 用于广告应用的行为瞄准系统

在一个实施例中，行为瞄准系统被用于基于用户对广告类别或主题



的兴趣来向用户提供在线广告。在线广告被用于营销多种产品和服务。一般地，在线广告被用于在潜在顾客间构建品牌知名度并且促进对产品和服务的在线购买。一类在线广告被称为品牌营销，用于向目标受众推广品牌。从而，品牌营销的目的在于增大广告主的品牌对于顾客来说的知名度。在线广告主的另一目的在于诱发来自顾客的动作或响应。这类广告被称为直接响应广告。在诸如网页之类的消息上显示的、包括将用户定向到广告主网站的链接的广告是直接响应广告的示例。

横幅广告和赞助列表广告是当前使用的两类在线广告。一般地，横幅广告指的是在诸如网页之类的在线消息的预定位置处显示的一类广告。例如，横幅广告可以被显示在网页顶部的水平矩形内，但横幅广告也可出现在整个网页中的任何位置处。横幅广告可以包括动态的或静态的图形图像和/或文本。通常，横幅广告包括 URL 链接，以便如果用户点击该链接，用户的浏览器会被重定向到与横幅广告相关联的新位置。

赞助列表广告是响应于用户的搜索标准、用户正在访问的页面或者用户的简档而显示的广告。例如，如果用户在基于 web 的搜索引擎中输入搜索查询，则搜索引擎可以响应于该查询显示与搜索查询有关的广告主的超链接文本列表。赞助列表广告可以采取文本和/或图像的形式。本发明的行为瞄准系统是结合在线广告来描述的。虽然本发明是结合横幅广告和赞助列表广告的使用来描述的，但是任何类型的被包括在消息上并且通过网络分发的广告都可以被使用，而不脱离本发明的精神或范围。例如，本发明可以应用来将赞助列表与用户行为简档和网页内容相匹配。本发明的教导可应用于任何类型的在线广告，包括但不限于横幅广告、赞助列表广告、保证印象广告和基于性能的广告。广告本身可以包括任何类型的媒体，其中包括文本、图像、音频或视频。

图 8 是示出用于利用行为瞄准系统来提供广告的一个实施例的框图。对于此示例性实施例，示例性的基础设施包括网络 830、web 服务器 840、广告 (ad) 服务器 850 以及用户 810 的计算机 820。web 服务器 (840) 基于用户的行为 (例如事件) 接收信息。用户行为信息被输入到行为瞄准处理 (860)。如图 8 所示，行为瞄准处理 860 利用用户行为信息逐类别地

生成原始用户得分（870）。用户得分被输入到 ad 服务器（850），该服务器进而向 web 服务器 840 提供广告。结果，定制 ad 在计算机 820 上被提供给用户 810。

### 行为瞄准系统中的维度处理

图 9 是示出用于行为瞄准系统中的维度处理的一个实施例的框图。对于此实施例，维度处理涉及新近度处理（950）、强度处理（955）和频率处理（960）。模型和规则（970）接收类别和事件类型信息，并且基于类别和事件类型为新近度、强度和频率处理模块生成参数（例如权重）。在一个实施例中，模型使用线性回归。进行新近度、强度和频率处理的具体模式在下文中更详细描述。

在一个实施例中，模型和规则（970）为每种类别包括一个模型。从而，维度处理的权重参数是在每类别基础上的。模型定义权重参数，其中包括衰减参数。权重参数也是特定于事件类型的。在一个实施例中，事件的权重参数没有被同等加权。此外，某一事件类型（例如页面查看、搜索等等）内的所有事件没有被同等加权。从而，在针对不同事件类型的一个类别内或者在针对同一事件类型的多个类别中，所有的权重和衰减参数不一定是恒定的。

#### a. 长期用户行为简档：

在一个实施例中，本发明的行为瞄准处理生成长期用户行为简档。长期用户行为简档对用户在对较长的一段时间中（例如在至少 24 个小时中）的活动进行汇编以生成用户行为简档。在一个实施例中，长期用户行为简档用于预测用户在下一日内的兴趣。例如，长期用户行为简档可用于确定在下一日中向用户提供什么广告。

长期用户行为简档可应用来用于提供引起用户的长期兴趣的广告。例如，用户可能在几个月的一段时间中表现出购买汽车的兴趣。可以从由在线学习和获取关于汽车的信息的用户产生的用户活动（例如，对汽车站点的页面查看、对汽车广告的广告点击，等等）来汇编兴趣得分。对于此示

例，用户在汽车类别中的长期直接响应得分较高，因此广告应用可以向用户提供汽车销售广告。

以下式子给出用于计算在时间  $t_n$  用户的长期用户兴趣得分的公式：

$$s_{t_n} = intercept + \sum_{event} w_{event} Decay(\{Satu(A_{event,t}) | t_0 \leq t \leq t_n\}) + \sum_{event} r_{event} R_{event,t_n} + \sum_{event} f_{event} F_{event,t_n}$$

其中：

$s_{t_n}$  表示用户的长期直接响应（原始）得分 — 基于到时间间隔  $t_n$  为止记录的用户过去的行为数据的用户点击倾向；

$A_{event,t}$  是在某一类别中，对于某一用户，在第  $t$  日事件类型  $event$  的活动数目；

$R_{event,t_n}$  表示基于下述一日的某一用户的事件类型  $event$  的新近度信息：该日是过去对于该事件类型该用户有活动的最近一日；

$r_{event}$  表示事件类型  $event$  的新近度信息权重，其定义了事件类型就预测能力而言随时间“衰减”的速度。例如， $r_{Ad\_Click}$  定义了基于用户有广告点击活动的最近一日来预测广告点击倾向的能力；

$w_{event}$  表示事件类型  $event$  的权重，其给出了利用强度信息来预测广告点击倾向的能力；

$F_{event,t_n}$  表示基于下述一日的某一用户的事件类型  $event$  的频率信息：该日是过去对于该事件类型该用户有活动的最近一日；

$f_{event}$  表示事件类型  $event$  的频率信息的权重。

对于每个类别，用户得分  $s_{t_n}$  是利用饱和函数  $Satu()$  进行变换并且利用衰减函数  $Decay()$  进行聚集的事件活动  $A_{event,t}$  的强度的线性组合，并且使用过去的事件活动的新近度  $R_{event,t_n}$ 。事件活动的强度和新近度分别由  $w_{event}$  和  $r_{event}$  加权。

图 10 是示出用于生成长期直接响应用户兴趣得分的一个实施例的框图。对于此实施例，长期行为瞄准系统以指定的时间间隔记录用户事件  $A_{event,t}$ （图 10，块 1020）。在一个实施例中，长期行为瞄准系统记录 24 小时时段中的事件。长期行为瞄准系统对事件分类（图 10，块 1020）。长期行为瞄准系统从日志中选择经分类的事件以便处理（图 10，块 1025）。与长期直接营销简档相对应的模型被选择。模型一部分包括用于

维度处理的多个权重（例如，新近度、强度和频率）。

向事件应用饱和函数（图 10，块 1035）。在一个实施例中，饱和函数  $Satu()$  被应用到所有每日事件活动。它是一个简单的上盖（upper cap）函数。 $U_{event}$  表示事件类型  $event$  的上盖。它是过去 7 日中在最重的 0.5% 用户中的最低  $A_{event,t}$ 。

$$Satu(A_{event,t}) = \begin{cases} A_{event,t} & \text{如果 } A_{event,t} \leq U_{event} \\ U_{event} & \text{其他} \end{cases}$$

向饱和函数的输出应用衰减函数（图 10，块 1040）。在一个实施例中，衰减函数  $Decay()$  被应用来聚集用户的每日事件活动，并且表达为：

$$Decay(\{A_{event,t} | t_0 \leq t \leq t_n\}) = \sum_{t=t_0}^{t_n} \alpha^{t_n-t} A_{event,t} \quad \alpha \in (0,1]$$

$\alpha$  是用于所有事件类型和所有类别的常数。 $\alpha$  的值越小，历史数据就越快被淘汰。可以基于系统约束和业务要求来选择  $\alpha$ 。 $\alpha$  的不同值的衰减速度的示例在以下表 1 中示出。在一个实施例中， $\alpha$  被设定在 0.95。

表 1

日 / 衰减	0.998	0.98	0.97	0.96	0.95	0.94	0.93
1 日	100%	98%	97%	96%	95%	94%	93%
2 日	100%	96%	94%	92%	90%	88%	86%
3 日	99%	94%	91%	88%	86%	83%	80%
4 日	99%	92%	89%	85%	81%	78%	75%
5 日	99%	90%	86%	82%	77%	73%	70%
6 日	99%	89%	83%	78%	74%	69%	65%
1 星期	99%	87%	81%	75%	70%	65%	60%
1 月	95%	57%	43%	32%	24%	18%	13%
2 月	89%	32%	18%	10%	6%	3%	2%
3 月	85%	18%	8%	3%	1%	1%	0%

然后，向衰减函数的输出应用强度权重（图 10，块 1050）。在一个实施例中，通过建模（例如回归分析）来拟合饱和的强度特征  $w_{event}$  的权重。权重与标准误差和性能度量一起被建模系统输出。

为所记录的事件生成新近度参数（图 10，块 1055）。 $R_{event,t_n}$  是到时间  $t_n$  为止，在用户的活动中，用户有该类别中的类型  $event$  的活动的最近一日。例如，如果在到  $t_n$  为止加载的所有页面查看活动中，用户在汽车类别

中的最后页面查看发生在 2 日之前，则对于汽车来说用户在页面查看方面的新近度是 2。如果没有为用户记录页面查看活动，则新近度被设定到常数  $R_{UNKNOWN}$ 。在一个实施例中， $R_{UNKNOWN}$  被设定到 90 日。

$$R_{event,t_n} = \begin{cases} t_{current} - \max\{t \mid A_{event,t} > 0, t_0 \leq t \leq t_n\} & \text{如果 } \exists A_{event,t} > 0 \\ R_{unknown} & \text{其他} \end{cases}$$

接下来，向新近度函数应用基于事件类型和事件的类别的新近度权重  $r_{event}$ （图 10，块 1060）。在一个实施例中，通过建模（例如回归分析）来拟合饱和的强度特征  $r_{event}$  的权重。权重与标准误差和性能度量一起被建模系统输出。

如图 10 中的循环所示，行为瞄准系统通过执行块 1025、1030、1035、1040、1050、1055 和 1060 为每个类别处理事件。

在一个实施例中，可以简化用于生成长期用户兴趣得分的表达式。令  $I_{t_n}$  表示从第  $t_0$  日到第  $t_n$  日在所有事件类型上用户活动的强度维度。如果  $\alpha$  对于所有事件类型都是相同的，则有

$$I_{t_n} = \sum_{t=t_0}^{t_n} \alpha^{t_n-t} \sum_{event} w_{event} \text{Satu}(A_{event,t})$$

