(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101421728 B
(45)授权公告日 2011.09.07

(21)申请号 200780012760.2
(22)申请日 2007.03.09
(30)优先权数据
11/401,022 2006.04.10 US
(85)PCT申请进入国家阶段日
2008.10.09
(86)PCT申请的申请数据
PCT/US2007/006241 2007.03.09
(87)PCT申请的公布数据
WO2007/123606 EN 2007.11.01
(73)专利权人 微软公司
地址 美国华盛顿州
(72)发明人 D·密尔斯汀 K·王 L·克里德尔
L-C·楚 M·D·马卢厄
S·C·福布斯
(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 顾嘉运
(51)Int.Cl.
G06F 17/30(2006.01)

(56)对比文件
US 2006053010 A1,2006.03.09,
CN 1480876 A,2004.03.10,
CN 1496522 A,2004.05.12,
CN 1691019 A,2005.11.02,
US 2003023730 A1,2003.01.30,
US 2003140159 A1,2003.07.24,

审查员 梁艳

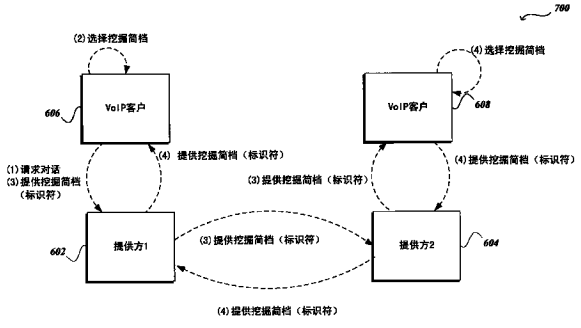
权利要求书 2 页 说明书 14 页 附图 15 页

(54)发明名称

挖掘关于服务的数据

(57)摘要

一般而言,本发明提供了处理数字语音对话以标识包含感兴趣内容的数据分组并且进一步处理所标识的数据分组的能力。更具体而言,可以开发标识要被挖掘的特定类型的内容,并进一步标识当定位到包含此类内容的数据分组时要完成什么的挖掘简档。一系统可在数字语音对话中搜索包含该内容的数据分组并且一旦标识了该数据分组即对其执行处理。



CN 101421728 B

1. 一种用于从数字语音对话中挖掘数据分组 [500] 的方法，包括 ：

选择用于从所述数字语音对话中挖掘数据分组的挖掘简档 [1315]，其中所述挖掘简档指定要从所述数字语音对话中挖掘的内容的类型，或，要对所述数据分组执行的处理的类型 ；

确定是否要挖掘所述数字语音对话的数据分组 [1305] ；以及

处理所述数字语音对话的数据分组 [1321] 以确定所述数据分组是否包括涉及所述挖掘简档的内容。

2. 如权利要求 1 所述的方法，其特征在于，还包括 ：

响应于确定所述数据分组包括涉及所述挖掘简档的内容，指示所述数据分组是所挖掘的数据分组。

3. 如权利要求 1 所述的方法，其特征在于，还包括 ：

响应于确定所述数据分组涉及所述挖掘简档，进一步处理 [1400] 所述数据分组以获得涉及所述数据分组的内容的附加信息。

4. 如权利要求 1 所述的方法，其特征在于，还包括 ：

确定挖掘规则 [1317] ；以及

基于所述挖掘规则来修改所述挖掘简档 [1319]。

5. 如权利要求 1 所述的方法，其特征在于，还包括 ：

将所述数据分组与所述挖掘简档一起储存 [1331] 以供后处理。

6. 如权利要求 5 所述的方法，其特征在于，还包括 ：

标识后处理规则 [1325] ；以及

将所述后处理规则与所述挖掘简档和所述数据分组一起储存以供后处理 [1331]。

7. 如权利要求 1 所述的方法，其特征在于，选择挖掘简档包括选择多个挖掘简档。

8. 一种用于从数字语音对话中挖掘数据分组 [500] 的系统，包括 ：

标识与挖掘简档相关的数据分组的所挖掘的数据分组管理组件 [1300]，其中所述挖掘简档指定要从所述数字语音对话中挖掘的内容的类型，或，要对所述数据分组执行的处理的类型 ；

确定所挖掘的数据分组是否要被更改的所挖掘的数据分组更改组件 [1405] ；

对与挖掘简档相关的数据分组执行处理的所挖掘的数据分组处理组件 [1409]。

9. 如权利要求 8 所述的系统，其特征在于，所挖掘的数据分组更改组件基于应用于挖掘简档的规则 [1407] 来更改数据分组的内容。

10. 如权利要求 9 所述的系统，其特征在于，所述规则由服务提供方 [132] 应用。

11. 如权利要求 8 所述的系统，其特征在于，所挖掘的数据分组的内容可通过从数字语音对话中移除所述数据分组、储存所述数据分组、替换所述内容或转发所述数据分组来更改。

12. 如权利要求 8 所述的系统，其特征在于，所述所挖掘的数据分组处理组件向个人提供涉及所述所挖掘的数据分组的内容的实时信息。

13. 如权利要求 8 所述的系统，其特征在于，对所述所挖掘的数据分组的处理是实时完成的。

14. 如权利要求 8 所述的系统，其特征在于，还包括 ；

用于将所挖掘的数据分组与后处理规则一起储存的后处理组件 [1331]。

15. 一种用于提供涉及数字语音对话的内容的附加信息的方法，包括 ：

处理所述对话的数据分组 [1300] 以标识所述数据分组内与挖掘简档相关的内容，其中所述挖掘简档指定要从所述数字语音对话中挖掘的内容的类型，或，要对所述数据分组执行的处理的类型 ；

获得与所标识的内容相关的附加信息 [1409] ；以及

提供所述相关的附加信息。

16. 如权利要求 15 所述的方法，其特征在于，所述处理是由第三方执行的。

17. 如权利要求 16 所述的方法，其特征在于，所述第三方被请求来从数字语音对话中挖掘特定的内容项。

18. 如权利要求 15 所述的方法，其特征在于，所述数据分组基于所述数据分组的头部被标识来进行处理。

挖掘关于服务的数据

技术领域

[0001] 本发明涉及挖掘数据，尤其，本发明涉及挖掘关于服务的数据。

背景技术

[0002] 一般而言，因特网电话系统为用户提供了拥有一呼叫连接的机会，该呼叫连接与常规的电话系统相比具有增强的呼叫特征。在通常被称为网际协议语音 (VoIP) 的典型因特网电话系统中，音频信息被处理成称为分组的数据块序列，以便利用网际协议 (IP) 数据网络进行对话。在 VoIP 呼叫对话期间，数字化的语音被转换成小的语音数据帧并通过向发送和接收的语音数据帧添加 IP 头部来组装语音数据分组。

[0003] VoIP 电话受到大家喜爱是因为其对话的灵活性和可移植性、建立和控制多媒体通信的能力等。VoIP 技术很可能因其提供增强的呼叫特征和高级服务的能力而继续得到青睐。

发明内容

[0004] 提供本概述以便以简化的形式介绍将在以下详细描述中进一步描述的一些概念。本概述并不旨在标识所要求保护的主题的关键特征，也不旨在用于帮助确定所要求保护的主题的范围。

[0005] 一般而言，本发明提供了处理数字语音对话以标识包含该对话内且与该对话相关联的感兴趣内容的数据分组以便进一步处理所标识的数据分组的能力。更具体而言，可以开发标识要被挖掘的特定类型的内容，并进一步标识当定位到包含此类内容的数据分组时要完成什么的挖掘简档。一系统可在数字语音对话中搜索包含该内容的数据分组并且一旦标识了该数据分组即对其执行处理。

[0006] 根据一方面，提供了一种用于从数字语音对话中挖掘数据分组的方法。可选择标识将要从该数字语音对话中挖掘的内容类型的挖掘简档。当客户设备中、服务提供方、服务器（内部部署或外部分机）或第三方中的任何一个或多个接收到数据分组时，该数据分组可被迅速检查以确定其是否可能包含值得挖掘的内容。例如，一数据分组可包含标识该数据分组是否包含值得挖掘的内容的头部。如果确定该数据分组可能包含值得挖掘的内容，则可分析该数据分组以确定其是否包含与挖掘简档相关的内容。

[0007] 根据另一方面，提供了一种具有用于处理所挖掘的数据分组的计算机可执行组件的计算机可读介质。该计算机可读介质包括被配置成标识对话中与一个或多个挖掘简档相关的数据分组的所挖掘的数据分组管理组件。还可包括确定所挖掘的数据分组或其内容是否要被更改，以及如果是，则如何被更改的所挖掘的数据分组更改组件。另外，还包括被配置成对所挖掘的数据分组执行处理的所挖掘的数据分组处理组件。

[0008] 根据另一方面，提供了一种用于提供与数字语音对话的内容相关的附加信息的方法。该方法包括处理对话的数据分组以标识与一挖掘简档相关的内容以及获得与所标识的内容相关的信息。例如，如果所标识的内容是关于“鞋”，则该方法可获得关于鞋的销售信息

并实时地或在稍后的时刻提供该信息。

附图说明

[0009] 当结合附图参考以下详细描述时，可以更容易领会并更好地理解本发明的上述方面和许多附加优点，附图中：

[0010] 图 1 是说明根据本发明的一个方面的用于在各客户之间建立对话信道的 VoIP 环境的框图；

[0011] 图 2 是说明根据本发明的一方面的对应于 VoIP 客户的各种 VoIP 设备的框图；

[0012] 图 3 是说明根据本发明的一方面的与 VoIP 客户设备相关联的各组件的框图；

[0013] 图 4A 和 4B 是说明根据本发明的一个方面的两个 VoIP 客户之间通过对话信道的数据交换的框图；

[0014] 图 5 是在图 1 的 VoIP 环境中所建立的对话信道上使用的数据分组的框图；

[0015] 图 6 是示出根据本发明的一方面的两个 VoIP 客户之间的用于传输由所标识的结构化分层结构定义的上下文信息的交互的框图；

[0016] 图 7 是示出根据本发明的一个实施例的两个 VoIP 客户之间用于交换挖掘简档的交互的框图；

[0017] 图 8-12 是说明根据本发明的一方面的对应于 VoIP 上下文信息的结构化分层结构的各属性和类的框图；以及

[0018] 图 13 是根据本发明的一方面的用于挖掘数字语音对话的挖掘例程的流程图；以及

[0019] 图 14 是根据本发明的一个实施例的用于处理已挖掘的数据分组的所挖掘的数据例程的流程图。

具体实施方式

[0020] 一般而言，本发明涉及对包含在诸如网际协议语音 (VoIP) 对话等数字语音对话中的内容的标识、提取和进一步使用。更具体而言，本发明涉及从对话和 / 或与对话相关的信息中“挖掘”数据。如此处所使用的“挖掘”或“挖掘服务”是对对话的处理以标识感兴趣的数据（对话和 / 或上下文的）。挖掘可以在对话期间在交换数据时对数据执行。或者，挖掘可对历史的或储存的数据执行（此类数据可涉及在过去发生、当前正在发生或在将来将会发生的事件或动作）。所标识的数据此处称为“所挖掘的数据”，其可由其它服务或应用程序使用、储存、转发、从对话中提取、替换、用附加数据补充等。类似于挖掘，对所挖掘的数据的处理可以在对话期间实时完成。或者，或除此之外，所挖掘的数据可被储存并被后处理。

[0021] 用于协助数据挖掘的一种技术是通过利用由“结构化分层结构”定义的类和属性，该结构化分层结构用于在网际协议 (IP) 网络环境中通过对话信道来表示上下文信息。此处使用的“结构化分层结构”是用于安排要在两个或更多 VoIP 设备之间交换的上下文信息的预定义的组织结构。例如，结构化分层结构可由诸如 XML 命名空间等各种类和属性的分层组织来定义。

[0022] 参考图 1，示出了用于在各“VoIP 客户”之间提供 IP 电话服务的 IP 电话环境 100

的框图。此处使用的“VoIP 客户”指的是具有一个或多个相关联的 VoIP 设备和唯一 VoIP 客户标识符的诸如个人、组织、公司等特定联系点。例如，单个个人、5 个相关联的 VoIP 设备和唯一 VoIP 客户标识符共同构成 VoIP 客户。类似地，包括 500 个个人和超过 1000 个相关联的 VoIP 设备的公司也可以被统称为 VoIP 客户，并且可以用唯一 VoIP 客户标识符来标识该 VoIP 客户。此外，VoIP 设备可以与多个 VoIP 客户相关联。例如，位于三个不同的个人（每一个人都与一单独的 VoIP 客户相关联）生活的住处的计算机（VoIP 设备）可以与这三个 VoIP 客户中的每一个相关联。不论设备的组合如何，都可以在语音系统中使用该唯一 VoIP 客户标识符来达到该 VoIP 客户的联系点。

[0023] 一般而言，IP 电话环境 100 可以包括诸如因特网、广域网（“WAN”）、局域网（“LAN”）等 IP 数据网络 108。IP 电话环境 100 还可以包括向 VoIP 客户 124、125、134 提供 VoIP 服务的 VoIP 服务提供方 126、132。VoIP 呼叫对话可以作为对应于语音信息、媒体信息、和 / 或上下文信息的数据分组流来交换。如将在以下更详细讨论的，数据可由客户、客户设备、服务提供方、第三方服务或其任意组合从对话中挖掘。在一个实施例中，可对对话中的没有客户和 / 或对第三方（例如，服务提供方）定义一个或多个“挖掘简档”。“挖掘简档”指定了对客户和 / 或第三方感兴趣的数据 / 内容的类型以及一旦标识了则将对该数据 / 内容要完成什么。例如，客户可创建指示其对鞋感兴趣的挖掘简档。该语音简档可用于处理对话并挖掘包括关于鞋的讨论的数据分组。所挖掘的数据分组然后可被进一步处理以确定该对话是 / 曾经是关于Nike®鞋子的，并且可以获得并提供关于Nike®鞋子的附加信息（例如，Nike®鞋子的销售）。

[0024] IP 电话环境 100 还可以包括第三方 VoIP 服务提供方 140。VoIP 服务提供方 126、132、140 可以提供各种呼叫特征，如来电过滤、文本数据、语音和媒体数据集成、挖掘服务、以及集成数据传输，作为 VoIP 呼叫对话的一部分。VoIP 服务提供方 126、132、140 还可生成、维护并提供挖掘简档以供客户在呼叫对话中传送来帮助标识要挖掘的数据。作为替换，或除此之外，VoIP 客户 104、124、125、136 可创建、维护和提供挖掘简档。

