

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

H04L 12/58 (2006.01)

G06F 17/30 (2006.01)



# [12] 发明专利申请公布说明书

[21] 申请号 200910038601.0

[43] 公开日 2009 年 9 月 23 日

[11] 公开号 CN 101540739A

[22] 申请日 2009.4.14

[21] 申请号 200910038601.0

[71] 申请人 腾讯科技(深圳)有限公司

地址 518044 广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室

[72] 发明人 丘正元 江宇闻 易文晟 廖海波

[74] 专利代理机构 广州华进联合专利商标代理有限公司

代理人 曾旻辉

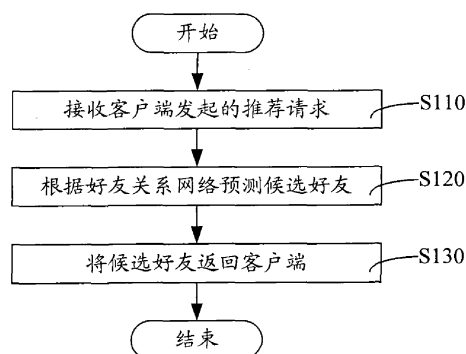
权利要求书 2 页 说明书 10 页 附图 2 页

## [54] 发明名称

用户推荐方法及用户推荐系统

## [57] 摘要

一种用户推荐方法，包括以下步骤：接收用户端发起的推荐请求；根据好友关系网络预测候选好友；将候选好友返回用户端。上述用户推荐方法依据用户之间的好友关系进行预测，不依赖于用户的个人信息，并且能够通过好友关系扩展交际面。此外，还提供了一种用户推荐系统。



1、一种用户推荐方法，其特征在于，包括以下步骤：

接收用户端发起的推荐请求；

根据好友关系网络预测候选好友；

将候选好友返回用户端。

2、根据权利要求1所述的推荐方法，其特征在于，所述根据好友关系网络预测候选好友的步骤包括如下步骤：

根据参选用户与发起请求的用户的紧密程度计算参选用户的分值；

根据参选用户的分值选择所述候选好友。

3、根据权利要求2所述的推荐方法，其特征在于，所述分值的计算采用以下计算方法中的一种或两种以上：共同邻接点数、Jaccard系数、聚类系数、Adamic / Adar系数、Katz方法、相遇时间估计。

4、根据权利要求2所述的推荐方法，其特征在于，所述根据参选用户的分值选择所述候选好友的步骤包括如下步骤：

根据所述参选用户的分值大小选取预设数目的候选用户；

计算所述候选用户成为好友的概率；

根据所述候选用户的概率分布随机地选取预定数量的候选好友。

5、根据权利要求4所述的推荐方法，其特征在于，所述概率的计算方法为用所述候选用户的分值除以所有候选用户的分值之和。

6、根据权利要求5所述的推荐方法，其特征在于，所述根据所述候选用户的概率分布随机地选取预定数量的候选好友的步骤包括如下步骤：

产生预定数量的0到1之间的随机数；

根据所述随机数落入的所述概率分布的范围选取所述候选好友。

7、根据权利要求1至6中任意一项所述的推荐方法，其特征在于，所述推荐请求中包含有限制条件，所述根据好友关系网络预测候选好友的步骤中，在满足所述限制条件的好友关系网络中预测所述候选好友。