然后有到  $t_n$  为止记录的活动的用户  $u$  的用户得分可以改写为：

$$s_{t_n} = I_{t_n} + \sum_{event} r_{event} R_{event,t_n}$$

在一些实施例中，对于特定类别，用户的长期得分可以被递增地更新，而不存储特定类别中用户的每日得分。因此，可以随着时间的过去（以预定的更新间隔）更新得分，以反映从用户计分的第一日开始用户的过去类别兴趣，而无需重新处理用来计算原始得分的所有先前事件和计算。

如果假定  $t_k$  表示最后为用户记录事件活动的那一日，并且递增更新间隔是每  $d$  日，则  $t_n = t_k + d$ 。例如，对于每日递增更新， $t_n = t_k + 1$ 。在一些实施例中，随后可以利用以下式子来确定更新后的得分  $s_{t_n}$ ：

$$s_{t_n} = \sum_{t=t_k+1}^{t_n} \alpha^{t_n-t} \sum_{event} w_{event} \text{Satu}(A_{event,t}) + \alpha^{t_n-t_k} I_{t_k} + \sum_{event} r_{event} R_{event,t_n}$$

$$R_{event,t_n} = \begin{cases} t_n - \max\{t \mid A_{event,t} > 0, t_k < t \leq t_n\} & \text{如果 } \exists A_{event,t} > 0, t_k < t \leq t_n \\ R_{event,t_k} + t_n - t_k & \text{其他} \end{cases}$$

### b. 短期用户行为简档:

在另一个实施例中，行为瞄准系统生成短期用户行为简档。在一个实施例中，短期用户行为简档用于在下一小时内向用户提供广告。短期用户行为简档可以用来提供引起用户的超短期购买意图的广告。例如，用户可以在过去的几分钟里表现出在线购买花卉的兴趣。通常，诸如花卉这样的小型购买通常是通过相对较短的一段时间内的购买来进行的。对于此示例，行为瞄准系统实时累积用户活动，并且在较短的一段时间内向用户提供广告。

以下式子给出了每小时间隔  $t_n$  的用户短期直接响应得分的公式：

$$s_{i_n} = intercept + \sum_{event} w_{event} Decay(\{Satu(A_{event,t}) | t_0 \leq t \leq t_n\}) + rR_{i_n} + r'R'_{i_n}$$

其中：

$s_{i_n}$  表示每小时间隔的用户短期直接响应（原始）得分，其基于在  $t_n$  更新的过去行为数据来衡量用户的点击倾向；

$A_{event,t}$  是在某一类别中，对于某一用户，在每小时间隔  $t$  中事件类型  $event$  的活动数目；

$R_{i_n}$  表示基于用户有活动的每日间隔的用户新近度信息，其中排除了广告类别查看事件类型；

$R'_{i_n}$  表示基于用户有活动的每小时日内间隔的用户新近度信息，其中排除了广告类别查看事件类型；

$r, r'$  表示新近度信息的权重，其定义了活动就预测能力而言随时间“衰减”的速度。例如， $r_{Ad\_Click}$  定义了基于过去用户有点击活动的最近时间间隔来预测未来广告点击的能力；并且

$w_{event}$  表示事件类型  $event$  的权重，其给出了利用强度信息来预测广告点击倾向的能力。

对于每个类别，用户得分  $s_{i_n}$  是（利用饱和函数  $Satu()$  进行变换并且利用衰减函数  $Decay()$  进行聚集的）事件活动  $A_{event,t}$  的强度和活动  $R_{i_n}$ 、 $R'_{i_n}$  的新近度的线性组合。活动的强度和新近度分别由  $w_{event}$  和  $r, r'$  加权。

图 11 是示出用于生成短期直接响应用户兴趣得分的一个实施例的框

图。对于此实施例，短期行为瞄准以指定的时间间隔实时接收用户事件  $A_{event,t}$  (图 11, 块 1110)。短期行为瞄准系统对事件分类 (图 11, 块 1120)。选择与短期直接营销/用户目标相对应的模型 (图 11, 块 1130)。模型一部分包括用于维度处理的多个权重 (例如, 新近度、强度和频率)。

向事件应用饱和函数 (图 11, 块 1135)。在一个实施例中, 饱和函数  $Satu()$  被应用到所有每小时事件活动。它是一个简单的上盖函数。  $U_{event}$  表示事件类型  $event$  的上盖。它是在每小时的间隔期间在最重的 0.5% 用户中的最低  $A_{event,t}$ 。

$$Satu(A_{event,t}) = \begin{cases} A_{event,t} & \text{如果 } A_{event,t} \leq U_{event} \\ U_{event} & \text{其他} \end{cases}$$

向饱和函数的输出应用衰减函数 (图 11, 块 1140)。在一个实施例中, 衰减函数  $Decay()$  被应用来随着时间过去聚集用户的事件活动, 并且表达为:

$$Decay(\{A_{event,t} | t_0 \leq t \leq t_n\}) = \sum_{t=t_0}^{t_n} \alpha^{t_n-t} A_{event,t} \quad \alpha \in (0,1]$$

$\alpha$  是用于所有类型 and 所有类别的常数。 $\alpha$  的值越小, 历史数据就越快被淘汰。可以基于系统约束和业务要求来选择  $\alpha$ 。 $\alpha$  的不同值的每小时和每日衰减速度的示例在以下表 1 中示出。在一个实施例中,  $\alpha$  被设定在 0.998。

表 2

小时 / 衰减	0.999	0.998	0.997	0.996	0.970	0.960	0.950
1 小时	100%	100%	100%	100%	97%	96%	95%
12 小时	99%	98%	96%	95%	69%	61%	54%
1 日	98%	95%	93%	91%	48%	38%	29%
2 日	95%	91%	87%	82%	23%	14%	9%
3 日	93%	87%	81%	75%	11%	5%	2%
4 日	91%	83%	75%	68%	5%	2%	1%
5 日	89%	79%	70%	62%	3%	1%	0%
6 日	87%	75%	65%	56%	1%	0%	0%
1 星期	85%	71%	60%	51%	1%	0%	0%
2 星期	71%	51%	36%	26%	0%	0%	0%
3 星期	60%	36%	22%	13%	0%	0%	0%
1 月	51%	26%	13%	7%	0%	0%	0%
2 月	26%	7%	2%	0%	0%	0%	0%

3月	13%	2%	0%	0%	0%	0%	0%
日 / 衰减	<b>0.998</b>	<b>0.98</b>	<b>0.97</b>	<b>0.96</b>	<b>0.95</b>	<b>0.94</b>	<b>0.93</b>
1日	100%	98%	97%	96%	95%	94%	93%
2日	100%	96%	94%	92%	90%	88%	86%
3日	99%	94%	91%	88%	86%	83%	80%
4日	99%	92%	89%	85%	81%	78%	75%
5日	99%	90%	86%	82%	77%	73%	70%
6日	99%	89%	83%	78%	74%	69%	65%
1星期	99%	87%	81%	75%	70%	65%	60%
1月	95%	57%	43%	32%	24%	18%	13%
2月	89%	32%	18%	10%	6%	3%	2%
3月	85%	18%	8%	3%	1%	1%	0%

然后，向衰减函数的输出应用强度权重（图 11，块 1150）。在一个实施例中，通过建模（例如回归分析）来拟合饱和的强度特征  $w_{event}$  的权重。权重与标准误差和性能度量一起被建模系统输出。

为事件生成新近度参数（图 11，块 1155）。在一个实施例中，为了简单，使用单个新近度值。 $R_{t_n}$  是在  $t_n$  最后更新的用户活动中，用户有该类别中的活动的最近的每小时间隔。例如，如果用户在两小时之前有在汽车类别中的页面查看，并且在两小时内没有其他活动，则对于汽车来说用户在页面查看方面的新近度是 2。如果从最后更新起用户没有活动，则新近度  $R_{t_n}$  被设定到常数  $R_{UNKNOWN}$ 。在一个实施例中， $R_{UNKNOWN}$  被设定到 504。

$$R_{t_n} = \begin{cases} t_{current} - \max\{t \mid A_{event,t} > 0, t_0 \leq t \leq t_n\} & \text{如果 } \exists A_{event,t} > 0 \cap event \neq adcat\_views \\ R_{unknown} & \text{其他} \end{cases}$$

从  $R_{t_n}$  导出  $R'_{t_n}$ ，其将  $R_{t_n}$  变换成 V 形日内模式：

$$R'_{t_n} = \begin{cases} |R_{t_n} \bmod 24 - 12| + 1 & R_{t_n} \neq R_{unknown} \\ 0 & \text{其他} \end{cases}$$

接下来，向新近度函数应用基于事件类型和事件的类别的新近度特征  $r, r'$ （图 11，块 1160）。在一个实施例中，通过建模来拟合强度特征  $r, r'$  的权重。权重与标准误差和性能度量一起被建模系统输出。

如图 11 中的循环所示，行为瞄准系统通过执行块 1130、1135、1140、1150、1160 和 1170 为每个类别处理事件。

也可以实时地递增更新针对特定类别用户的短期得分。可以在不确定



或存储特定类别中用户的事件活动的计数  $A_{event,t}$ （每小时间隔  $t$  的事件类型  $event$  的活动数目）的情况下执行得分的实时递增更新。得分的实时递增更新允许了基于（事件类型  $event$  的）新近事件来调节得分，而无需重新处理用来计算原始得分的所有先前事件和计算。

如上所述， $I_{t_n}$  表示得分的强度分量（即，从第  $t_0$  日到第  $t_n$  日在所有事件类型上用户活动的强度维度），其中：

$$I_{t_n} = \sum_{t=t_0}^{t_n} \alpha^{t_n-t} \sum_{event} w_{event} \text{Satu}(A_{event,t})$$

同样如上所述，有到  $t_n$  为止记录的所有活动的用户的得分可以简化为：

$$s_{t_n} = I_{t_n} + \sum_{event} r_{event} R_{event,t_n}$$

如果强度分量被初始化为零（ $I_{t_0}=0$ ）并且最后用户活动的每小时间隔被设定到未知（ $R_{t_0}=R_{unknown}$ ），则初始用户得分  $s_{t_0}$  被计算为  $s_{t_0}=R_{UNKNOWN}$ 。另外，假定  $s_{t_{n-1}}$  表示在每小时间隔  $t_{n-1}$  最后更新的用户得分。当在每小时间隔  $t_n$  接收到事件类型  $event$  的新近活动时，可以将用户得分  $s_{t_{n-1}}$  递增地更新到用户得分  $s_{t_n}$ ，而无需重新处理在特定类别中用户的所有过去的事件活动  $A_{event,t}$ 。在一些实施例中，递增更新的用户得分  $s_{t_n}$  是利用以下式子来确定的：

$$s_{t_n} = \alpha^{t_n-t_{n-1}} I_{t_{n-1}} + w_{event} + r R_{t_n} + r' R'_{t_n}$$

$$R_{t_n} = \begin{cases} R_{t_{n-1}} & \text{如果 } event = adcat\_views \\ R_{t_{n-1}} - t_n + t_{n-1} & \text{其他} \end{cases}$$

$$R'_{t_n} = \begin{cases} |R_{t_n} \bmod 24 - 12| + 1 & R_{t_n} \neq R_{unknown} \\ 0 & \text{其他} \end{cases}$$