[0025] VoIP 服务提供方 132 可以被耦合到诸如公司 LAN 136 等专用网络，从而向通信地连接到公司 LAN 136 的若干 VoIP 客户 134 提供 IP 电话服务（例如，该专用网络中的内部呼叫、该专用网络外的外部呼叫等）和多媒体数据服务。类似地，诸如 VoIP 服务提供方 126 等 VoIP 服务提供方可以被耦合到因特网服务供应商 (ISP)122，从而向 ISP 122 的客户提供 IP 电话服务和 VoIP 服务。

[0026] 在一实施例中，一个或多个 ISP 106、122 可被配置为向 VoIP 客户 104、124、125 提供因特网接入以使 VoIP 客户 104、124、125 可以维护在因特网上建立的对话信道。连接到 ISP 106、122 的 VoIP 客户 104、124、125 可以使用有线和 / 或无线通信线路。此外，每一 VoIP 客户 104、124、125、134 都可以与通信地连接到公共交换电话网 (PSTN)112 的普通老式电话业务 (POTS)115 进行通信。诸如 PSTN 网关等 PSTN 接口 114 可以在 PSTN 和 IP 数据网络 108 之间提供访问。PSTN 接口 114 可以为 PSTN 将 VoIP 数据分组转换成电路交换语音话务，反之亦然。PSTN 112 可以包括陆线设备 116、移动设备 117 等。

[0027] 诸如陆线 116 等常规的设备可以基于 VoIP 客户的唯一 VoIP 标识符来请求与该 VoIP 客户连接，然后将使用与该 VoIP 客户相关联的适当的 VoIP 设备来建立连接。在一示例中，与 VoIP 客户相关联的个人可以基于各种条件（例如，基于呼叫方、当天的时间等的连

接）来指定在连接呼叫时要使用哪一设备。

[0028] 应当理解，环境 100 中的上述配置仅仅是示例性的。本领域普通技术人员将认识到，具有各种 VoIP 实体的任何数目和组合的合适的配置都可以是环境 100 的一部分。例如，耦合到 LAN 136 的 VoIP 客户 134 可以能够使用或不使用 VoIP 服务提供方 132 或 ISP 106、122 来与其它 VoIP 客户 104、124、125、134 进行通信。此外，ISP 106、122 还可以向其客户提供 VoIP 服务。

[0029] 现参考图 2，示出根据本发明的一实施例的包括若干 VoIP 设备和一唯一 VoIP 标识符的示例性 VoIP 客户 200 的框图。每一 VoIP 设备 202、204、206 都可以包括用来维护语音消息、地址簿、客户指定的规则、语音简档、挖掘简档等的存储。另选地或另外地，例如由服务提供方维护的独立存储可以与该 VoIP 客户相关联且可由包含涉及该 VoIP 客户的信息的每一 VoIP 设备访问。在一实施例中，诸如无线电话 202 和 IP 电话 204、或具有适当的 VoIP 应用程序的计算机 206 等任何合适的 VoIP 设备都可以是 VoIP 客户 200 的一部分。VoIP 客户 200 还维护一个或多个唯一客户标识符 208。唯一客户标识符 208 可以是常量或可以随时间改变。例如，唯一标识符 208 在每次呼叫时都可以改变。该唯一 VoIP 标识符被用来标识客户并与同该 VoIP 客户相关联的联系点 210 进行连接。该唯一 VoIP 标识符可以在包括在该 VoIP 客户中的每一 VoIP 设备上进行维护和 / 或由包括与该 VoIP 客户中所包括的每一 VoIP 设备的关联的服务提供方来维护。在唯一 VoIP 标识符由服务提供方维护的情况中，该服务提供方可以包括关于每一相关联的 VoIP 设备的信息和关于要为传入的通信连接哪一（哪些）设备的知识。在替换实施例中，VoIP 客户 200 可以维护多个客户标识符。在该实施例中，可以为每一呼叫会话向 VoIP 客户 200 临时分配唯一 VoIP 标识符。

[0030] 该唯一 VoIP 标识符可类似于 PSTN 中的电话号码来使用。然而，取代拨打典型的电话号码来使诸如家庭电话等特定 PSTN 设备响铃，该唯一 VoIP 标识符被用来达到诸如个人或公司等与该 VoIP 客户相关联的联系点。基于客户的安排，将连接适当的设备来达到该联系点。在一实施例中，包括在 VoIP 客户中的每一 VoIP 设备还可以具有其自己的网络中的物理地址或唯一设备号码。例如，如果个人使用个人计算机 (VoIP 设备）给 POTS 客户打电话，则该 VoIP 客户标识号码连同该个人计算机的 IP 地址将最后一起被转换成在 PSTN 中可以识别的电话号码。

[0031] 图 3 是可以与一个或多个 VoIP 客户相关联并与本发明的各实施例一起使用的 VoIP 设备 300 的框图。应该注意，VoIP 设备 300 是作为示例来描述的。可以理解，具有各种其它组件的任何合适的设备都可以与本发明的各实施例一起使用。为利用 VoIP 服务，VoIP 设备 300 可以包括适用于接收、发送和处理各种类型的数据分组的组件。例如，VoIP 设备 300 可以包括多媒体输入 / 输出组件 302 和网络接口组件 304。多媒体输入 / 输出组件 302 可以被配置为输入和 / 或输出多媒体数据（包括音频、视频等）、用户生物测定、文本、应用程序文件数据等。多媒体输入 / 输出组件 302 可以包括任何合适的用户输入 / 输出组件，如话筒、摄像机、显示屏、键盘、用户生物测定识别设备等。多媒体输入 / 输出组件 302 还可以经由网络接口组件 304 接收和发送多媒体数据。网络接口组件 304 可以支持诸如以太网接口、帧中继接口、电缆接口、DSL 接口、令牌环接口、射频（空中接口）等接口。VoIP 设备 300 可以包括含有诸如只读存储器设备 (ROM)、随机存取存储器 (RAM)、硬盘驱动器、光盘驱动器等永久的和 / 或可移动的存储的硬件组件 306。该存储可被配置为存储用于控制操作

系统和/或一个或多个应用程序的操作的程序指令，以及存储涉及与包括该设备的VoIP客户相关联的个人的上下文信息（例如，挖掘简档）。在一实施例中，硬件组件306可以包括允许非VoIP客户设备发送和接收VoIP对话的VoIP接口卡。

[0032] 设备300还可以包括用于设备300的操作的软件应用程序组件310和用于支持各种VoIP服务的VoIP服务应用程序组件308。VoIP服务应用程序组件308可以包括诸如数据分组组装器/拆解器应用程序、结构化分层结构解析应用程序、音频编码器/解码器(CODEC)、视频CODEC和用于提供VoIP服务的其它合适的应用程序等应用程序。软件应用程序组件300还可包括利用一个或多个挖掘简档来从对话中挖掘数据的挖掘组件。客户可维护和使用一个以上挖掘简档。例如，可对个人、商业、家庭、朋友等建立，并基于对话选择和使用标识要挖掘的不同类型的内容的不同挖掘简档。挖掘组件可利用诸如但不限于，语音识别、关键字搜索、图像识别、语音识别、上下文数据搜索和/或分组分析等技术来从对话中挖掘数据。

[0033] 参考图4A，示出说明根据本发明的一实施例的两个不同的VoIP客户的VoIP设备之间通过对话信道的对话流程400的框图。尽管此处提供的示例集中于两个VoIP客户之间的通信，但是可以理解，对话也可以是在任意数量的VoIP客户之间的。在连接建立阶段，第一VoIP客户406的VoIP设备请求启动与第二VoIP客户408的对话信道。在一个说明性实施例中，对于第一VoIP客户406的VoIP服务提供方402(提供方1)接收启动对话信道的请求，并将该请求转发给对于第二VoIP客户406的VoIP服务提供方404(提供方2)。

[0034] 尽管该示例利用了两个VoIP服务提供方和两个VoIP客户，但是对本发明的实施例可以使用任意数量和组合的VoIP客户和/或服务提供方。例如，在建立该连接时可以只利用一个服务提供方。而在另一示例中，通过利用公共和专用线路，VoIP设备之间的通信可以是直接的，从而消除对VoIP服务提供方的需求。在对等上下文中，VoIP设备之间的通信也可以是直接的，而不涉及任何服务提供方。

[0035] 有各种可被选择以供在各VoIP客户、各VoIP设备、和/或各VoIP服务提供方之间交换信息时使用的协议。例如，在选择会话发起协议(SIP)作为信令协议时，将通过SIP信令路径/信道来交换会话控制信息和消息，并通过实时传输协议(RTP)路径/信道来交换媒体流。出于讨论的目的，此处使用的通信信道一般指的是任何类型的数据或信号交换路径/信道。因此，可以理解，取决于协议，连接建立阶段和连接终止阶段可以在对话流程400中需要额外步骤。

[0036] 为易于解释，将利用其中第一VoIP客户406和第二VoIP客户408各自都只包括一个VoIP设备的示例。因而，在此提供的讨论将涉及两个VoIP设备的连接。使用第一VoIP客户406的设备的个人可以选择或输入要被呼叫客户的唯一VoIP标识符。如果要由第一客户的设备完成挖掘，则可选择适当的挖掘简档。挖掘简档可以由个人手动选择。或者，挖掘简档可以基于例如所使用的设备、所呼叫的客户、进行呼叫的位置、一天中的时间等来自动选择。

[0037] 提供方1402从第一VoIP客户408的设备接收该请求并基于包括在该请求中的唯一VoIP标识符来确定终止服务提供方（例如，第二VoIP客户408的提供方2 404)。另外，如果提供方1402要挖掘数据，则使用以上讨论的技术中的任一种来选择适当的挖掘简档。随后将该请求转发到提供方2 404。该呼叫发起将被转发给第二VoIP客户的设备。随后可

以在第一 VoIP 客户 406 的设备和第二 VoIP 客户 408 的设备之间建立对话信道。如果提供方 2404 或第二 VoIP 客户中的任一个或两者要挖掘数据，则可对该对话标识关于第二客户的适当的挖掘简档。

[0038] 在一说明性实施例中，在第一 VoIP 客户 406 和第二 VoIP 客户 408 的设备开始交换数据分组之前，可以交换上下文信息。如下更详细地讨论的，可以根据与该对话相关联的预定结构来将该上下文信息分组化。与第一 VoIP 客户 406、第一 VoIP 客户 406 的服务提供方或不同的设备 / 服务提供方相关联的任何设备都可以基于该上下文信息的内容来确定该结构。在一实施例中，所交换的上下文信息可以包括涉及呼叫 VoIP 客户 406、设备、和被呼叫 VoIP 客户 408、以及可任选的第三方服务的信息。

[0039] 可用媒体类型、呼叫客户和被呼叫客户的规则、呼叫客户和被呼叫客户的适当的挖掘简档等也可以是在连接建立阶段期间所交换的上下文信息的一部分。取决于该上下文信息的本质，该上下文信息可以由第一 VoIP 客户 406 的设备中的一个、第二 VoIP 客户 408 的设备中的一个、和 / 或由 VoIP 服务提供方（例如，提供方 1402 和提供方 2404）来处理和收集。在一实施例中，VoIP 服务提供方 402、404 在转发该客户的上下文信息之前，可以向 / 从该客户的上下文信息中添加和 / 或删除某些信息。

[0040] 响应于发起对话信道的请求，第二 VoIP 客户 408 可以经由提供方 2404 接受建立对话信道的请求或执行诸如拒绝该请求等其它适当的动作。可以基于所获得的上下文信息来确定这些适当的动作。在建立了对话信道时，第一 VoIP 客户 406 的设备和第二 VoIP 客户 408 的设备开始通过交换数据分组来互相通信。如下将更详细地描述的，包括对话数据分组和上下文数据分组的这些数据分组通过已连接的设备之间已建立的对话信道来传送。所交换的数据分组中的某一些或全部可被挖掘以满足挖掘简档的数据匹配要求。

[0041] 对话数据分组携带涉及对话的数据，例如，语音数据分组或多媒体数据分组。上下文数据分组携带涉及除对话数据之外的数据的信息。一旦建立了对话信道，或者第一 VoIP 客户 406 或者第二 VoIP 客户 408 都可以请求终止该对话信道。在终止之后，可以在第一 VoIP 客户 406 和第二 VoIP 客户 408 之间交换某些上下文信息。

[0042] 图 4B 是说明根据本发明的一实施例在两个 VoIP 客户的设备之间经由若干服务提供方的对话流 400 的框图。如图 4A 一样，此处描述的示例将利用其中每个客户只有一个与之相关联的设备且连接发生在那两个设备之间的场景。在连接建立阶段，第一 VoIP 客户 406 的设备请求启动与第二 VoIP 客户 408 通信的对话信道。在一个说明性实施例中，对于第一 VoIP 客户 406 的 VoIP 服务提供方 402（提供方 1）接收启动对话信道的请求，并将该请求转发给对于第二 VoIP 客户 408 的 VoIP 服务提供方 404（提供方 2）。如同以上相对于图 4A 所描述的示例一样，第一 VoIP 客户 406、提供方 1402、提供方 2404、和 / 或第二 VoIP 客户 408 中的任一个或多个可被指定来从对话挖掘数据。由此，设备可标识并可任选地交换关于与其它设备的对话的适当的挖掘简档。

[0043] 在第一 VoIP 客户 406 的设备和第二 VoIP 客户 408 的设备开始交换语音数据分组之前，可以在第一 VoIP 客户 406 和第二 VoIP 客户 408 之间交换上下文信息。上下文信息可使用由第一 VoIP 客户 406 定义的结构化组织来交换。在一个实施例中，提供方 1402 可以标识提供方 1402 期望从第一 VoIP 客户 406 获得的特定上下文信息。第一 VoIP 客户 406 可以基于上下文信息的内容来指定对应的结构。用于交换信息的结构的标识以及附加上下

文信息可以经由提供方 2404 和提供方 1402 发送到第二 VoIP 客户 408。

[0044] 取决于该上下文信息的本质，该上下文信息可以在第一 VoIP 客户的客户、第二 VoIP 客户的设备和 / 或 VoIP 服务提供方（例如，提供方 1 和提供方 2）处处理和收集。例如，挖掘简档可由服务提供方 402、404 选择和接收，且仅临时提供给各设备。此外，第三方服务提供方 410、412 也可维护用于从对话挖掘数据的挖掘简档。例如，第一 VoIP 客户可以请求第三方提供方在所有对话中挖掘特定内容。类似地，服务提供方 410、412 可以是第一 VoIP 客户正从中作出呼叫的公司或第二 VoIP 客户 408 所处的公司。公司可具有用于从对话挖掘与该公司的贸易秘密信息相关的特定关键字的强制挖掘简档。在标识了包括这些关键字的对话之后，所挖掘的数据可被储存以供将来使用，并且可执行其它处理（例如，终止对话、在继续之前需要进一步的认证 / 许可设置等等）。在一个实施例中，提供方 1402、提供方 2404 和第三方服务提供方 410、412 中的任一个可以在将上下文信息转发给包括其它服务提供方在内的下一 VoIP 设备之前添加、修改和 / 或删除上下文信息。