8、根据权利要求7所述的推荐方法，其特征在于，所述限制条件为限制候选好友所在的地区、职业、爱好、性别、年龄中的一种或两种以上。

9、根据权利要求7所述的用户推荐方法，其特征在于，所述限制条件为与所述发起请求的用户的好友关系间隔层数。

10、一种用户推荐系统，其特征在于，包括：

推荐管理模块，用于接收用户端发起的推荐请求；

存取模块，用于读取数据库中用户与用户之间的好友关系网络信息；

预测模块，与所述推荐管理模块及存取模块相连，用于根据所述推荐请求及好友关系网络信息预测候选好友并将所述候选好友通过所述推荐管理模块返回所述用户端。

11、根据权利要求10所述的系统，其特征在于，所述预测模块采用权利要求2至6中任意一项所述的根据好友关系网络预测候选好友的步骤预测候选好友。

## 用户推荐方法及用户推荐系统

### 【技术领域】

本发明涉及计算机技术领域，尤其是计算机技术中的用户推荐方法及用户推荐系统。

### 【背景技术】

互联网技术的发展，极大地改变了人们的工作生活。其中，即时通讯技术的出现使人与人之间的沟通变得非常便捷，同时也提供了多样化的沟通方式，例如文字、语音、图片、视频等。

在进行即时通讯之前，如果需要进行即时通讯的对象不在用户的好友列表中，用户需要查找到即时通讯的对象，查找的方式通常有以下几种：

随机查找：当用户发起查找好友的请求时，即时通讯服务器随机地返回一组候选好友供用户选择。例如，在“查找/添加好友”中通过看谁在线上的方式查找好友。随机查找获得的候选好友可以认为是系统从所有在线用户中随机推荐获得的结果。

关键词查找：根据用户输入的关键词，查找与该关键词匹配的候选好友返回供用户选择，关键词查找又可分为以下两种：精确查找：用户输入即时通讯的唯一标识（例如即时通讯号码或者邮件地址）进行好友查找时，即时通讯服务器精确地返回该标识对应的好友；条件查找：条件查找的方式允许用户通过输入个人信息资料来进行好友查找。个人信息资料包括省份、城市、年龄、性别、职业等等。关键词查需要用户输入关键词，不属于系统推荐。

自动查找：将用户自己的某些个人信息与其他用户的个人信息进行匹配，最后返回能匹配上的候选好友。这种好友查找方式无需用户输入查找关键词而自动搜索到候选好友。自动查找获得的候选好友属于系统推荐获得的结果。

即时通讯或社会性网络服务（Social Network Service, SNS）使用的上述查找方式中，随机查找无法返回满足用户特定需求的候选好友；关键词查找和自动查找的方式都是利用用户的个人信息，对于个人信息不完整的用户，在查找

好友的过程中不便使用自动查找功能，而且被查找到的概率也大大降低。

### 【发明内容】

有鉴于此，有必要提供一种不依赖用户的个人信息向用户推荐好友的用户推荐方法。

此外，还有必要提供一种不依赖用户的个人信息向用户推荐好友的用户推荐系统。

一种用户推荐方法，包括以下步骤：接收用户端发起的推荐请求；根据好友关系网络预测候选好友；将候选好友返回用户端。

一种用户推荐系统，包括：推荐管理模块，用于接收用户端发起的推荐请求；存取模块，用于读取数据库中用户与用户之间的好友关系网络信息；预测模块，与所述推荐管理模块及存取模块相连，用于根据所述推荐请求及好友关系网络信息预测候选好友并将所述候选好友通过所述推荐管理模块返回所述用户端。

上述用户推荐方法及系统依据用户之间的好友关系进行预测，不依赖于用户的个人信息向用户推荐好友，并且能够通过好友关系扩展交际面。

### 【附图说明】

图 1 为用户推荐方法的流程图；

图 2 为根据参选用户的分值选择候选好友的流程图；

图 3 为用户推荐系统的模块图；

图 4 为好友关系示意图。

### 【具体实施方式】

在以下实施方式中，通过对存储的已有的用户之间的好友关系进行分析，预测当前还不是好友关系的用户之间成为好友的可能性，并依据这种可能性来推荐一些用户作为某一用户的候选好友。这里的好友关系是指在即时通讯或社会性网络服务（Social Network Service, SNS）中，用户与用户之间已经建立的

联络关系。如图 1 所示，一种用户推荐方法包括如下步骤：

步骤 S110，接收用户端发起的推荐请求。用户端可以是即时通讯工具的用户端，也可以是 SNS 用户的浏览器用户端等。即时通讯用户或 SNS 用户可以通过用户端主动发出该推荐请求，也可以由用户端定时自动发出该请求。