## 建模过程

如上所述，模型包括用于应用权重以生成用户简档得分的权重参数。在一个实施例中，权重参数是从用户数据集合生成的。从用户过去的活动汇编的用户数据集合将事件信息与用户行为相互关联起来（例如，点通率与用户事件相关联）。可以分析用户数据集合以识别正面用户（例如，用户活动符合瞄准目标）和负面用户（例如，不符合瞄准目标的用户）。利用数据挖掘技术来分析用户数据集合，以确定什么动作（例如事件信息）

对于为瞄准目标预测行为最为有用。可以分析正面用户的事件信息以确定对瞄准目标最有贡献的事件。例如，如果瞄准目标是直接响应广告并且用户数据集合标识出一种趋势，即提交被分类在“运动”类别中的搜索查询的用户也点击直接响应广告，那么对于类别“运动”，事件类型“搜索”的权重参数被给予相对较高的值。对于此实施例，分析用户数据集合以为分类体系的每个类别中的每个事件类型生成权重。

与生成模型的这种方法相关联的一个问题是与某一类别相关联的数据可能不足或者没有数据与该类别相关联。在一个实施例中，行为瞄准系统为具有不充足数据集合的类别利用模型继承技术。利用模型继承技术，从一个类别创建的模型被用作分类体系上的有关类别或者父类别的模型。例如，分类体系可以包括位于分类体系中的父类别“金融”之下的子类别“产权投资”。如果对于类别“产权投资”来说用户数据集合不足，则“金融”类别的模型可用作“产权投资”类别的模型。模型继承技术还可用来通过生成一些模型并随后将这些模型用于有关类别来促进支持大量类别的系统中的构建模型的过程。

在一个实施例中，建模过程使用用户权重来提高模型的准确度。一般地，用户权重标识出用户数据集合中的提供最高质量的信息用以构建模型的那些用户。示例性的用户数据集合可以包括下表：

表 3

用户	广告查看	广告点击
John	100	0
Mary	1	0
Peter	100	1
Sue	1	1

此示例性数据集合表明，就点击倾向而言，John 和 Mary 是负面用户，而 Peter 和 Sue 是正面用户。但是，Mary 只进行了一次广告查看，而 John 进行了 100 次广告查看。从而，该数据提供了有关以下事实的更强置信度：比起 Mary 来，John 更是非点击者（non-clicker）。类似地，Sue 仅进行了广告查看一次就点击了，而 Peter 在点击之前已经有了 100 次广告查看。

从而，作为点击者的 Sue 的数据的置信度要高于 Peter 是点击者的置信度。在用于生成模型的一个实施例中，向来自用户数据集的数据赋予用户权重以衡量数据的质量。对于上述示例，在生成模型时，比起与 Peter 相关联的数据来，与 Sue 相关联的数据被指派以更高的用户权重。

在一个实施例中，基于就瞄准目标而言用户的预期表现（例如点通率）来向用户指派用户权重。在数据挖掘算法中使用的目标变量被重新调节，并且正面和负面用户的权重被正规化以使其和为 1。具体而言，目标变量  $t_u$  可以如下被设定到 1 或 0：

如果  $(adclick - \overline{CTR} * f(adview)) > 0$ ,

$$\text{则 } t_u = 1$$

否则，

$$t_u = 0$$

在一个实施例中，

$$f(adview) = adview + k \text{ (其中 } k \text{ 是常数)}。$$

在其他实施例中，

$$f(adview) = adview$$

$$f(adview) = (1 - e^{-k * adview})$$

$$f(adview) = \text{bin}(adview)$$

然后用户加权计算为：

$$u_u = |adclick - \overline{CTR} * f(adview)|$$

利用目标变量和用户权重，机器学习算法根据以下关系来最小化预测误差：

$$\text{error} = \sum_u (t_u - \tilde{t}_u)^2 * u_u$$

其中， $t_u$  是预测性目标变量， $\tilde{t}_u$  是实际目标变量。

在一个实施例中，数据清洁技术被用于从用户数据集合中去除用户数据。对于此实施例，在建模过程之前排除“机器人”事件，以确保它们的行为不会使模型出现偏差。机器人事件会使建模出现错误，尤其如果存在相对于其他用户的任何计分的话。从而，从模型中排除机器人事件和具有最低限度活动的 cookie。

### 实时和分批处理实现方式：

图 12 是示出生成长期和短期用户兴趣得分的行为瞄准系统的一个实施例的框图。与上述基础设施类似，使用了示例性的用户计算机（1220）、网络 1230、web 服务器 1240、服务器 1250。在一个实施例中，数据库 1255、分批行为瞄准处理 1260 和仓库简档 1280 被用于计算长期用户兴趣得分。数据库 1255 耦合到（一个或多个）web 服务器 1240，以存储记录的事件。记录的事件随后在分批行为瞄准处理 1260 中以预定的间隔被处理。分批行为瞄准处理 1260 的输出包括针对每个类别的多个用户兴趣得分。得分被存储在仓库简档 1280 中。

行为瞄准系统 1200 还包括实时行为瞄准处理 1270 和用户数据存储 1290，用于生成短期用户兴趣得分。具体而言，在短期时间间隔（例如一个小时）中，在实时行为瞄准处理 1270 中累积事件。实时行为瞄准处理 1270 生成短期用户兴趣得分。如上所述，短期用户兴趣得分可用来为各种用户和营销目标选择广告。在另一个实施例中，用户数据存储包括长期用户兴趣得分和短期用户兴趣得分两者。对于此实施例，行为瞄准系统 1200 生成组合的长期和短期用户兴趣得分，如下所述。

### 将用户得分映射到输出度量：

对于一些应用，用户简档得分的范围被转换成跨类别的相对得分（即，相对于不同类别中的其他用户简档得分）。在一个实施例中，在每个类别中对用户简档得分进行百分比排名。例如，“金融”类别中的用户简档得分“.2”可给出所有用户中的前 1%内的用户简档得分，而“运动”类别中的用户简档得分“.6”可给出所有用户中的前 10%内的用户简档得分。对于此示例，“运动”中的用户简档得分“.6”被转换成等于 1 的相对用户简档得分。利用此技术，相对用户简档得分允许了相对于不同用户的其他兴趣比较“运动”和“金融”中的用户兴趣。在另一个实施例中，选择类别，以通过利用输出度量（例如点通率）跨所有类别对用户的得分进行排名，来最大化关于用户的 ROI。

在另一个实施例中，可以将用户简档分割成层。例如，营销者可能希

望只利用指定类别中的前 10%的用户来启动一个计划。然后，可以为另一目的对接下来的 10%的用户进行排名。用户的排名可以基于出于任何目的对用户数据的任何分类。

如上所述，产生了瞄准目标的原始得分。例如，可以通过品牌模型/规则集合来产生品牌得分，并且可以通过直接响应模型/规则集合来产生直接响应得分。可以利用长期或短期数据来产生原始得分。通过长期数据产生的原始得分被称为长期原始得分，通过短期产生的原始得分被称为短期原始得分。例如，直接响应得分可以利用长期数据和短期数据来产生。

原始得分指示出在针对某一瞄准目标的类别中用户的相对兴趣量度。但是，原始得分不直接衡量用户的动作倾向。这样，原始得分被转换或映射到用于瞄准优化目的的更有用的得分，在这里它被称为映射得分。在一些实施例中，映射得分是反映出用户进行特定动作的趋势的面向行为的得分。映射得分对于瞄准优化目的是有用的，因为它们提供了关于用户相对于感兴趣的类别的动作的预测性信息。

映射得分的一些示例是表达为百分比的点击倾向、点通率、以及也表达为比率的转换倾向。在其他实施例中，也可以产生其他映射得分。一般地，针对某一类别的用户点通率（CTR）反映了用户将要选择（“点击”）与该类别相关联的内容（例如广告、链接等等）的概率。一般地，在某一类别中用户的转换率反映了用户将会买/购买与该类别相关联的产品或服务的概率。在一些实施例中，对于特定类型的原始得分，长期和短期原始得分被转换成长期和短期映射得分。例如，长期直接响应得分（LTDR）可被转换成长期 CTR 得分（LTCTR），并且短期直接响应得分（STDOR）可被转换成短期 CTR 得分（STCTR）。

对于同一用户和类别，原始得分通常不与映射得分具有线性关系。例如，基于从多个用户聚集的统计行为数据，用户的直接响应得分不与其在相同类别中的 CTR 得分具有线性关系。图 13 示出了一曲线图，该曲线图基于从多个用户聚集的统计数据示出了作为直接响应得分的函数的 CTR 得分。如图 13 所示，在 y 轴上示出的 CTR 得分随着在 x 轴上示出的直接响应得分增大而非线性地增大。直接响应和 CTR 得分分别被下限  $s_{\min}$  和  $c_{\min}$

所限，其中  $s_{\min}$  指示出以  $c_{\min}$  作为其相应 CTR 得分的非积极用户的直接响应得分。

在一些实施例中，利用非线性多项式映射方程来将原始得分转换成映射得分。可以利用从多个用户聚集的行为数据来确定映射方程，所述行为数据将映射得分与原始得分相关联起来。在一些实施例中，该映射方程表示为：

$$\text{RawScoreToMappedScore}(s_s) = c_{\min} + k_1(s_s - s_{\min}) + k_2(s_s - s_{\min})^2 = c_s$$

其中：

$s_s$  = 原始得分值；

$s_{\min}$  = 在某一类别中，过去未表现出任何活动（页面查看、搜索、点击等等）的用户的得分；

$c_{\min}$  = 与  $s_{\min}$  相对应的点通过率。注意， $c_{\min} > 0$ ，因为即使是过去未表现出行为兴趣的用户仍可以有非零的点击倾向。

$k_1$  和  $k_2$  = 常数；并且

$c_s$  = 映射得分值。

在一些实施例中，在映射方程中使用的常数的值（ $s_{\min}$ 、 $c_{\min}$ 、 $k_1$  和  $k_2$ ）依据要映射的原始得分的类型而变化。在这些实施例中，用于产生原始得分的特定模型/规则集合计算和存储这些随后被用于映射方程中的常数值。例如，用于将品牌得分转换到映射得分的常数值可以不同于用于将直接响应得分转换到映射得分的常数值。用于品牌得分的常数值可以在品牌模型/规则集合中计算和存储，用于直接响应得分的常数值可以在品牌模型/规则集合中计算和存储。在一些实施例中，用于映射方程中的常数的值（例如  $s_{\min}$ 、 $c_{\min}$ 、 $k_1$  和  $k_2$ ）依据映射得分的类型而变化。例如，用于产生 CTR 得分的常数值可以不同于用于产生转换率得分的常数值。