[0045] 响应于发起对话信道的请求，第二 VoIP 客户 408 可以经由提供方 2404 接受建立对话信道的请求或拒绝该请求。在建立了对话信道时，第一 VoIP 客户 406 和第二 VoIP 客户 408 的设备如上所述开始通过交换数据分组来互相通信。在一个实施例中，上下文和 / 或对话数据分组可从提供方 1402、提供方 2404 或从任一 VoIP 客户 406、408 转发到第三方服务提供方 410、412。此外，所转发的上下文和 / 或对话数据分组可在各种第三方服务提供方 410、412 之中交换。VoIP 客户 406、408、服务提供方 402、404 或第三方服务提供方 410、412 中的任一个可以基于所选挖掘简档来挖掘上下文和 / 或对话数据分组。

[0046] 图 5 是根据本发明的一实施例的在通信（对话）信道上使用的数据分组结构 500 的框图。数据分组结构 500 可以是用于适用于被利用来携带对话数据（例如，语音、多媒体数据等）或上下文数据（例如，涉及 VoIP 服务的信息等）的 IP 数据分组的数据分组结构。然而，可以利用任何其它合适的数据结构来携带对话数据或上下文数据。数据分组结构 500 包括头部 502 和净荷 504。头部 502 可以包含将对应的数据分组传送到目的地所必需的信息。另外，头部 502 可以包括在对话的过程中所利用的信息。这一信息可以包括用于标识对话（例如，呼叫）的对话 ID 506、诸如被呼叫的客户的唯一 VoIP 标识符等目的地 ID508、源 ID 510（呼叫客户的唯一 VoIP 标识符或设备标识符）、用于标识净荷类型的净荷 ID 512（例如，对话或上下文）、用于标识该净荷可能包含值得挖掘的内容的挖掘 ID 513、用于标识该对话数据所涉及的个人的个人 ID（未示出）等。在一替换实施例中，头部 502 可以包含关于因特网协议版本和净荷长度的信息，以及其它信息。净荷 504 可以包括涉及所标识的对话的对话或上下文数据。本领域普通技术人员将认识到，可以将额外的头部用作更上层的头部，如 TCP 头部、UDP 头部等。

[0047] 在本发明的一实施例中，可以预定义一结构化分层结构以通过 VoIP 对话信道来传送上下文信息。该上下文信息可以包括涉及 VoIP 客户、VoIP 客户设备、对话信道连接（例如，呼叫基础）、对话上下文（例如，呼叫上下文）等的任何信息更具体地，该上下文信息可以包括个人客户挖掘简档或挖掘简档标识符、客户规则、客户的位置（例如，用户的位置、设备位置等）、生物测定信息、用户的机密信息、VoIP 客户设备功能、VoIP 服务提供方信息、媒体类型、媒体参数、呼叫号码优先级、关键字、涉及应用程序文件的信息等。取决于该上下文数据的本质，可以在每一 VoIP 客户和 / 或 VoIP 服务提供方处处理和收集该上下文

信息。

[0048] 在一方面，VoIP 服务提供方在转发该上下文信息之前可以添加、修改和 / 或删除 VoIP 客户的上下文数据。例如，VoIP 客户的机密信息将被与该客户相关联的 VoIP 服务提供方删除，除非客户授权发送这一信息。在某些情况下，在内联网网络之外发送最小量的上下文信息。另外，如下所述，诸如办公室等某些内联网网络可以实现指示可从对话中挖掘什么以及可在对话中包括什么以供可能的挖掘的挖掘简档。

[0049] 参考图 6，示出了图解根据本发明的一实施例的两个 VoIP 客户之间的用于传输上下文信息的交互的框图 600。如图 4A 和 4B 一样，此处描述的示例将利用其中每个客户只有一个与之相关联的设备且连接发生在那两个设备之间的场景。在一实施例中，VoIP 客户 606 和 VoIP 客户 608 的设备已建立了 VoIP 对话信道。VoIP 客户 606 可以标识将使用哪些结构化分层结构来携带特定上下文信息。关于所标识的结构化分层结构的信息可以包括关于哪些结构化分层结构被用来携带该上下文信息、如何标识结构化分层结构等的信息。在交换对应的上下文信息之前，将在 VoIP 客户 606 和 VoIP 客户 608 之间交换这一信息。在接收到关于将使用哪一结构化分层结构来携带上下文信息的信息时，VoIP 客户 608 查找预定义的结构化分层结构（例如，XML 命名空间等）以选择所标识的结构化分层结构。在一实施例中，可以全局地存储并在可从一组 VoIP 客户访问的集中式位置管理这些预定义的结构化分层结构。在该实施例中，该集中式位置的统一资源标识符 (URI) 地址可从 VoIP 客户 606 发送到 VoIP 客户 608。

[0050] 在另一实施例中，每一 VoIP 客户都可以具有存储在任何设备的本地存储或所有设备可以共享的专用本地存储中的一组预定义的结构化分层结构。在交换上下文信息之前，可以在 VoIP 客户之间声明这些预定义的结构化分层结构并对其达成一致。以此方式，可以消除对提供上下文数据分组的结构的需求，并因而减少对应于该上下文数据的所发送的数据分组的量。此外，通过利预定义的结构化分层结构，可以用硬件和 / 或软件无关的方式来发送数据分组。另外，对预定义的结构化分层结构的使用由于已知可能相关的信息的位置而增强了高效地挖掘数据的能力。

[0051] 在检索所标识的结构化分层结构时，VoIP 客户 608 期望接收数据流以便根据所标识的结构化分层结构来定义对应于该数据流的数据分组。VoIP 客户 606 可以开始发送根据所标识的结构化分层结构来表示的上下文信息。在一实施例中，VoIP 客户 608 开始关于该上下文信息的数据绑定过程。例如，可以用所接收到的上下文信息来构造所标识的结构化分层结构的实例。

[0052] 参考图 7，框图 700 示出了根据本发明的一实施例的两个 VoIP 之间用于建立对话信道并选择在从对话中挖掘数据时使用的挖掘简档的交互。在一个实施例中，VoIP 客户 606 在请求与 VoIP 客户 608 的连接时可以选择将在该对话期间使用的挖掘简档。例如，如果 VoIP 客户 608 是个人朋友，则 VoIP 客户 606 可选择被创建来挖掘与两个朋友之间的感兴趣的共同项目有关的数据的挖掘简档。作为连接请求和对话信道建立的一部分，VoIP 客户 606 经由提供方 1602 和提供方 2604 可将所选的挖掘简档提供给 VoIP 客户 608。或者，如果客户先前已经交换了挖掘简档，则 VoIP 客户 606 可以只需通过发送挖掘简档标识符来标识所选的挖掘简档。类似地，VoIP 客户 608 在呼叫建立期间或在呼叫期间的任何时刻可以向提供方 604、602 和 VoIP 客户 606 提供由 VoIP 客户 608 选择的用于在该对话期间挖掘

数据的挖掘简档的挖掘简档（或其标识符）。

[0053] 如以下所讨论的，挖掘简档可以被交换，因为一方或双方都可能希望在对话完成之后对数据执行进一步的后处理。各方可具有关于可执行的后处理的类型和范围的限制。这些限制可以被包含在交换的挖掘简档中，并被应用于被储存来进行后处理的数据分组。在一替换实施例中，挖掘简档可在提供方 1602 和提供方 2604 之间维护和交换，并且数据可被挖掘和 / 或储存以供提供方 602、604 进行稍后的后处理。再者，如果不执行后处理，则可以不交换挖掘简档。挖掘简档可在对话中的任一时刻选择、使用和交换。另外，挖掘简档可在对话期间被改变、替换或移除。在一个实施例中，在对话期间改变或替换挖掘简档之前可能需要参与对话的个人的批准。

[0054] 在建立了对话信道之后，在对话期间，VoIP 客户 606、608 和 / 或提供方 602、604 可挖掘所交换的对话和上下文数据分组以寻找与挖掘简档相关的内容。例如，如果挖掘简档之一指定要挖掘涉及“高尔夫”的对话，使用 VoIP 客户 608 的个人提及其将要在本周日下午的 1:00 在 Pebble 海滩打高尔夫并且邀请使用 VoIP 客户 606 的个人加入他们，则可挖掘并进一步处理来自该数字语音对话的数据分组。例如，根据挖掘简档挖掘数据分组的设备可获得关于下午 1:00 加利福尼亚的 Pebble 海滩的天气信息，并在该对话期间将该信息提供给一个或多个 VoIP 客户。同样，可定义另一挖掘简档以标识日历项目。可挖掘涉及在周日下午 1:00 打高尔夫的请求的相同的数据分组并可激活日历程序。然后可向 VoIP 客户呈现对于周日下午 1:00 在 Pebble 海滩打高尔夫的日历请求，由此允许他们接受、修改或拒绝该请求并将其置于其日历上。

[0055] 除了 VoIP 客户 606 和 608 和 / 或提供方 602、604 挖掘对话的数据分组之外，客户之一或两者也可请求或同意允许第三方提供方（未示出）从该对话中挖掘数据分组。例如，如果客户之一同意允许高尔夫商店从该对话中挖掘数据分组，则该第三方可挖掘相同的数据分组样本，并且作为响应，可在该呼叫期间和 / 或之后向客户提供关于高尔夫服装或装备的广告。由于挖掘对话的特定部分的能力，可提供聚焦的广告。例如，高尔夫商店还可执行关于周日下午 1:00 加利福尼亚州 Pebble 海滩的天气搜索，并确定在该时段可能下雨。结果，提供给客户的广告可包括关于下雨用具的信息。

[0056] 可向一个对话应用多个挖掘简档，并且可实时挖掘并处理数据分组，且处理结果在对话期间提供给客户。也可对挖掘简档施加规则，由此修改可挖掘的数据的数量和 / 或类型。再者，可在稍后的时刻挖掘并处理数据分组（后处理）。

[0057] 除了交换挖掘简档之外，如上所述，可指定结构化分层结构以便通过指定发送的数据的结构而在对话期间用于协助高效地挖掘数据分组。VoIP 客户 606 还可定义预定义的结构化分层结构集的附加类和 / 或属性。例如，VoIP 客户 606 可标识具有关于特定上下文信息的附加属性和类的结构化分层结构。VoIP 客户 606 将对应于所标识的分层结构及其附加属性和类的信息发送给提供方 1602。提供方 1602 将对应于所标识的结构化分层结构及其附加属性和类的信息转发给提供方 2604。VoIP 客户 608 从提供方 2604 接收关于所标识的结构化分层结构及其附加属性和类的信息。或者，提供方 1602 将此类信息转发给 VoIP 客户 608。VoIP 客户 608 基于附加属性和类信息来更新本地储存的结构化分层结构。结果，所标识的结构化分层结构在 VoIP 客户 606 和 VoIP 客户 608 处扩展。VoIP 客户 606 开始向提供方 1602 发送根据所扩展的结构化分层结构表示的上下文信息，后者作为返回，向提供

方 2604 发送所接收到的上下文信息。在一个实施例中，在从提供方 2604 接收到上下文信息之后，VoIP 客户 608 通过构造所接收的上下文信息的所扩展的结构化分层结构的实例来处理所接收到的上下文信息。

[0058] 在一实施例中，可以通过可扩展标记语言 (XML) 来定义结构化分层结构。然而，可以理解，可以通过适用于实现和维护可扩展结构化分层结构的任何语言来定义该结构化分层结构。一般而言，XML 因为其是跨平台的、软件和硬件无关的发送信息的工具而众所周知。此外，XML 将其数据作为分层地结构化的节点树来维护，每一节点包括可以包含描述性属性的标签。通常，提供 XML 命名空间来给予该命名空间一个唯一名字。在一些实例中，该命名空间可被用作指向包含关于该命名空间的默认信息的集中式位置的指针。

[0059] 在一特定实施例中，VoIP 客户 606 通过将 XML 命名空间属性置于发送元素的起始标签中来标识上下文信息的 XML 命名空间。在 VoIP 客户 608 接收到 XML 命名空间信息之后，VoIP 客户 606 向 VoIP 客户 608 发送根据所标识的 XML 命名空间来定义的一组上下文数据分组。当命名空间在元素的起始标签中定义时，具有同一前缀的所有子元素都与该同一命名空间相关联。因此，VoIP 客户 608 和 VoIP 客户 606 可以发送上下文信息而不必在所有子元素中包括前缀，从而减少对于该上下文信息的所发送的数据分组的量。可以理解，可对本发明的各实施例使用用于标识结构化分层结构的任何技术。通过起始标签来标识 XML 命名空间的示例是仅作为示例来提供的，并且不被认为是限制。

[0060] 参考图 8-12，示出了说明对应于 VoIP 上下文信息的结构化分层结构的各种类和属性的框图。如上所述，“结构化分层结构”是用于安排要在两个或更多 VoIP 设备之间交换的上下文信息的预定义的组织结构。结构化分层结构可通过重定义各种类和属性来定义、更新和 / 或修改。在各 VoIP 实体（例如，客户、服务提供方等）之间交换的 VoIP 上下文信息可以对应于 VoIP 命名空间 800。在一实施例中，VoIP 命名空间 800 被表示为分层地结构化的节点树，每一节点对应于与 VoIP 上下文信息的子集相对应的子类。例如，VoIP 命名空间 800 可以被定义为分层地结构化的树，包括呼叫基础类 802、呼叫上下文类 810、设备类型类 820、VoIP 客户类 830 等。

[0061] 参考图 9，示出了呼叫基础类 802 的框图。在一说明性实施例中，呼叫基础类 802 可以对应于涉及对话信道连接（例如，PSTN 呼叫连接、VoIP 呼叫连接等）的 VoIP 上下文信息的子集。涉及对话信道连接的 VoIP 上下文信息的子集可以包括始发号码（例如，呼叫者的 VoIP ID 号码）、目的地号码（例如，被呼叫者的 VoIP ID 号码或电话号码）、呼叫连接时间、VoIP 服务提供方相关信息、和 / 或诸如 IP 地址、MAC 地址、命名空间信息等 ISP 相关信息等。另外，涉及对话信道连接的上下文信息可以包括呼叫优先级信息（其定义目的地号码的优先级）、呼叫类型信息等。呼叫类型信息可以指示该对话信道是否是为紧急通信、广播通信、计算机到计算机的通信、计算机到 POTS 设备的通信等建立的。在一实施例中，涉及对话信道连接的上下文信息可以包括表示情感、声音（例如，“啊”、“啊呀”、“哎唷”等）和图形符号形式的面部表情等预定义标识符。在一实施例中，呼叫基础类 802 可被定义为 VoIP 命名空间 800 的子树结构，其包括诸如呼叫优先级 803、命名空间信息 804、呼叫类型 805、目的地号码 806、服务提供方 807、预定义标识符 808 等节点。

[0062] 参考图 10，示出了呼叫上下文类 810 的框图。在一实施例中，涉及对话上下文的 VoIP 上下文信息的子集可以对应于呼叫上下文类 810。该涉及对话上下文的上下文信息可

以包括诸如客户提供的关键字、从文档文件数据中标识的关键字、从对话数据分组中标识的关键字（例如，对话关键字）、作为对话的一部分交换的文档和/或多媒体文件的文件名、游戏相关信息（如游戏类型、特定游戏中的虚拟邻近性）、使用频率（包括涉及特定文件、特定主题和特定客户的呼叫的频率和持续时间）以及文件标识（如涉及对话的件号、事件号等等）等信息，以及许多其它信息。根据一说明性实施例，呼叫上下文类 810 可被定义为 VoIP 命名空间 800 的子树结构，其包括与文件标识 812、客户提供的关键字 813、对话关键字 814、使用频率 815、对话主题 816 等相对应的节点。

[0063] 参考图 11，描绘了设备类型类 820 的框图。在一实施例中，设备类型类 820 可以对应于涉及用于对话信道连接的 VoIP 客户设备的 VoIP 上下文信息的子集。该涉及 VoIP 客户设备的 VoIP 上下文信息的子集可以包括处理由该 VoIP 客户设备生成的音频数据可能需要的音频相关信息。该音频相关信息可以包括涉及设备的音频功能和能力的信息，如采样速率、机器类型、输入/输出类型、话筒、数字信号处理 (DSP) 卡信息等。该涉及 VoIP 客户设备的 VoIP 上下文信息的子集可以包括处理由该 VoIP 客户设备生成的视频数据可能需要的视频相关信息。该视频相关信息可以包括分辨率、刷新率、视频数据的类型和大小、图形卡信息等。另外，如果确定了系统不兼容，则可添加一转换层以重新配置所交换的信息以便允许不同类型的系统之间的通信。涉及 VoIP 客户设备的上下文信息还可以包括其它设备专用信息，如计算机系统的类型、处理器信息、网络带宽、无线/有线连接、该计算机系统的可移植性、该计算机系统的处理设置等。在一说明性实施例中，设备类型类 820 可被定义为 VoIP 命名空间 800 的子树结构，其包括与音频 822、视频 824、设备专用 826 等相对应的节点。

[0064] 参考图 12，描绘了 VoIP 客户类 830 的框图。根据一个说明性实施例，VoIP 客户类 830 可以对应于涉及 VoIP 客户的上下文信息的子集。在一个实施例中，涉及 VoIP 客户的 VoIP 上下文信息的子集可以包括语音简档信息（例如，指定个人用户的声调和发音特性的信息的集合）、数字签名信息、生物测定信息、以及挖掘简档信息（例如，标识要在挖掘对话时使用的主要简档或挖掘简档集）。挖掘简档可包括关于要挖掘的内容的类型（挖掘要求）、如何挖掘数据分组（例如，语音识别、关键字搜索等）、何时挖掘数据分组（实时或储存并稍后挖掘）、要对所挖掘的数据分组完成的处理（例如，更新日志、生成任务、检索相关信息、生成广告等）、处理何时完成（实时或稍后）、以及允许从对话中挖掘的内容的类型的信息。

[0065] 生物测定信息可以包括涉及生物测定认证的用户标识信息（例如，指纹）、用户压力程度、用户情绪等。另外，涉及 VoIP 客户的 VoIP 上下文信息的子集可以包括位置信息（包括客户定义的位置、VoIP 定义的位置、GPS/ 三角测量位置、以及个人用户的逻辑/虚拟位置）、所分配的电话号码、用户联系信息（如名字、地址、公司等）、客户定义的规则、用户偏好、数字权限管理 (DRM)、个人用户在一组织中的成员等级、与该成员等级相关联的优先级等。与该成员等级相关联的优先级可被用来向客户分配关于电话会议的优先级。在一实施例中，VoIP 客户类 830 可被定义为 VoIP 命名空间 800 的子树结构，其包括与用户生物测定 831、位置 832、客户规则 833、用户标识 834、成员优先级 835、用户偏好 836、挖掘简档标识 837 等相对应的节点。

[0066] 图 13 是根据本发明的一实施例的用于从数字语音对话中挖掘数据分组的挖掘例

程的流程图。挖掘例程 1300 在框 1301 处开始。在框 1303 处，接收来自语音对话的数据分组。可以理解，挖掘例程 130 可使用单个数据分组或利用一组数据分组来执行。在判定框 1305，确定是否要挖掘数据分组。如上所述，数据分组可在该数据分组的头部中包括标识该数据分组的净荷可包括值得挖掘的内容的挖掘标识符 513（图 5）。或者，可对本发明的各实施例利用用于标识数据分组可包括值得挖掘的内容的任何技术。例如，如果数据分组是特定大小，由此指示它可能包含值得挖掘的有用内容，则可在判定框 1305 处选择该数据分组来进行挖掘。

[0067] 如果在判定框 1305 处确定所接收的数据分组要被挖掘，则在框 1307 处，确定用于挖掘的客户简档。如上所述，客户可创建在数字语音对话中使用的标识要挖掘的内容的类型的一个或多个简档。例如，客户可创建在个人对话中使用的个人挖掘简档。则以个人挖掘简档可包括标识要从对话本身中挖掘的内容的类型的话题或类别。例如，如果客户对鞋感兴趣，则个人挖掘简档可包括一标识符来搜索关键字或利用语音识别以标识包括对鞋或鞋类型的引用的对话的标识符。另外，可根据以上讨论的结构化分层结构以及用于确定数据分组的内容或结构是否包括对鞋的引用的上下文信息的相关部分来搜索对话的上下文信息。

[0068] 在判定框 1309 处，确定要用于挖掘在框 1303 处接收到的数据分组的一个或多个客户挖掘简档是否存在。如果在框 1309 处确定一个或多个客户挖掘简档的确存在，则在框 1311 处，获得这些简档。在判定框 1313，或者在框 1311 处获得了一个或多个简档之后，或者如果在判定框 1309 处确定不存在简档，确定要用于挖掘在框 1303 处接收到的数据分组的外部挖掘简档是否存在。外部简档可以是例如，但不限于，由客户生成的简档或要用于挖掘对话的其它简档。例如，客户可请求第三方在关于该个人的对话中挖掘特定的感兴趣内容。另外，外部挖掘简档可包括，但不限于，由服务提供方、政府、公司等管理的安全简档。

[0069] 如果在判定框 1313 处确定要使用外部挖掘简档来挖掘所接收到的数据分组，则在框 1315 处，获得这些简档。在框 1315 处获得了这些外部挖掘简档之后，或者如果在框 1313 处确定没有外部挖掘简档，则在判定框 1317 处，确定是否有要被应用来挖掘的任何挖掘规则。挖掘规则可以是，例如，由从中生成呼叫的公司指定的规则。例如，如果一客户正在从公司进行呼叫，则该公司可以具有不允许对源自该公司的对话的任何挖掘或包括该公司内的对话的挖掘规则。或者，挖掘规则可指定何时可从诸如公司等特定位置进行呼叫。如果在判定框 1317 处确定存在挖掘规则，则在框 1319 处，基于这些挖掘规则来修改在框 1311 和框 1315 处获得的挖掘简档。

[0070] 在框 1321 处，在于框 1319 处修改了所获得的挖掘简档之后，或者如果在框 1317 处确定没有挖掘规则，基于所获得的简档来挖掘在框 1303 处接收到的数据分组。例如，如果挖掘简档指定要使用语音识别和关键字搜索来挖掘数据分组（或一组数据分组），并且如果标识了包括相关关键字的内容，则该数据分组可被标记或以其它方式标识为要被进一步处理的“所挖掘的数据分组”。对所挖掘的数据分组的处理将参考图 14 来更详细地描述。

[0071] 除了在对话期间挖掘数据分组之外，可以在判定框 1323 处确定是否要执行对对话或所挖掘的数据分组的任何后处理。如果在判定框 1323 处确定要执行后处理，则在判定框 1325 处，确定是否有任何后处理规则。如果在判定框 1325 处确定有后处理规则，则在框 1327 处将这些后处理规则应用于所挖掘的数据分组。后处理规则可以是，例如，关于要对数

据分组执行什么类型的后处理、在后处理期间将对从数据分组挖掘的内容完成什么等的规则。然而，如果在判定框 1325 处确定没有后处理规则，或者在判定框 1327 处应用了后处理规则之后，应用在框 1311 和框 1315 处获得的、可能在框 1319 处修改的简档，并将其与要后处理的所挖掘的数据分组相关联。要被后处理的数据分组与挖掘简档和规则相关联，使得这些简档和规则可在对数据分组的后处理发生时被标识和获得。或者，可以不将挖掘简档应用于所挖掘的数据分组。在框 1331 处，保存所挖掘的数据分组、后处理规则和可任选的所应用的简档以供后处理。在于框 1331 处保存了数据分组和适当信息以供后处理之后，如果在判定框 1323 处确定不要实现后处理，或者如果在判决框 1305 处确定不要挖掘所接收到的数据分组，则 VoIP 挖掘例程 1300 完成，如框 1333 所示。

[0072] 图 14 是根据本发明的一实施例的用于处理从数字语音对话挖掘到的数据分组的所挖掘的数据分组例程的流程图。所挖掘的数据分组例程 1400 在框 1401 处开始。在框 1403 处，标识从数字语音对话挖掘的数据分组。类似于图 13，所挖掘的数据分组例程 1400 可以对单个数据分组或一组数据分组执行。另外，如以上所讨论的，所挖掘的数据分组可包含对话数据或上下文数据。在判定框 1405，确定所挖掘的数据分组或所挖掘的数据分组的内容是否要被更改。如果在判定框 1405 处确定所挖掘的数据分组 / 内容要被更改，则在框 1407 处，更改所挖掘的数据分组 / 内容。所挖掘的数据分组或内容可基于应用于挖掘简档的规则和 / 或基于后处理规则来更改。例如，规则可指定所有包含涉及公司的贸易秘密信息的内容的所挖掘的数据分组要从对话中移除并转发给公司的官员。可对所挖掘的数据分组和 / 或这些分组的内容执行任何类型的更改，包括但不限于，从对话中移除数据分组、用其它内容替换数据分组的内容、储存数据分组以供将来使用、记录数据分组或数据分组的内容等。

[0073] 在框 1407 处更改了所挖掘的数据分组或内容之后，或者如果在框 1405 处确定该数据分组或内容不要更改，则在框 1409 处，以各种方式中的任一种处理所挖掘的数据分组。所挖掘的数据分组可以用多种方式来处理以向客户提供附加服务。例如，来自所挖掘的数据分组的内容可用于反馈诸如日记、任务列表等可在稍后由客户访问来审阅该对话或获得附加信息的另一应用程序或服务。同样，来自所挖掘的数据分组的内容可用于获得涉及该内容的其它信息。例如，再次参考挖掘包含涉及鞋的内容的数据分组的示例，如果内容是关于Nike®鞋子的对话，则挖掘服务提供方可获得关于Nike®鞋子的附加信息。例如，附加信息可以是Nike®鞋子的销售、购买Nike®鞋子的位置、Nike®鞋子的质量等。该信息可以与对话一起储存以供稍后由客户查看和 / 或在对话期间提供给客户。

[0074] 除了反馈其它应用程序之外，来自所挖掘的数据分组的内容可用于重定向特定对话的各部分。例如，如果内容涉及对话的个人部分并且是高质量视频，但是接收设备是没有视频能力的公司设备，则该对话的视频部分可被重新路由到提供必要能力的不同客户设备并被查看或储存以供稍后在该设备上查看。另外，网络可利用所挖掘的数据分组来进行负载平衡以便基于如由所挖掘的数据分组指定的对话的内容和大小来重新路由或重新处理这些对话。

[0075] 相关领域的技术人员可以理解，存在可用于处理所挖掘的数据分组并向客户提供附加服务的无穷数量的处理技术和变体。因此，此处提供的示例仅用于说明，而不被解释为限制。

[0076] 在于框 1409 处处理了所挖掘的数据分组之后，在判定框 1411 处，确定是否接收到要处理的额外的所挖掘的数据分组。如果在判定框 1411 处确定存在要处理的额外的所挖掘的数据分组，则例程 1403 返回到框 1403 并继续。然而，如果在判定框 1411 处确定没有要处理的额外的所挖掘的数据分组，则例程 1400 在框 1413 处完成。