如上所述，对于一类原始得分（例如，直接响应得分），可能存在长期和短期原始得分（例如 LTDR 和 STDR）。对于此实施例，长期原始得分（例如 LTDR）可以被转换成长期映射得分（例如 LTCTR），并且短期原始得分（例如 STDR）可以被转换成短期映射得分（例如 STCTR）。这些长期和短期映射得分被组合成单个组合映射得分。在一些实施例中，组合映射得分是利用组合方程来计算的，该组合方程是利用从多个用户聚集

的统计行为数据来确定的。在一些实施例中，组合方程表示为：

$$\text{CombinedMappedScore}(c_{st}, c_{lt}) = (c_{st} * c_{sl}) / (\text{EMS}) = c_c$$

其中：

$c_{st}$  = 短期映射得分值；

$c_{sl}$  = 长期映射得分值；

EMS = 多个用户中的预期映射得分；并且

$c_c$  = 组合映射得分值。

可以利用从多个用户聚集的行为数据来确定 EMS 值。在一些实施例中，EMS 值在特定类别内是恒定的，但在不同类别之间是变化的。例如，如果长期和短期映射得分是长期和短期 CTR 得分，则组合方程表示如下：

$$\text{CombinedCTRScore}(c_{st}, c_{lt}) = (c_{st} * c_{sl}) / (\text{ECTR}) = c_c$$

其中：

$c_{st}$  = 短期 CTR 倾向；

$c_{sl}$  = 长期 CTR 倾向；

ECTR = 多个用户中的预期 CTR；并且

$c_c$  = 组合 CTR 倾向。

这样，当产生长期和短期映射得分时，可以使用组合方程来提供反映长期和短期映射得分的单个组合映射得分。单个组合映射得分可用于稍后的瞄准优化处理。

组合长期和短期得分的过程是递增的。此技术允许了在无需重新处理与长期得分相关联的事件信息的情况下对长期和短期得分进行组合。这带来了处理和时间上的巨大节省，因为产生大量事件信息需要大量时间。因此，通过将长期用户简档得分与短期用户简档得分相组合，可以在以后使用被汇编以生成长期用户简档得分的事件信息。例如，事件信息可以被计算为某一时间  $time_0$  的长期用户简档得分。一天之后，即  $time_0$  加一天，可以为过去的一天计算短期用户简档得分。 $time_0$  的两天之后，即  $time_0$  加两天，可以实时计算第二短期用户简档得分。对于此示例，行为瞄准系统可以重新计算新长期得分，以便与  $time_0$  的新短期得分相组合，而无需重新处理事件信息。

如上所述，在一个实施例中，强度维度包括衰减函数。一般地，强度

参数基于所经过的时间量来衰减事件信息加权。在递增性长期和短期得分组合过程的一个实施例中，通过按计算长期得分与生成新的长期得分的当前时间之间经过的时间量增大衰减函数来生成新的长期得分。新近度维度也可以被递增地更新。在一个实施例中，新近度参数是时间的线性函数，使得新近度参数在每单位时间中被增大固定的量（例如，新近度参数可以每天递增 1 点）。从而，可以通过增大新近度参数以与新的时间相一致（即，在长期得分的计算与当前时间之间）来生成与长期得分相关联的新的新近度参数。

长期和短期得分的组合提供了更准确的行为瞄准系统。长期得分捕捉综合行为数据。但是，长期得分相对于当前行为兴趣有相对较长的滞后时间（例如一天）。短期数据实时捕捉行为数据，因此捕捉了用户的当前行为兴趣。但是，与长期得分不同，短期得分不捕捉用户行为的整体简档。从而，通过组合长期和短期得分，用户行为简档既捕捉了当前兴趣又捕捉了过去的兴趣，以提供用户兴趣的完整视图。

#### 用于行为瞄准系统的网络环境：

图 14 示出了用于本发明的行为瞄准系统的操作的网络环境 1400 的一个实施例。网络环境 1400 包括耦合到网络 1430（例如因特网、内联网、外联网、虚拟专用网、非基于 TCP/IP 的网络、任何 LAN 或 WAN，等等）的客户端系统 1420，以及服务器系统 1440<sub>1</sub> 至 1440<sub>N</sub>。服务器系统可以包括单个服务器计算机或者数个服务器计算机。客户端系统 1420 被配置为与服务器系统 1440<sub>1</sub> 至 1440<sub>N</sub> 中的任何一个通信，以例如请求和接收基本内容和额外内容（例如，以网页的形式）。

客户端系统 1420 可以包括桌面型个人计算机、工作站、膝上型计算机、PDA、蜂窝电话、任何具有无线应用协议（WAP）能力的设备、或者任何其他能够直接或间接与网络通信的设备。客户端系统 1420 通常运行 web 浏览程序，该程序允许了客户端系统 1420 的用户通过网络 1430 向服务器系统 1440<sub>1</sub> 至 1440<sub>N</sub> 请求内容并从其接收内容。客户端系统 1420 通常包括一个或多个用户接口设备 22（例如键盘、鼠标、滚动球、触摸屏、



笔，等等），用于与显示器（例如监视屏、LCD 显示器等等）上的 web 浏览器的图形用户界面（GUI）交互。

在一些实施例中，客户端系统 1420 和/或系统服务器 1440<sub>1</sub> 至 1440<sub>N</sub> 被配置为执行这里描述的方法。一些实施例的方法可以用被配置为优化对将要显示给用户的额外内容的选择的软件或硬件来实现。

图 15 示出了行为瞄准系统 1500 的概念图。行为瞄准系统 1500 包括客户端系统 1505、基本内容服务器 1510（包含基本内容）、额外内容服务器 1515（包含额外内容）、用户简档数据库 1520 以及行为瞄准服务器 1535。行为瞄准服务器 1535 包括接收事件信息的优化器模块 1537。行为瞄准系统 1500 被配置为基于用户的简档来选择将要发送给用户的额外内容。客户端系统 1505 被配置为接收基本和额外内容并将基本和额外内容显示给用户（例如，作为公布的网页）。优化系统的各个部分可以驻留在一个或多个服务器（例如服务器 1440<sub>1</sub> 至 1440<sub>N</sub>）和/或一个或多个客户端系统（例如客户端系统 1420）中。

用户简档数据库 1520 为多个用户/客户端系统存储用户简档，每个用户简档具有为用户使用的特定客户端系统 1505 指派的唯一用户标识号。用户标识号例如可以被存储在用户使用的客户端系统 1505 上的 cookie 中。当用户向基本内容服务器 1510 请求一条基本内容时，该 cookie 被从客户端系统 1505 传送到基本内容服务器 1510，然后被传送到行为瞄准服务器 1535。行为瞄准服务器 1535 随后使用 cookie 中的信息来从用户简档数据库 1520 取回特定的用户简档。

行为瞄准系统可以用硬件或软件来实现。对于软件实现方式，行为瞄准系统是包括用于在通用计算机系统上实现的多个计算机可执行指令的软件。在加载到通用计算机系统中之前，行为瞄准系统软件可以作为经编码的信息驻留在计算机可读介质上，例如磁软盘、磁带、以及致密盘只读存储器（CD-ROM）。

虽然已经按特定的示例性实施例描述了本发明，但是应当认识到，在不脱离本发明的精神和范围的情况下，本领域的技术人员可以进行各种修改和变更。

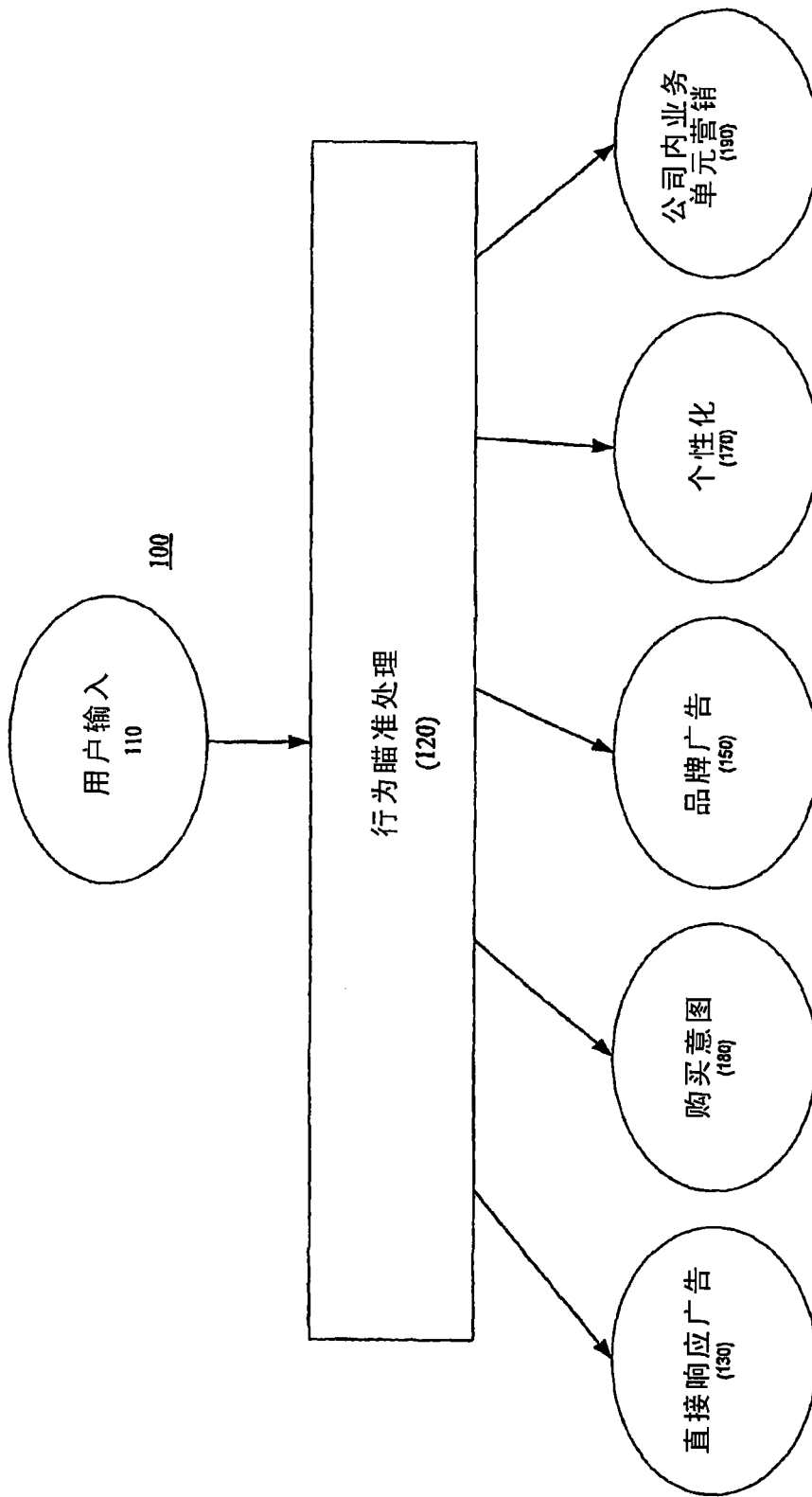


图1

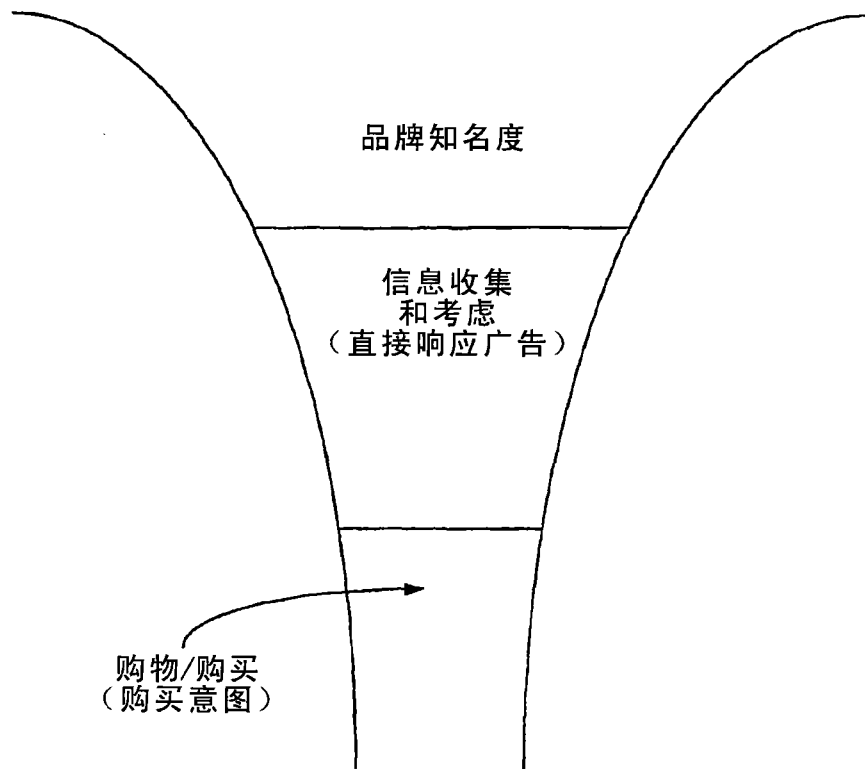


图2

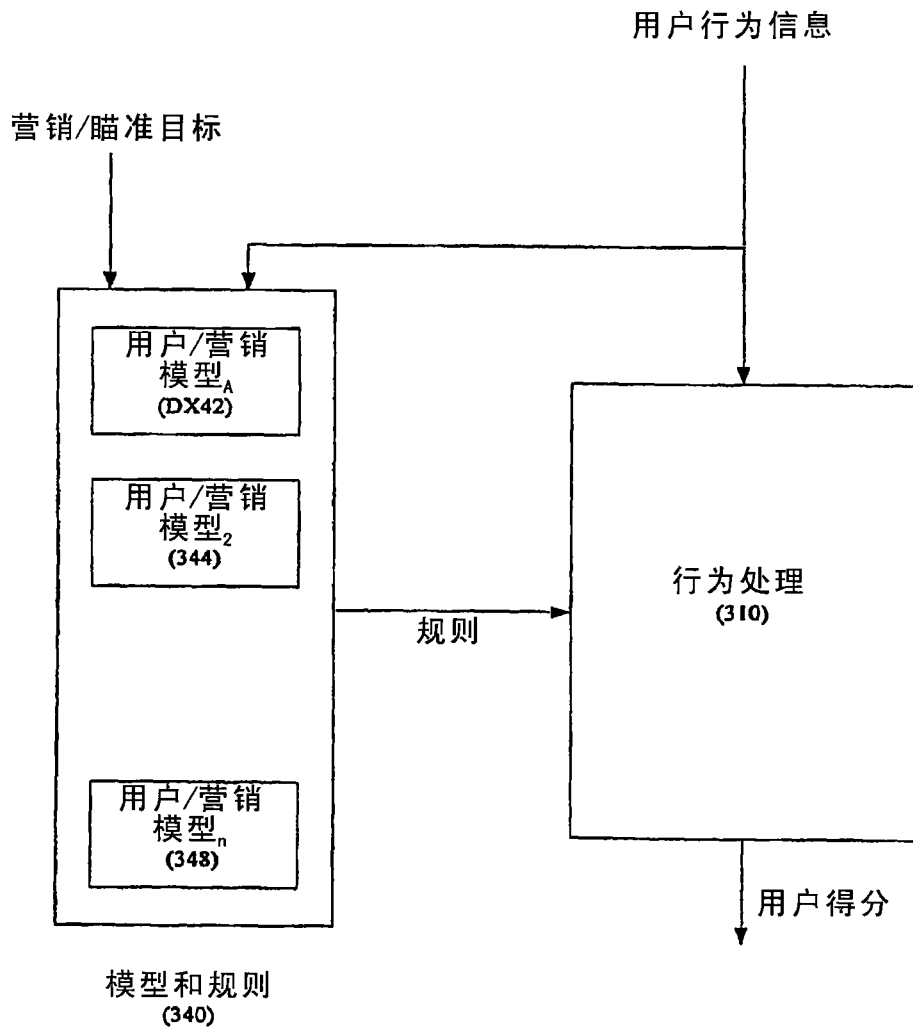