[0077] 尽管示出和描述了各说明性实施例，但可以理解，可在其中作出各种改变而不背离本发明的精神和范围。

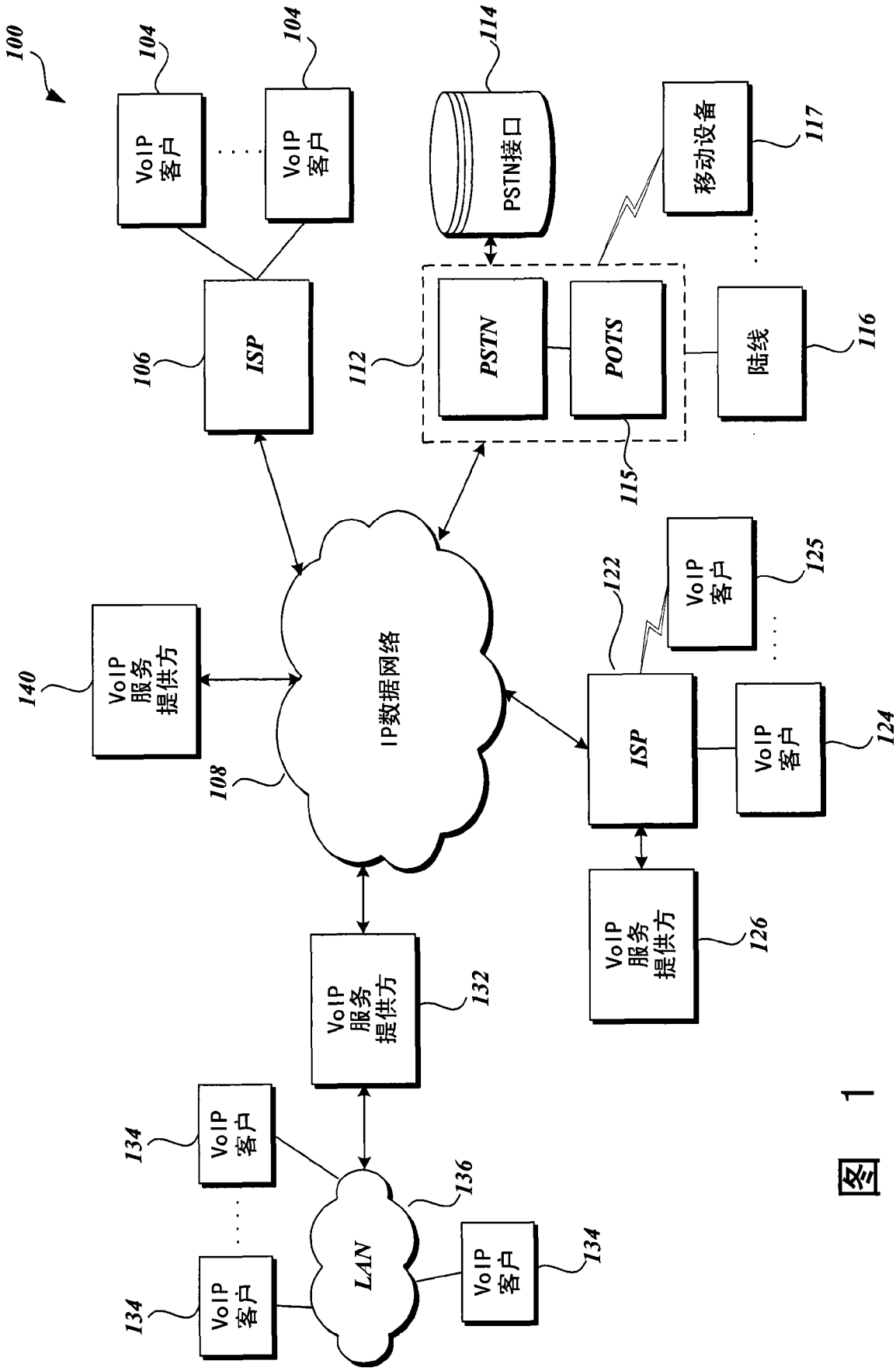

100
104
VoIP 客户
104
VoIP 客户
114
PSTN接口
移动设备
117
106
ISP
112
PSTN
POTS
115
陆线
116
140
VoIP 服务提供方
IP数据网络
108
122
VoIP 客户
125
ISP
VoIP 客户
124
VoIP 服务提供方
126
132
VoIP 服务提供方
134
VoIP 客户
134
VoIP 客户
LAN
136
VoIP 客户
134

图 1

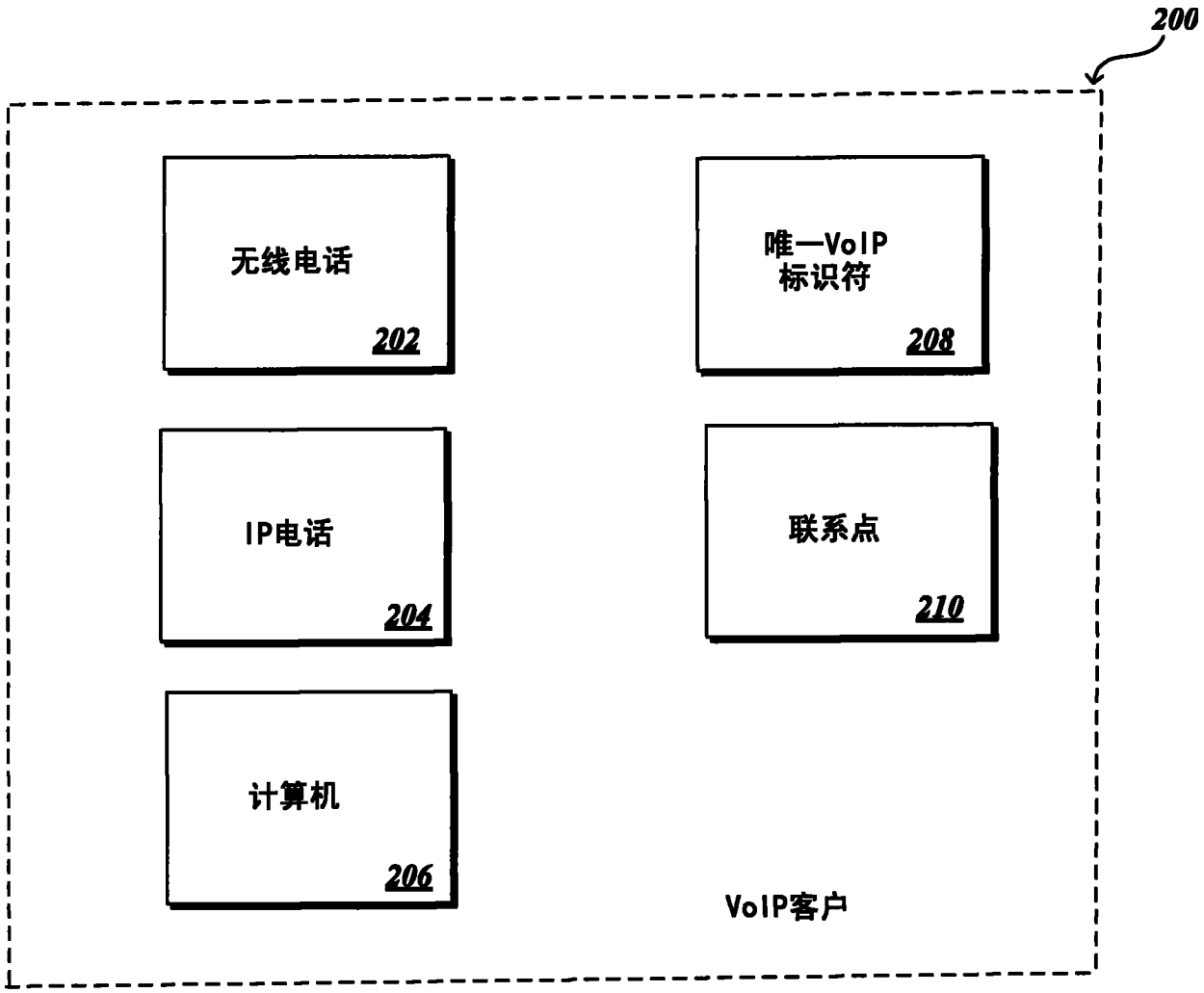


图 2

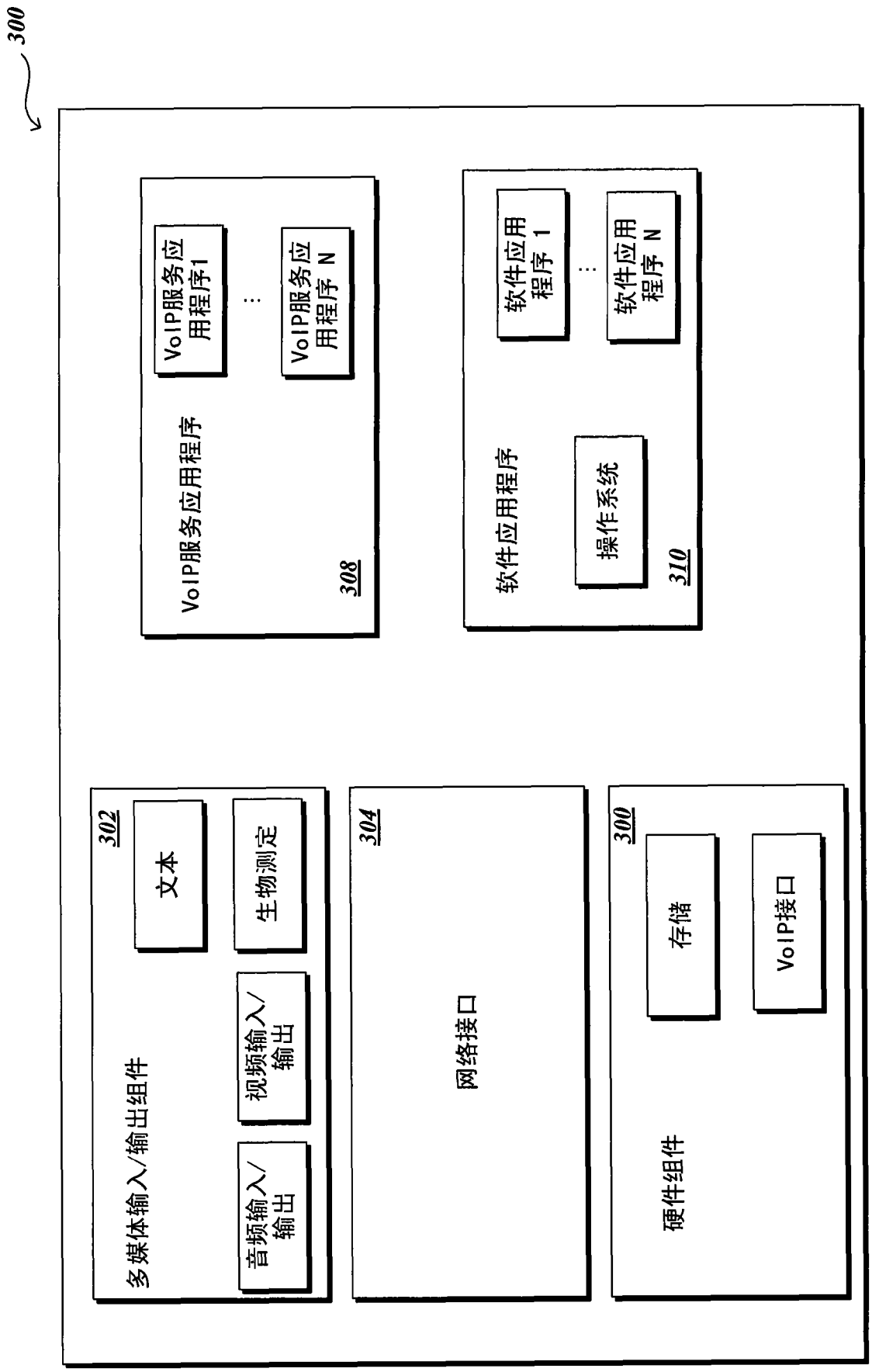

300
VoIP服务应用程序
VoIP服务应用程序1
VoIP服务应用程序 N
308
软件应用程序
软件应用程序 1
软件应用程序 N
操作系统
310
多媒体输入/输出组件
302
文本
生物测定
视频输入/输出
音频输入/输出
网络接口
304
硬件组件
300
存储
VoIP接口

图 3

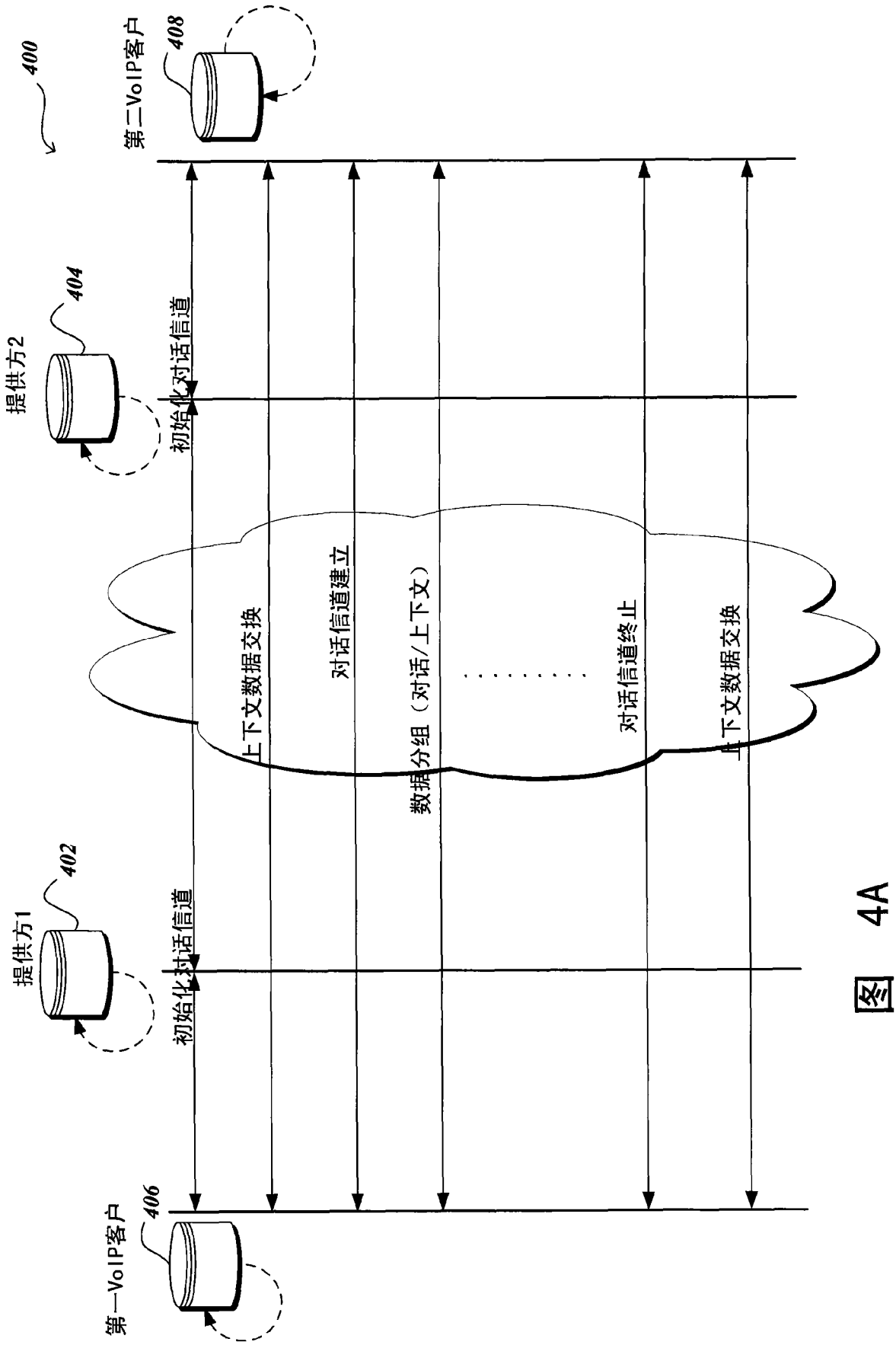

400
第二VoIP客户
408
提供方2
404
初始化对话信道
上下文数据交换
对话信道建立
数据分组（对话/上下文）
对话信道终止
上下文数据交换
提供方1
402
初始化对话信道
第一VoIP客户
406

图 4A

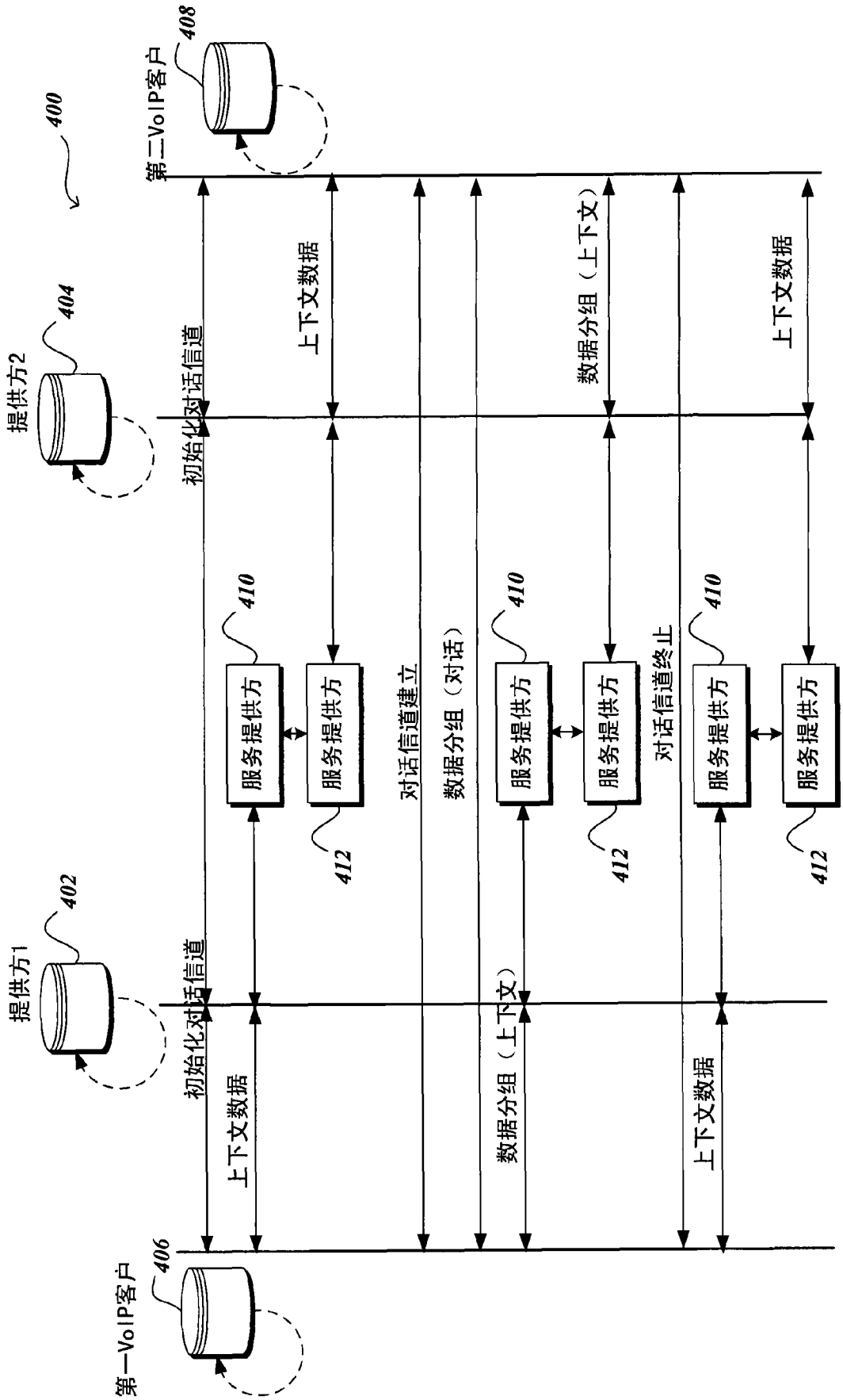

400
第二VoIP客户
408
提供方2
404
初始化对话信道
上下文数据
数据分组（上下文）
上下文数据
410
服务提供方
服务提供方
412
对话信道建立
数据分组（对话）
410
服务提供方
服务提供方
412
对话信道终止
410
服务提供方
服务提供方
412
提供方1
402
初始化对话信道
上下文数据
数据分组（上下文）
上下文数据
第一VoIP客户
406

图 4B

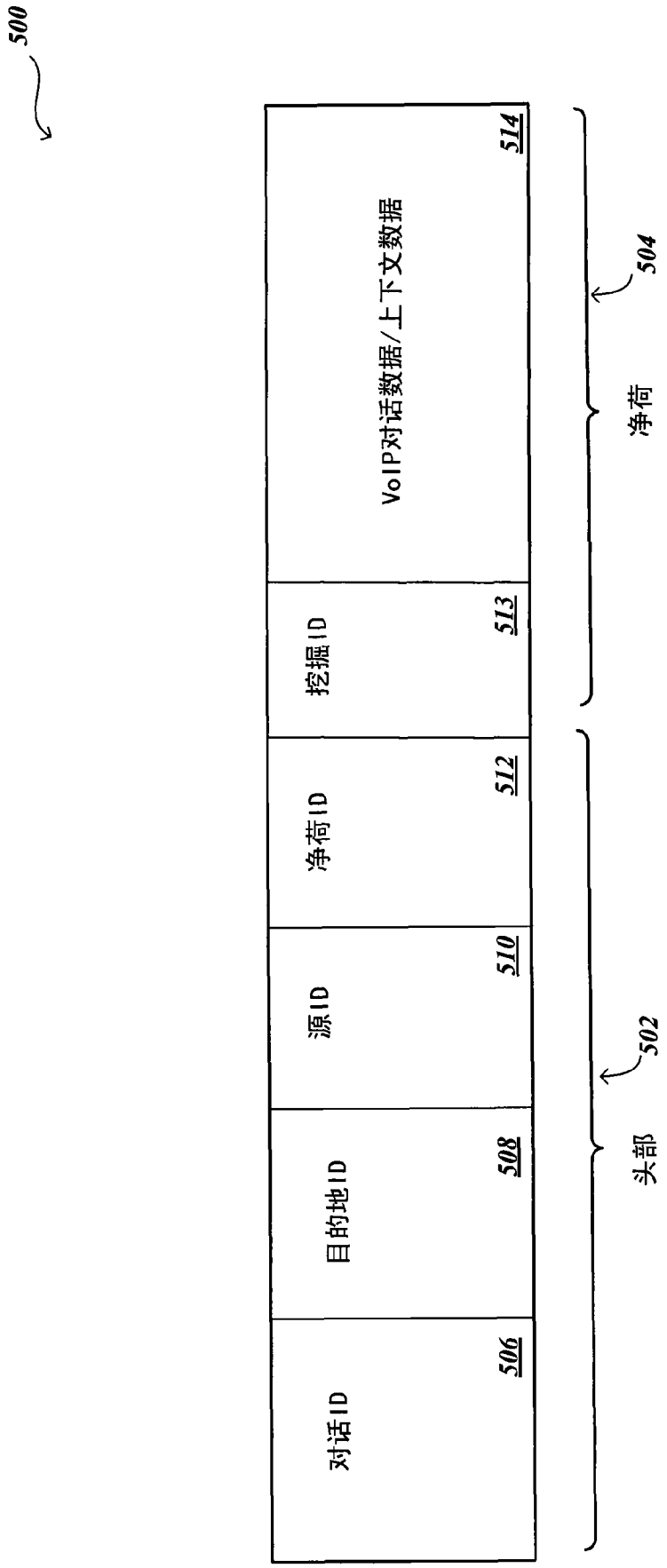

500
对话ID
506
目的地ID
508
源ID
510
净荷ID
512
挖掘ID
513
VoIP对话数据/上下文数据
514
头部
502
净荷
504

图 5

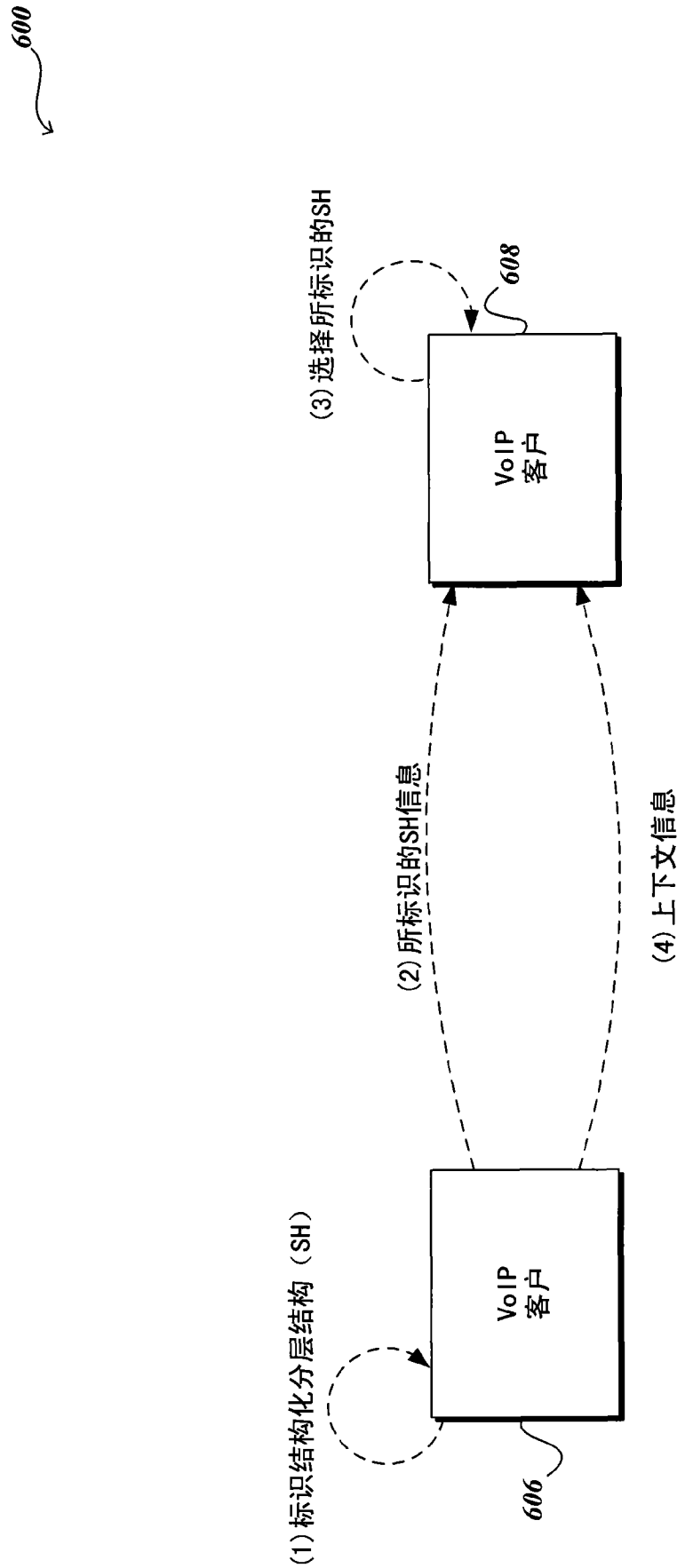

600
(3) 选择所标识的SH
608
VoIP
客户
(2) 所标识的SH信息
(4) 上下文信息
(1) 标识结构化分层结构（SH）
VoIP
客户
606

图 6

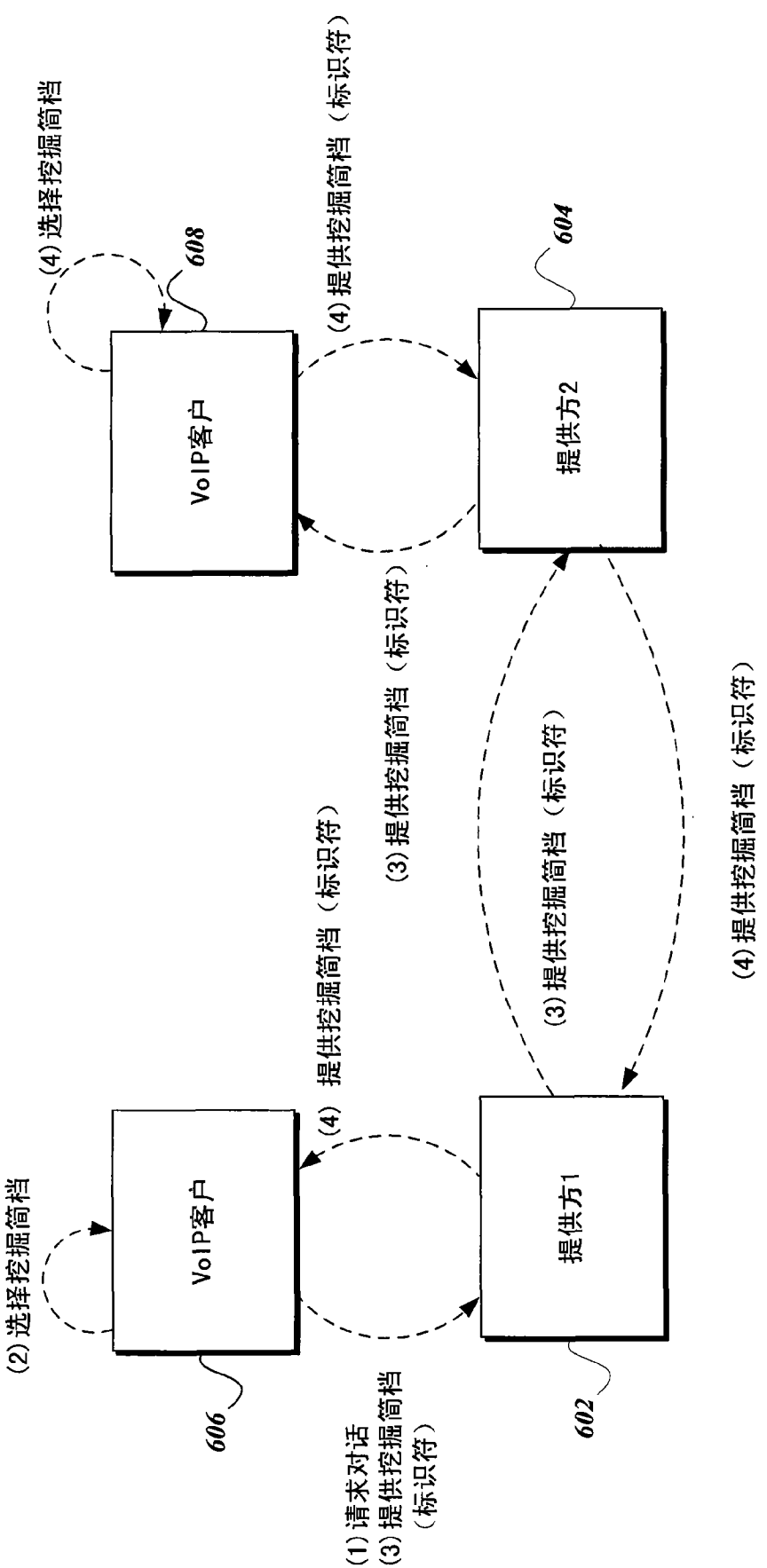

700
(4) 选择挖掘简档
608
VoIP客户
(4) 提供挖掘简档（标识符）
604
提供方2
(3) 提供挖掘简档（标识符）
(3) 提供挖掘简档（标识符）
(4) 提供挖掘简档（标识符）
(4) 提供挖掘简档（标识符）
(2) 选择挖掘简档
VoIP客户
606
(1) 请求对话
(3) 提供挖掘简档（标识符）
602
提供方1