图3

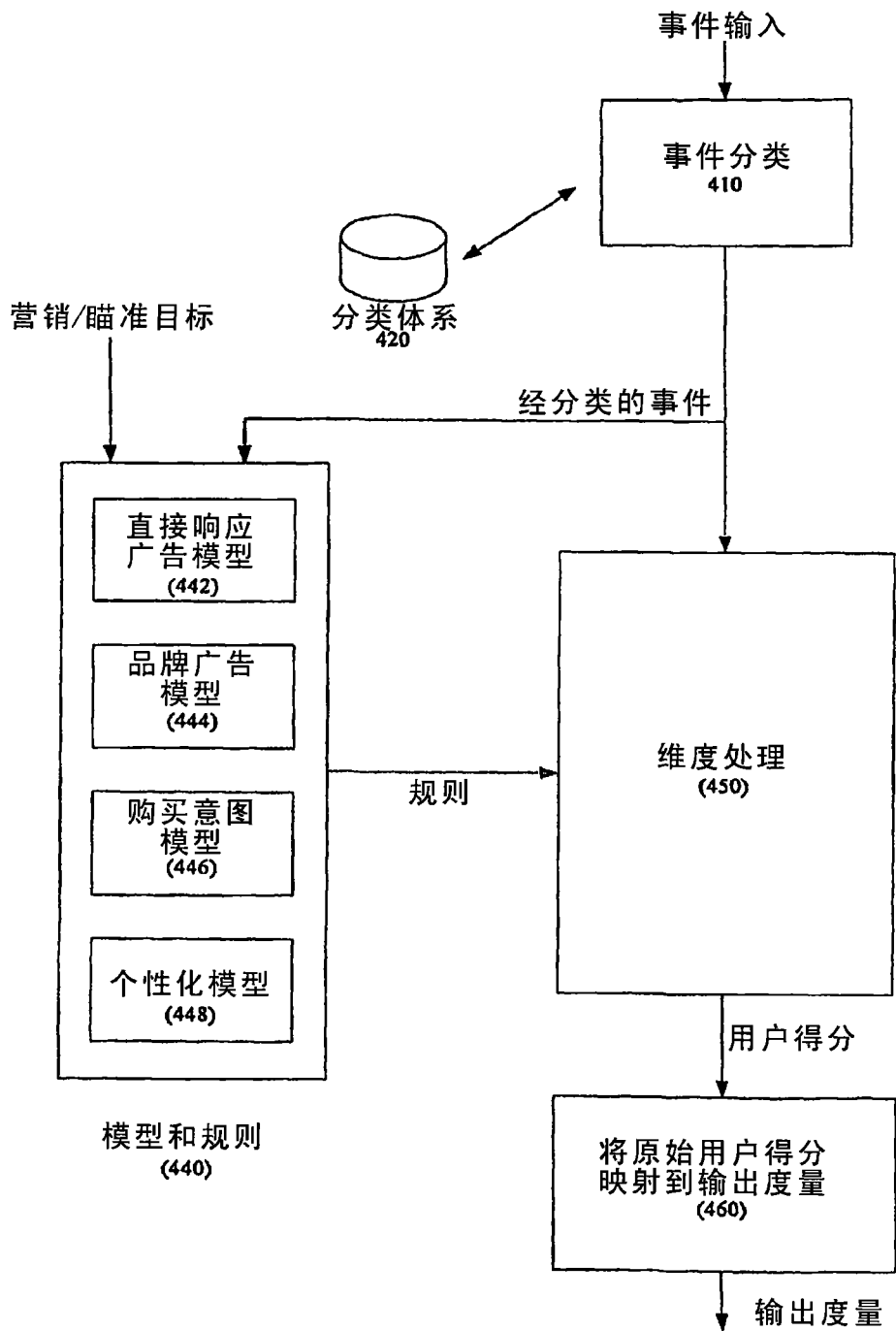


图4

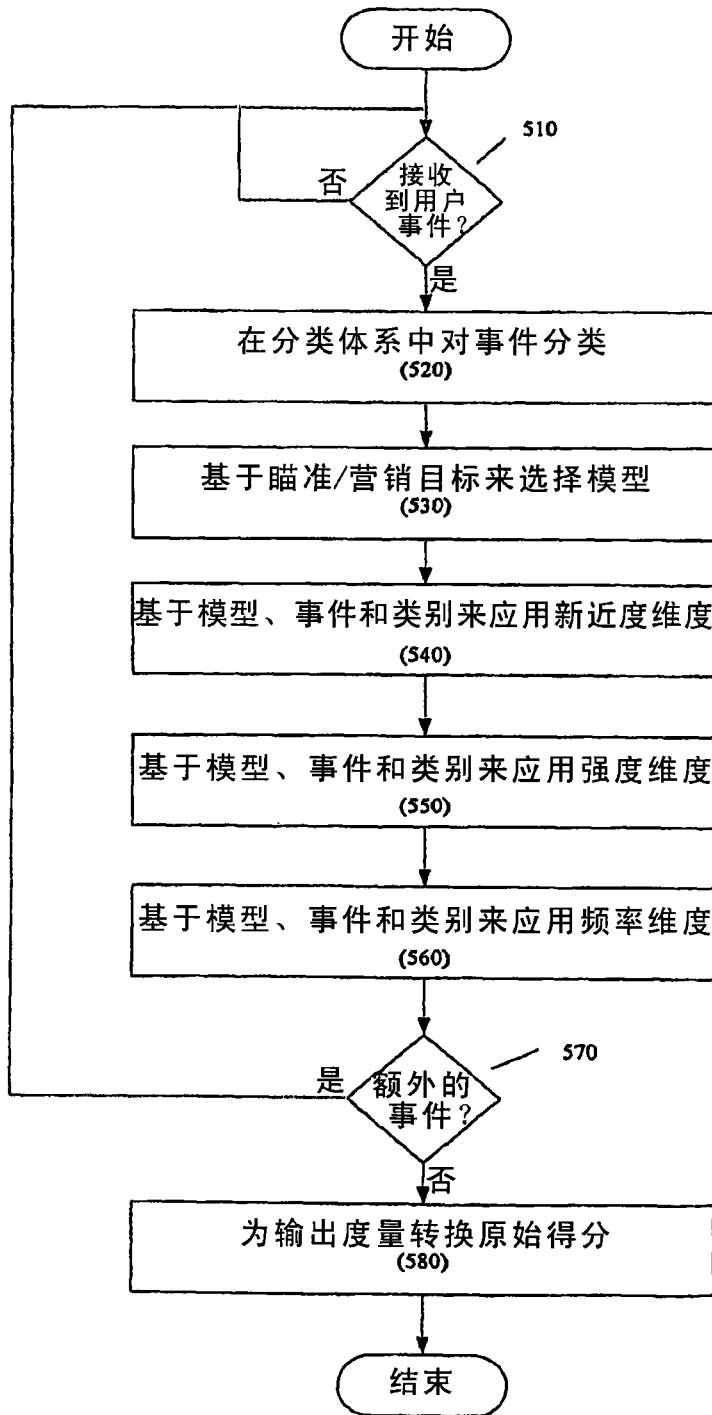


图5

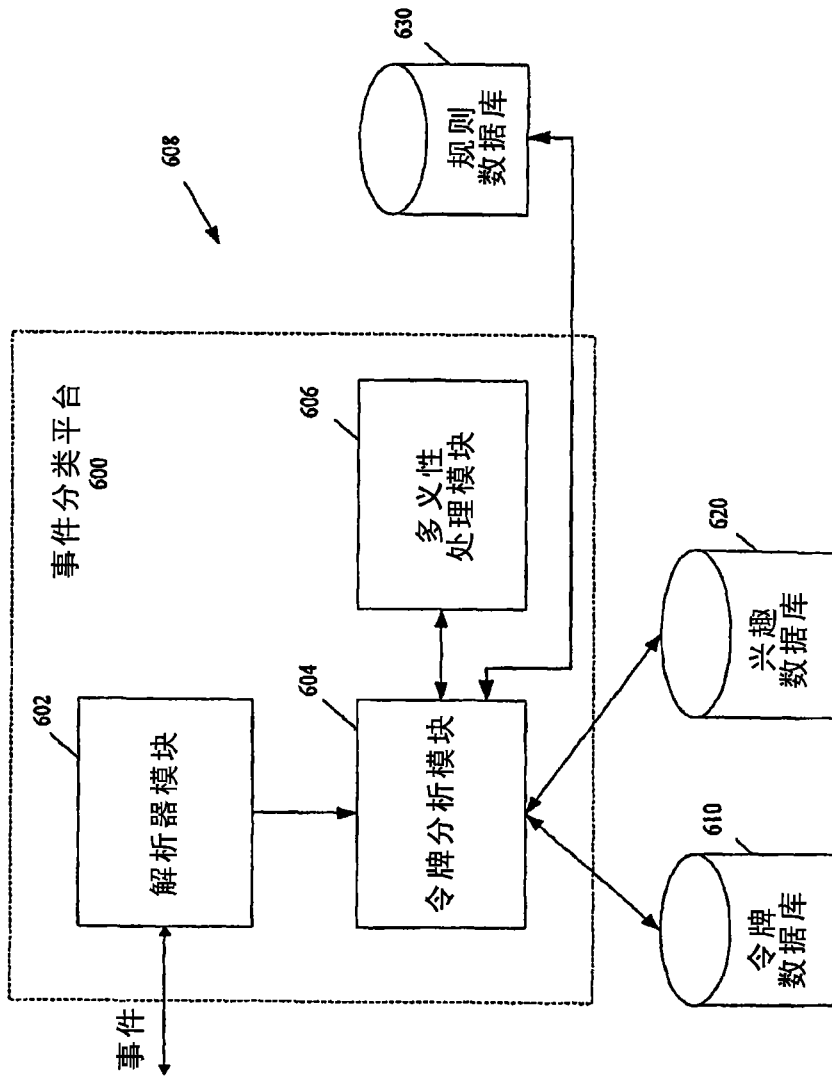


图6

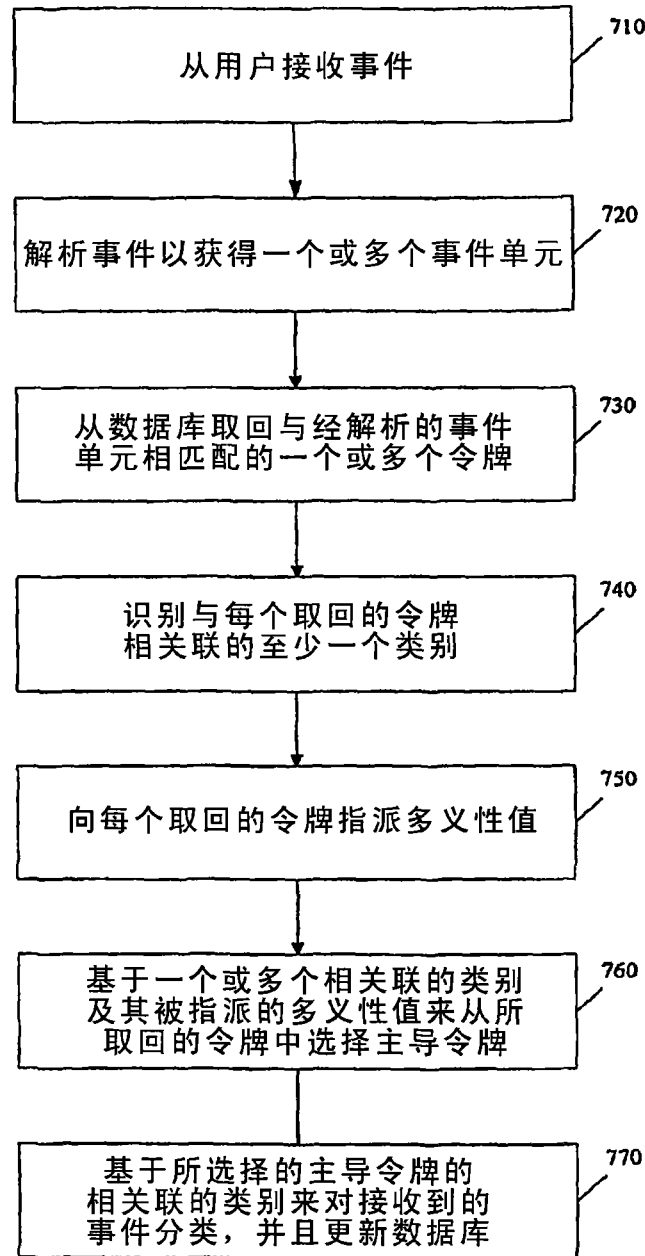