图 7

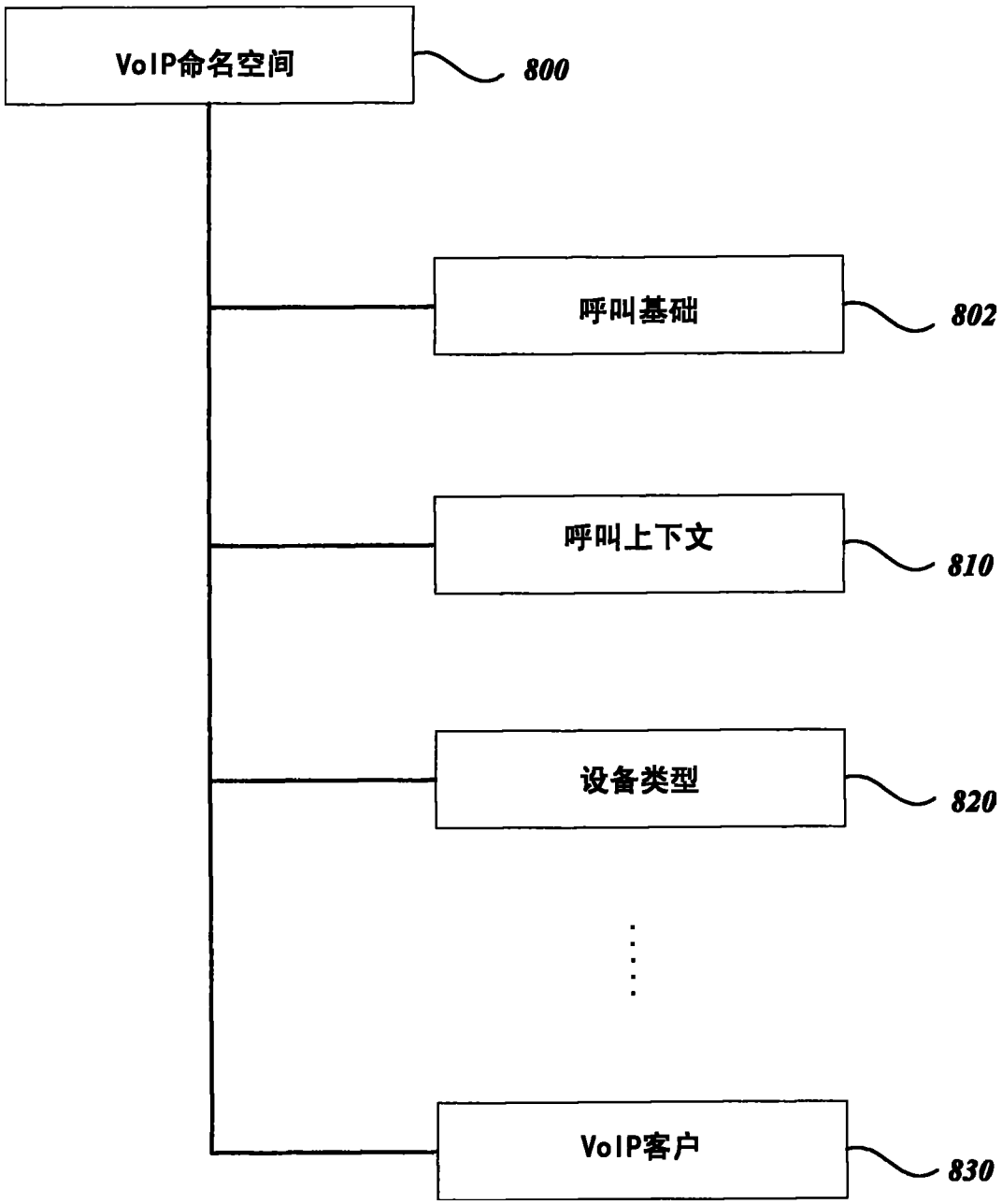


图 8

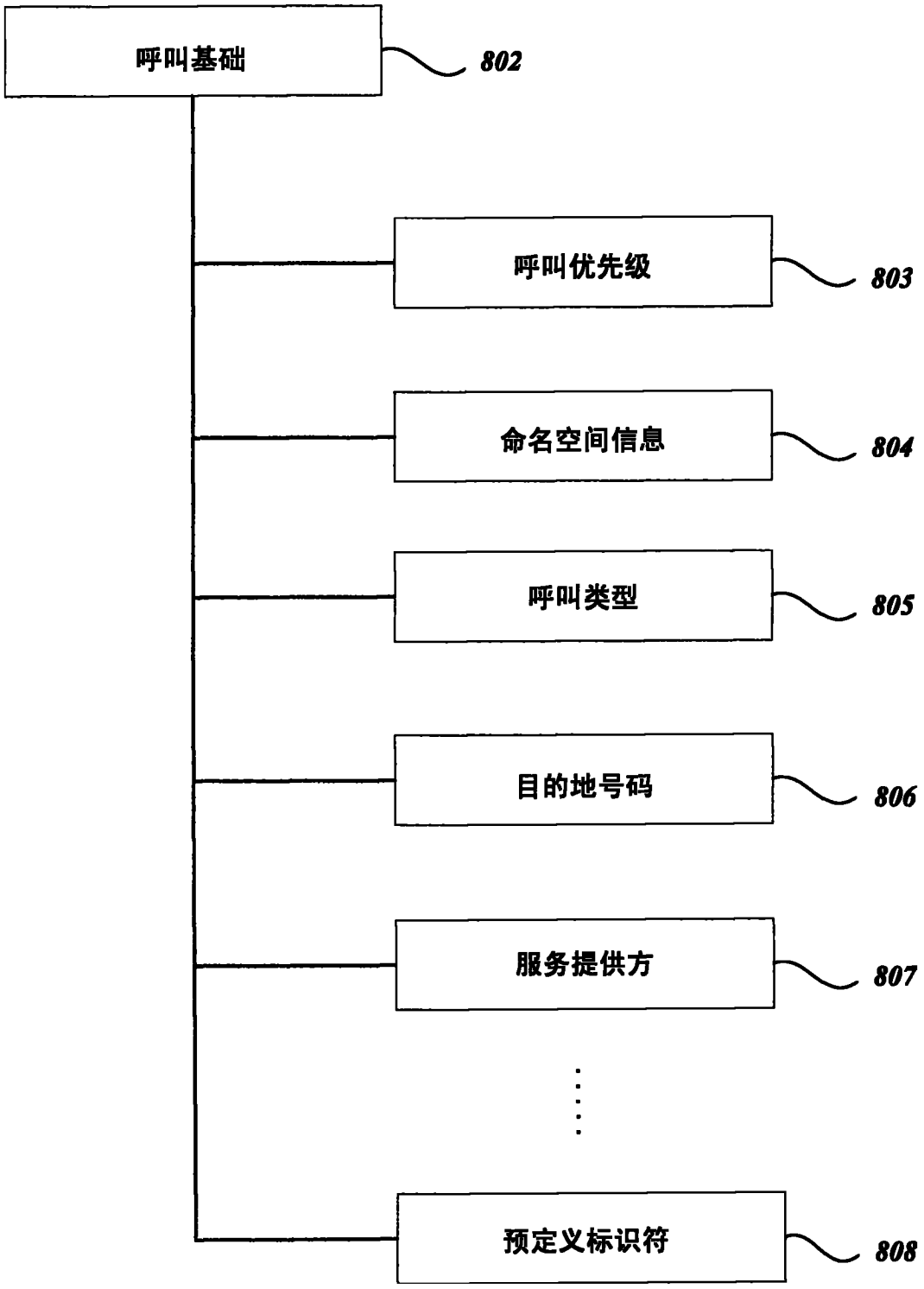

呼叫基础
802
呼叫优先级
803
命名空间信息
804
呼叫类型
805
目的地号码
806
服务提供方
807
预定义标识符
808

图 9

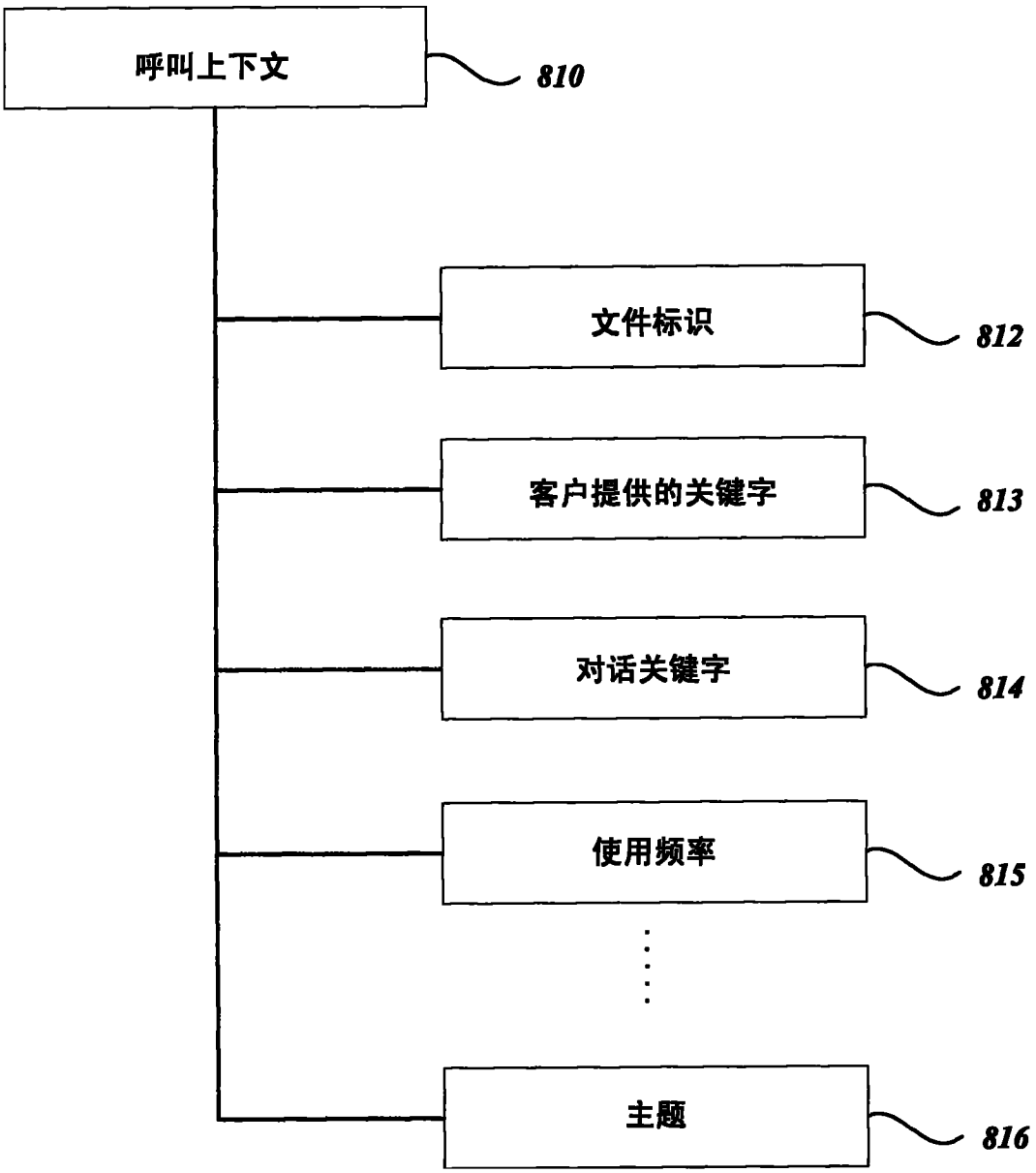


图 10

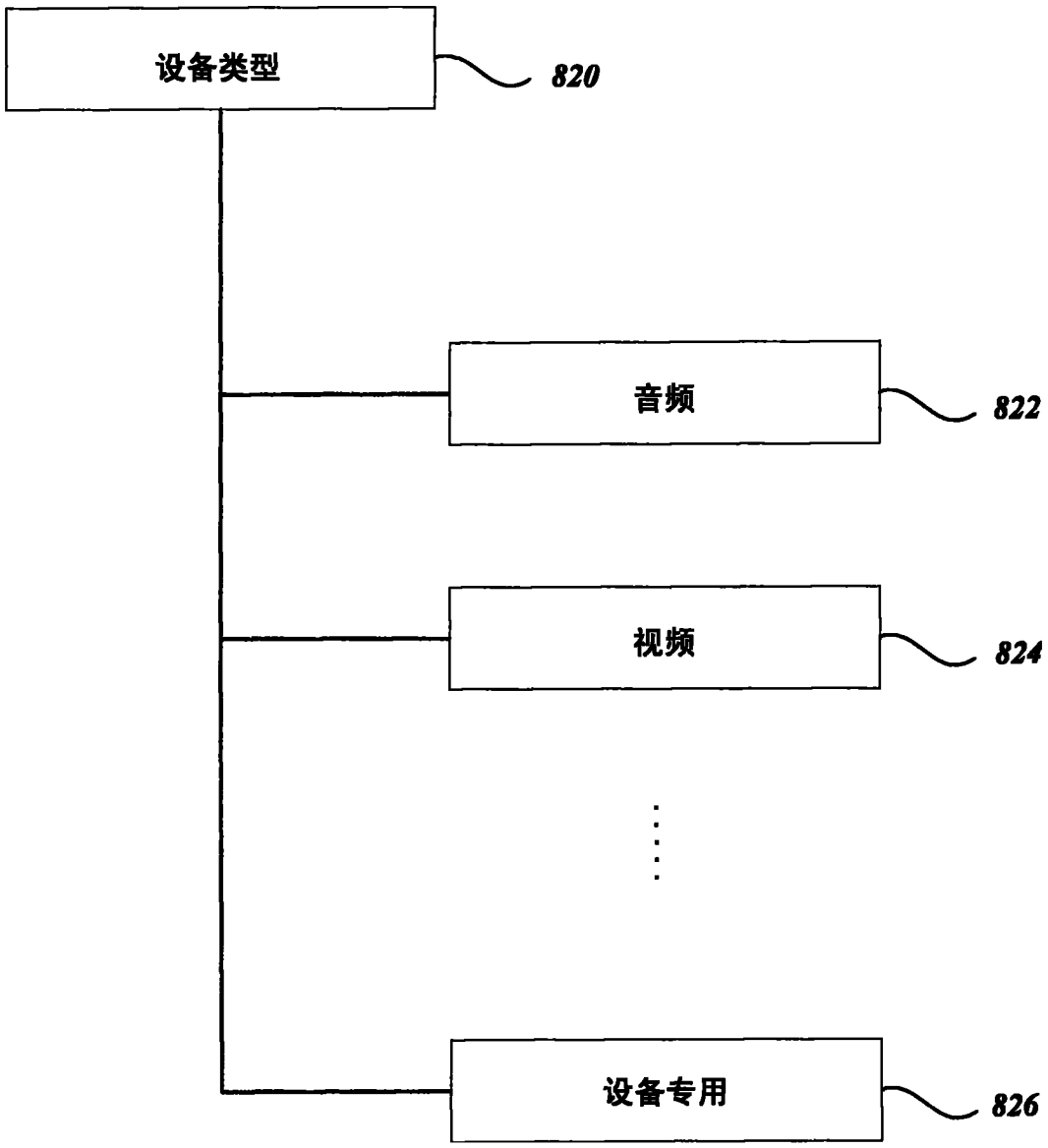


图 11

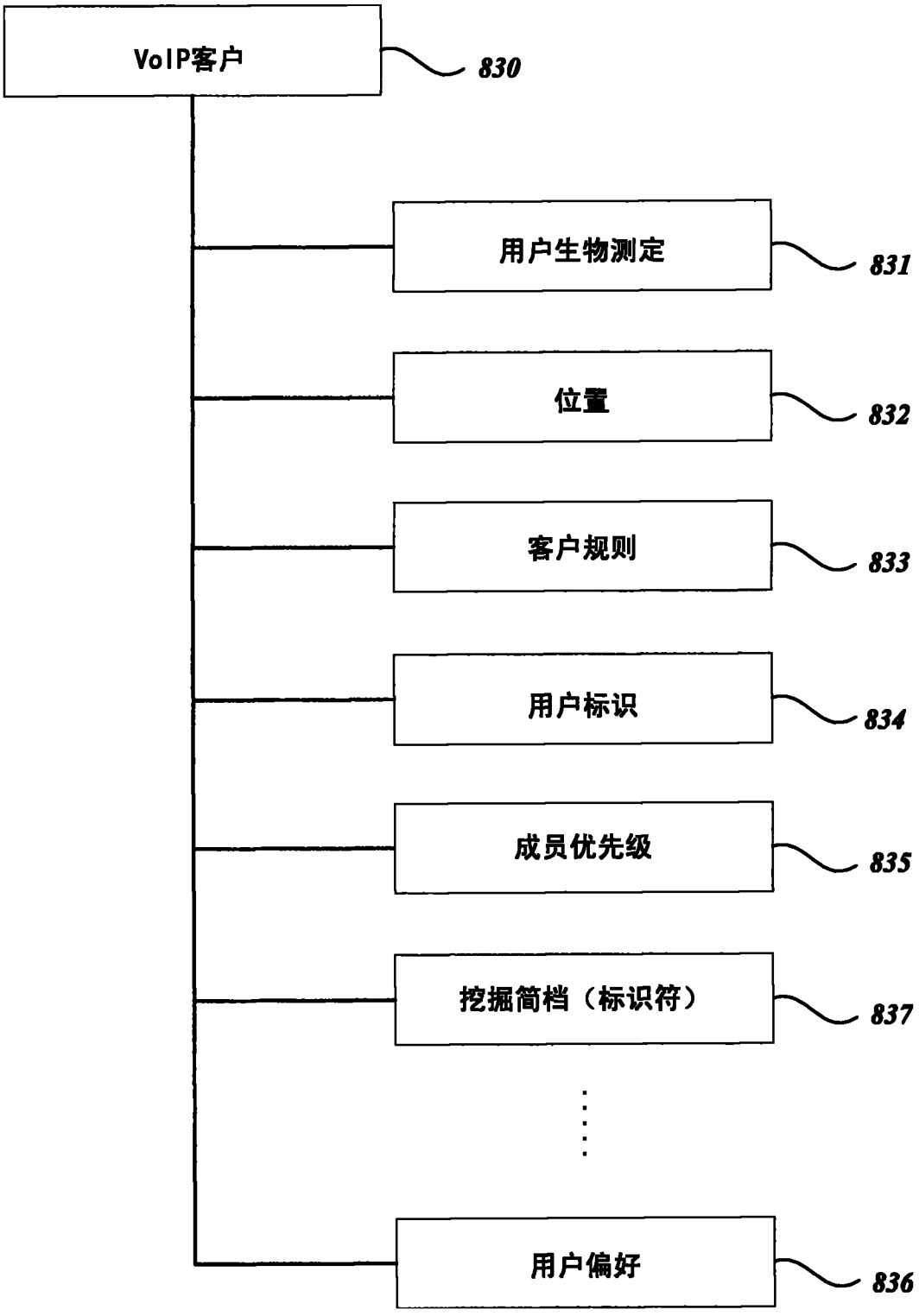

VoIP客户
830
用户生物测定
831
位置
832
客户规则
833
用户标识
834
成员优先级
835
挖掘简档（标识符）
837
用户偏好
836

图 12

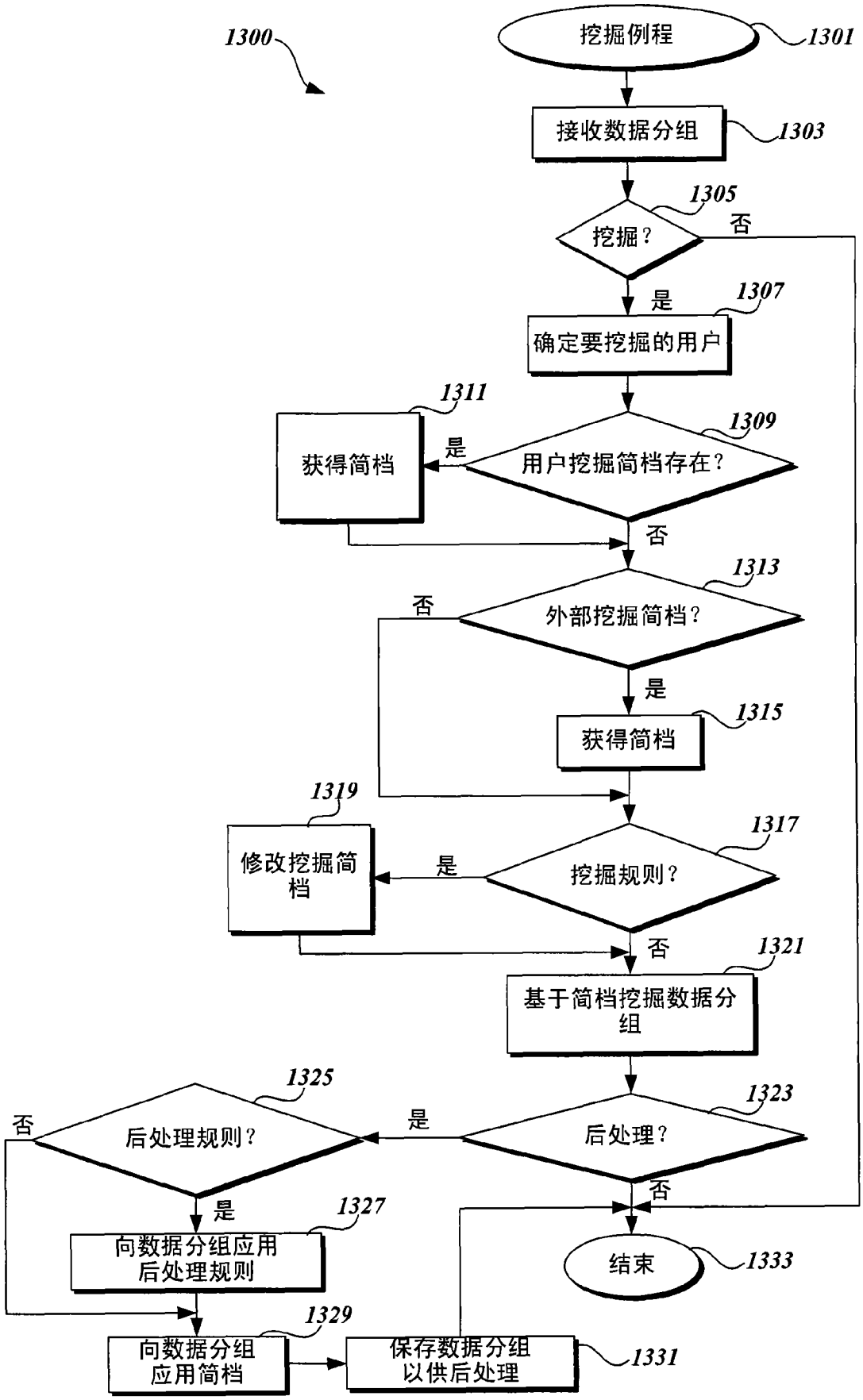