图7



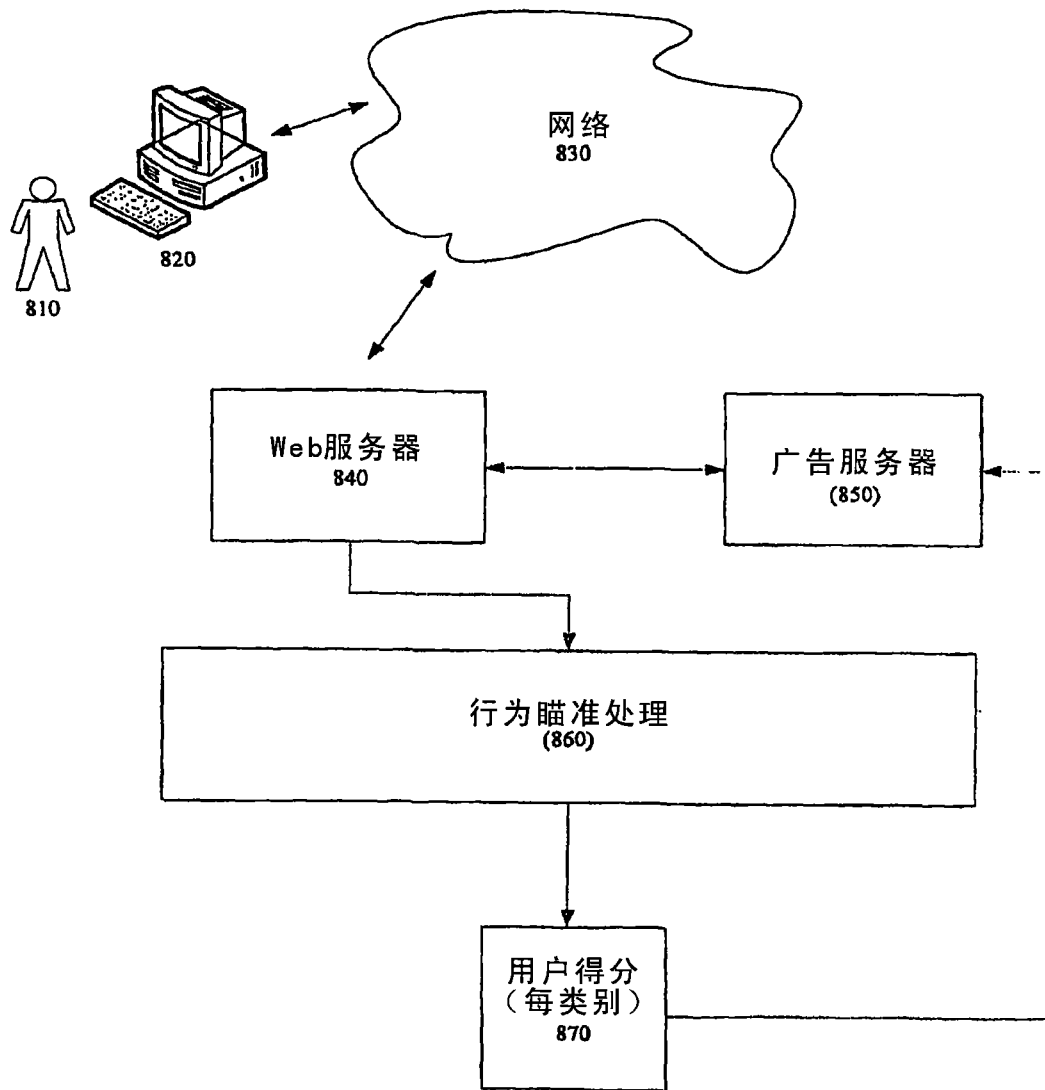


图8

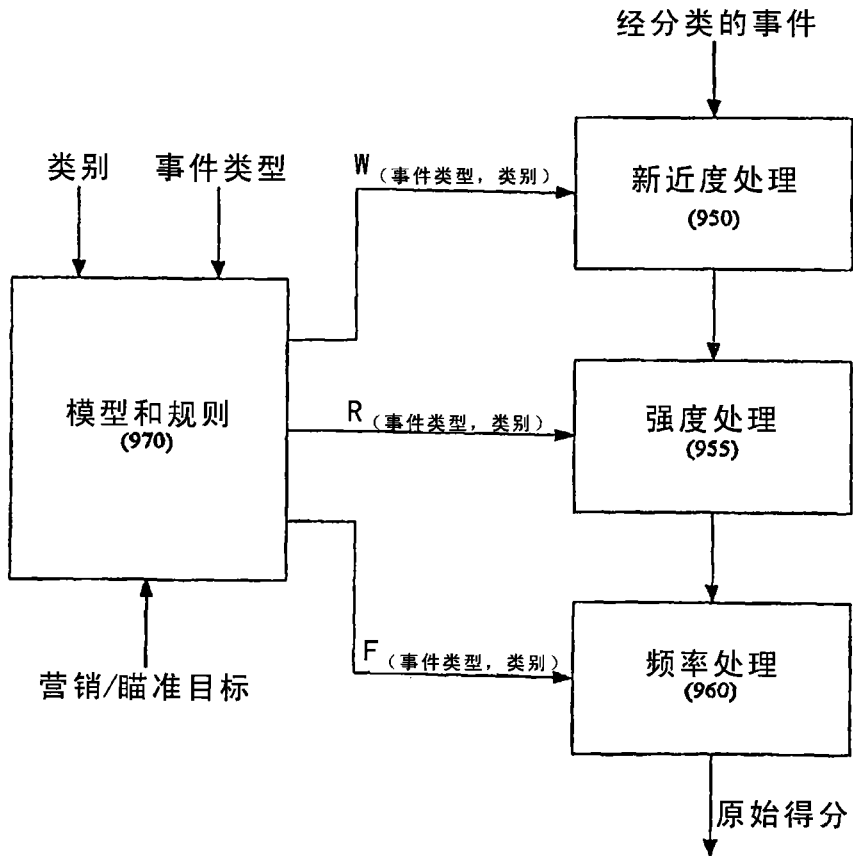


图9

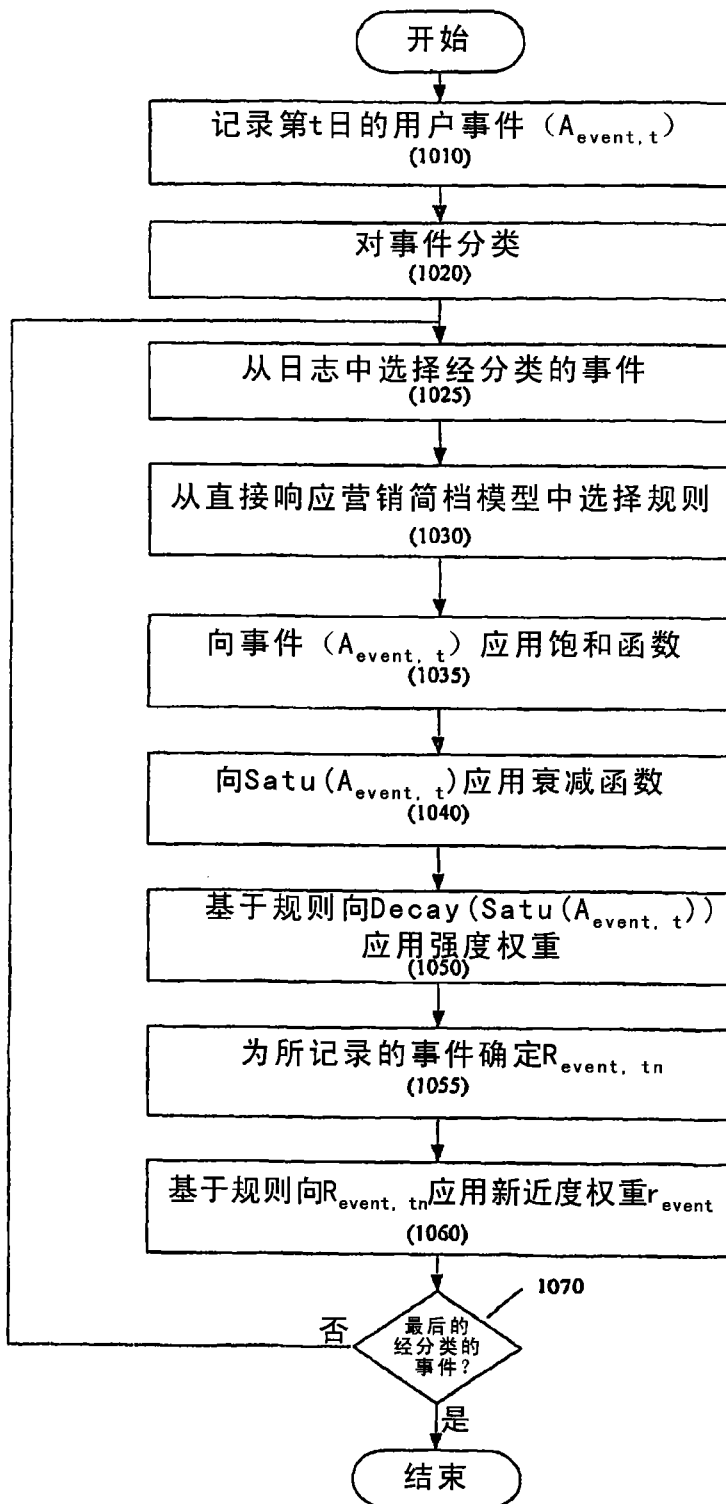


图10

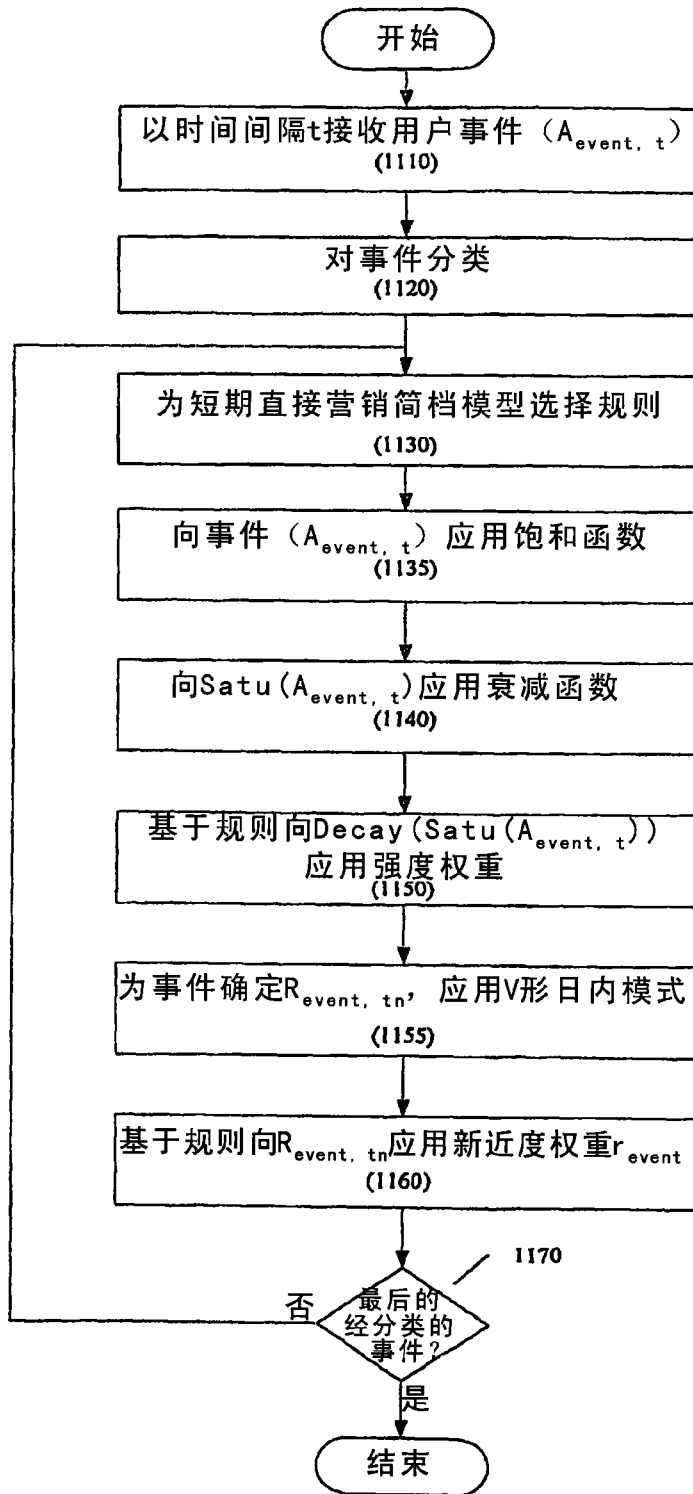


图11

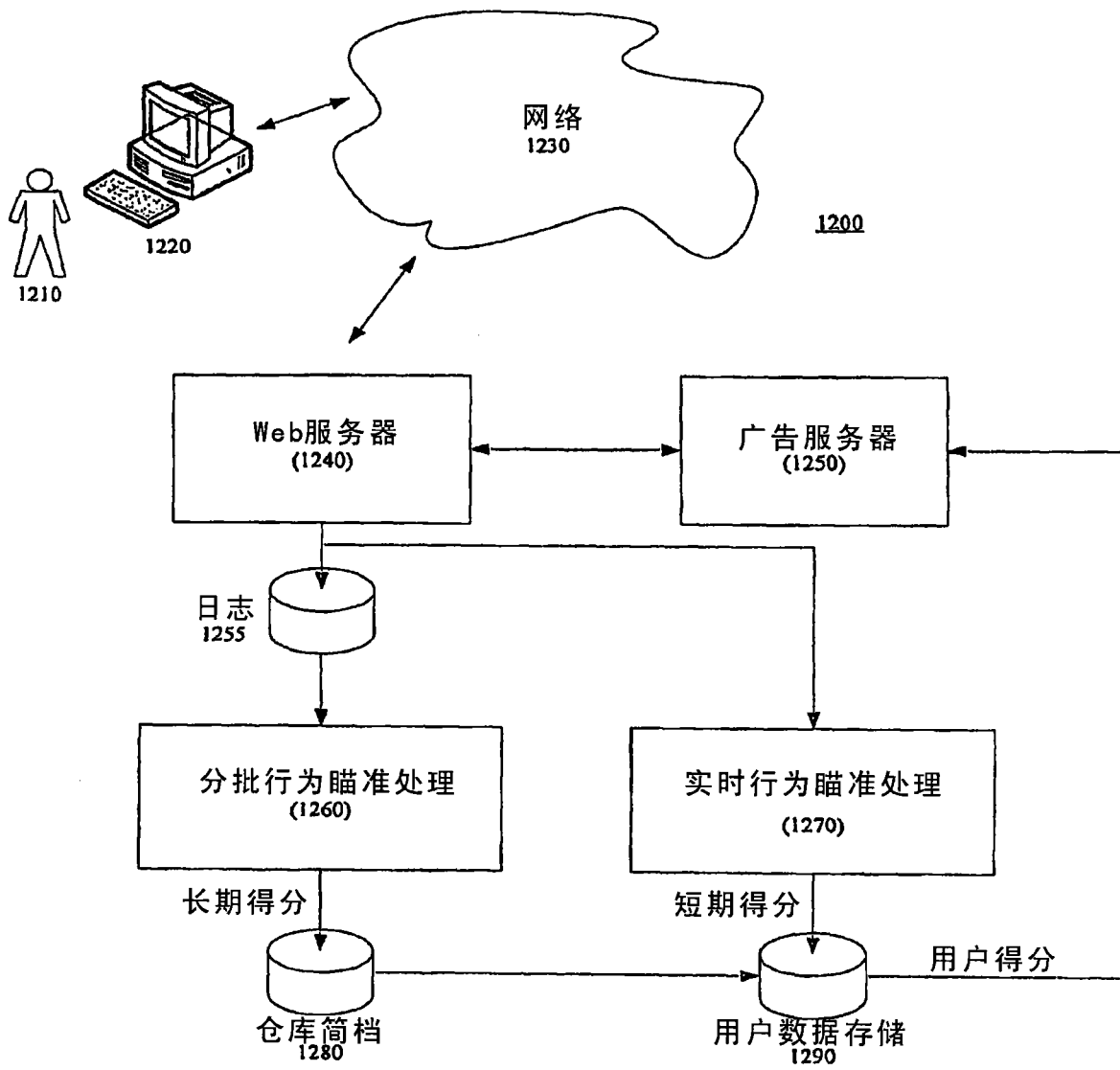


图12

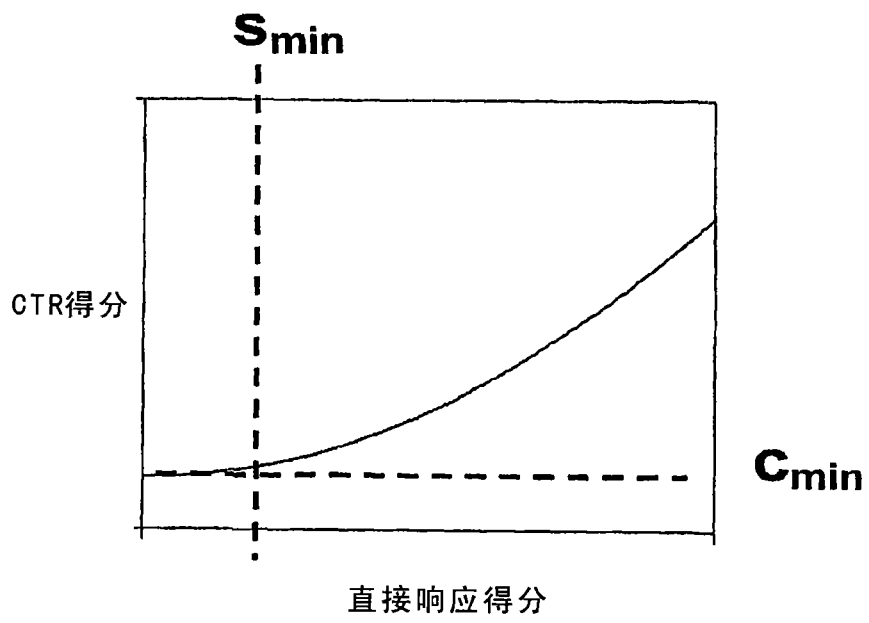


图13

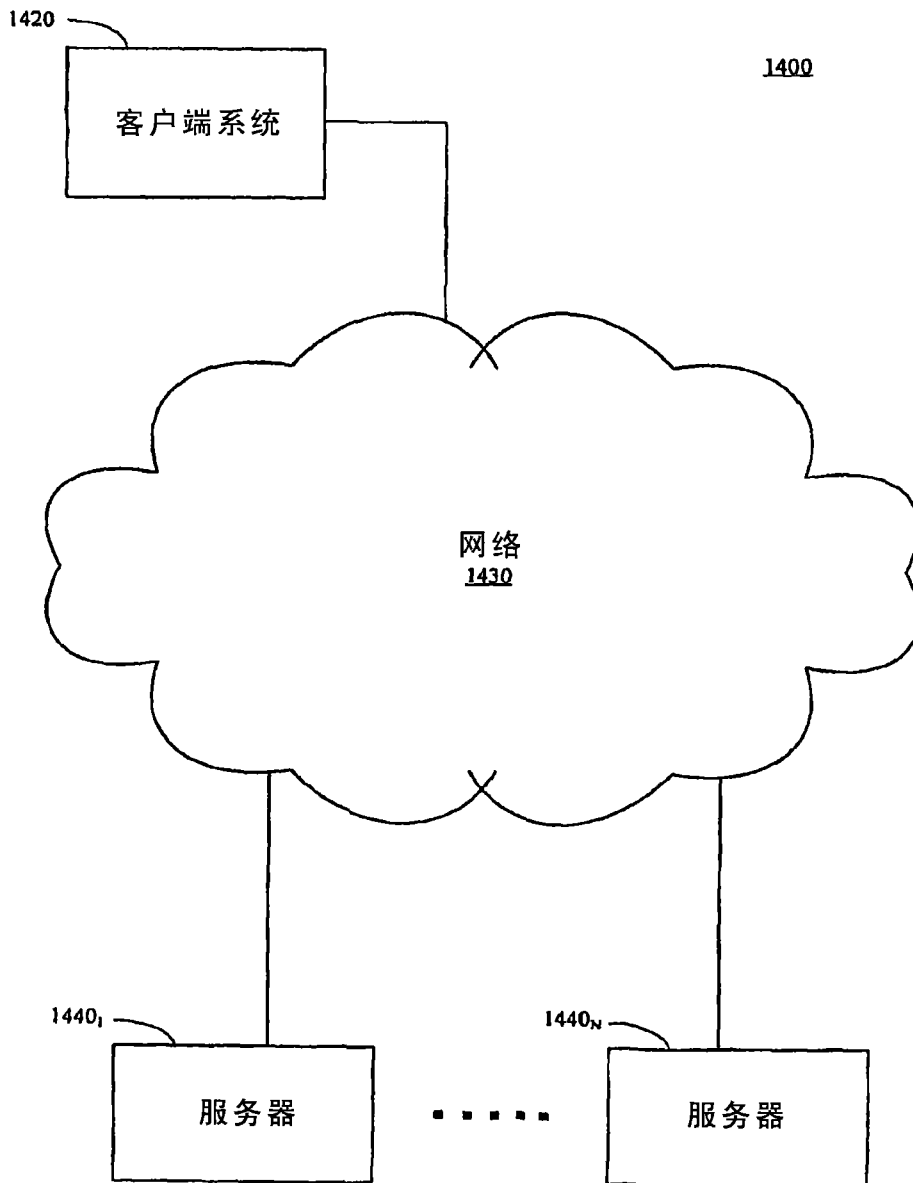


图14

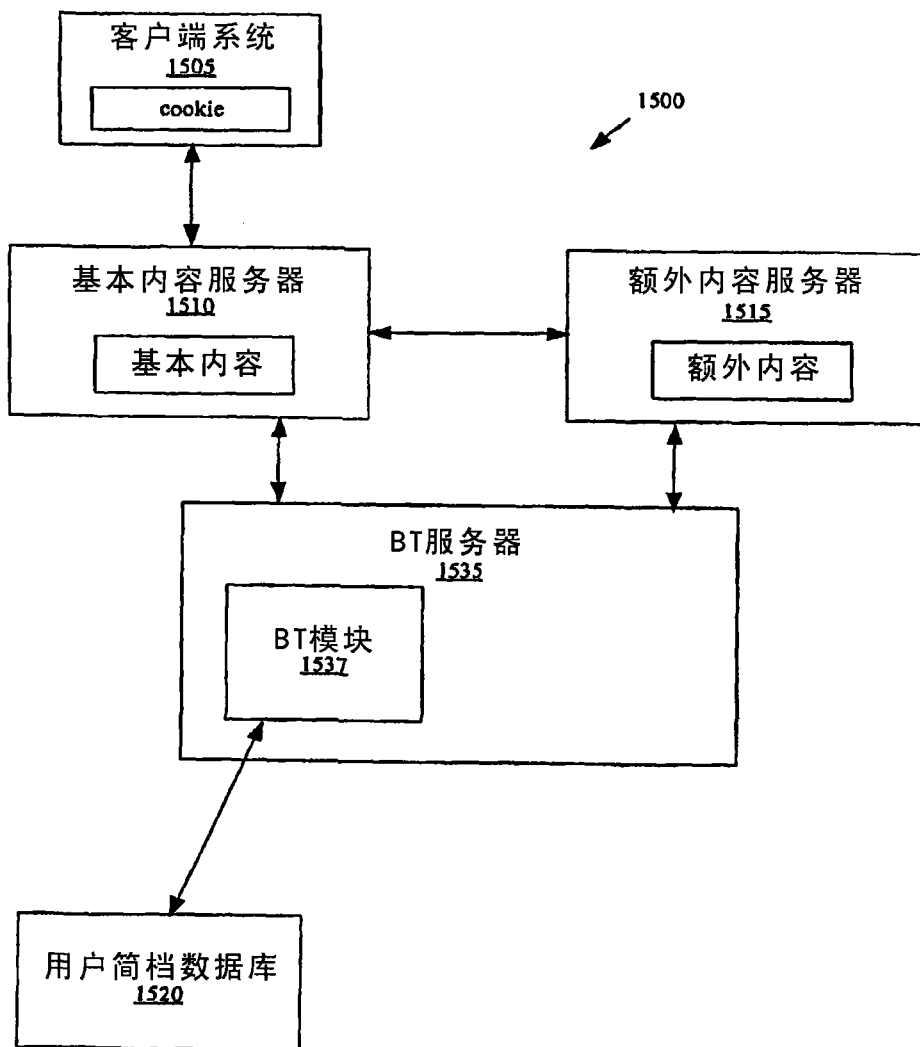


图15