1300
挖掘例程
1301
接收数据分组
1303
1305
挖掘？
否
是
1307
确定要挖掘的用户
1311
获得简档
是
1309
用户挖掘简档存在？
否
1313
否
外部挖掘简档？
是
1315
获得简档
1319
修改挖掘简档
是
1317
挖掘规则？
否
1321
基于简档挖掘数据分组
1325
否
后处理规则？
是
1323
后处理？
否
是
1327
向数据分组应用后处理规则
结束
1333
1329
向数据分组应用简档
保存数据分组以供后处理
1331

图 13

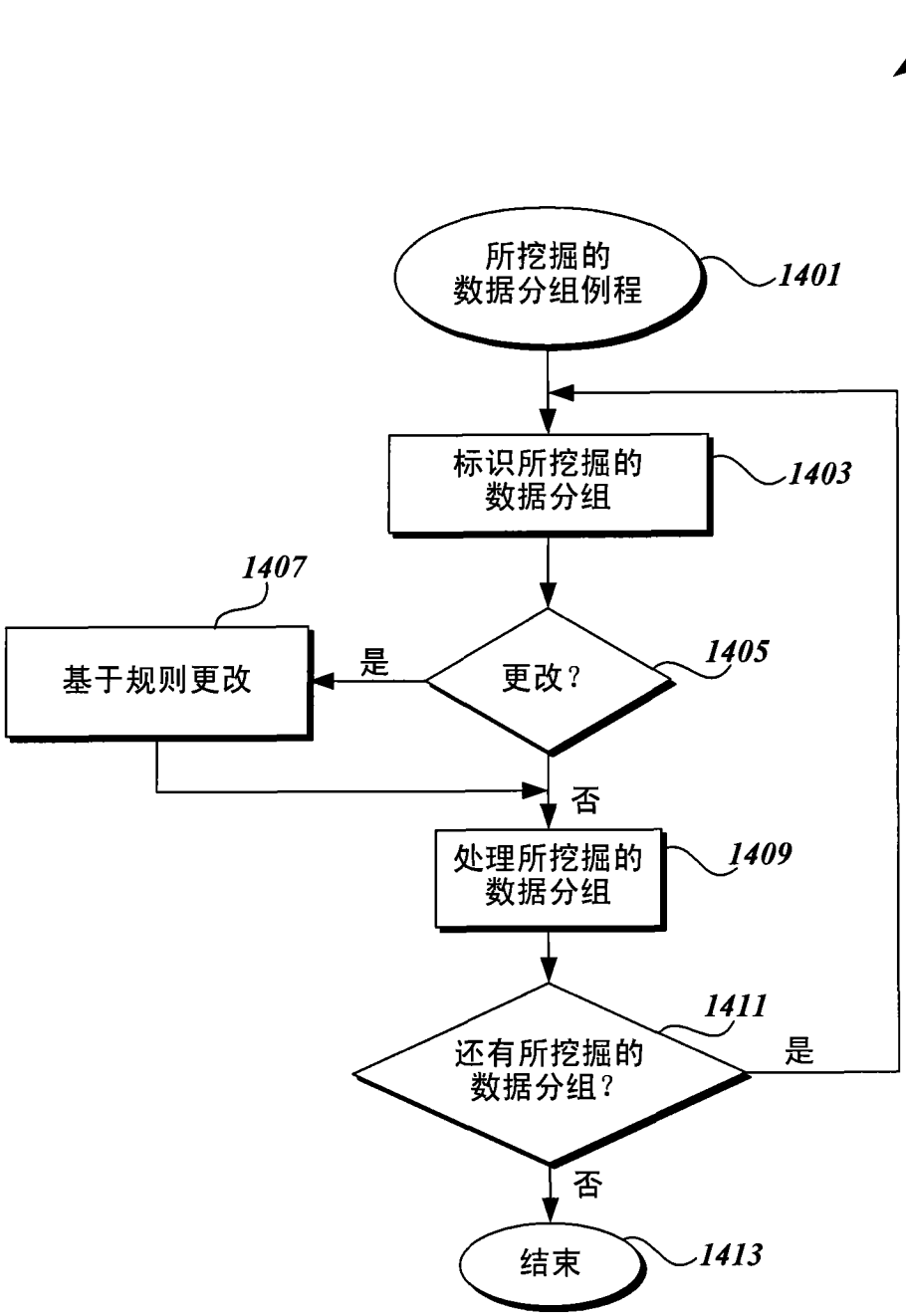

1400
所挖掘的
数据分组例程
1401
标识所挖掘的
数据分组
1403
1407
基于规则更改
是
更改？
1405
否
处理所挖掘的
数据分组
1409
1411
还有所挖掘的
数据分组？
是
否
结束
1413

图 14