步骤 S120，根据好友关系网络预测候选好友。在即时通讯或 SNS 等网络服务商的存储系统中，通常存储有大量用户与用户之间的好友关系，根据已有的用户之间的好友关系进行分析，预测当前还不是好友关系的用户之间成为好友的可能性，并依据这种可能性来推荐一些用户作为某一用户的候选好友。例如用户 A 的好友中有用户 B，但是没有用户 C，而用户 B 中的好友有用户 C，容易理解，用户 C 成为用户 A 的好友的可能性较大，因此可以将用户 C 作为候选好友推荐给用户 A。当然，在即时通讯或 SNS 网络中的好友关系中，用户与用户之间的好友关系非常复杂，相应的计算也会变得复杂，因此，可以根据需要来预测候选好友。

步骤 S130，将候选好友返回用户端。将预测产生的候选好友发送给用户端，同时可以将用户的个人资料发送给用户端，用户的个人资料可以是以下资料中的一种或多种：好友所在的地区、职业、爱好、性别、年龄等。

更具体地，在步骤 S110 中，该推荐请求可以包括该发起请求的用户的身分识别信息（ID）。用户端发起推荐请求的时机可以是用户登录之时，或用户打开用户推荐显示面板之时，或用户刷新用户推荐显示面板之时等。

另外，该推荐请求中还可以包含有限制条件，在根据好友关系网络预测候选好友的过程中，在满足该限制条件的好友关系网络中预测候选好友。例如，所述限制条件为限制候选好友所在的地区为某一城市，则预测候选好友的过程中只选择所在地区为该城市的用户作为候选好友。限制条件还可以是限制用户职业、爱好、性别、年龄中的一种或多种。

除了上述的依据用户的个人资料进行限制之外，还可以依据用户的好友关系进行限制。例如，所述限制条件为与所述发起请求的用户的好友关系间隔层数。其中，将发起请求的用户（例如用户 A）的好友（例如用户 B）作为第 0

层,将第0层用户(用户B)的好友同时又不是发起请求的用户的好友(用户C)作为第1层,以此类推。显然,层数越低,成为发起请求的用户的好友的可能性越大。通过设置好友关系间隔层数,用户可以设定需要的好友关系网络预测的范围,同时缩小预测的范围,也可以减少计算量,从而加快候选好友返回用户端的速度。

更具体地,在步骤S120中,所述根据好友关系网络预测候选好友是指根据参选用户与发起请求的用户的紧密程度计算参选用户的分值,再根据参选用户的分值选择所述候选好友。其中参选用户是指被纳入预测范围的用户,参选用户可以是即时通讯或SNS服务中所有的非发起请求的用户的好友的用户,也可能是满足上述限制条件的用户。紧密程度是参选用户与发起请求的用户通过其他用户连接的数量及连接的层数。连接的数量越多,连接的层数越少,紧密程度越高。在相同的连接数量下,连接的层数越少,则紧密程度越高;同样,在相同的连接层数下,连接的数量越多,紧密程度也就越高。例如发起请求的用户A的好友中有用户B1、B2,用户B1的好友中有用户C1、C2,用户B2的好友中有用户C1,用户C1中的好友有用户D。显然,C1与A的紧密程度高于C2,因为C1同时通过B1和B2与A连接,连接数量要多;C1与A的紧密程度高于D与A的紧密程度,因为连接的层数少一层。紧密程度越高,分值也就越高,分值越高,被选择成为候选好友的可能性也就越大。

上述分值的计算可以采用以下计算方法中的一种或两种以上:共同邻接点数、Jaccard系数、聚类系数、Adamic/Adar系数、Katz方法、相遇时间估计等。用 $\Gamma(x)$ 表示x的邻接点集(即发起请求的用户端用户x的好友的集合), $\Gamma(y)$ 表示y的邻接点集(即参选用户y的好友的集合)。

x和y之间联系的紧密程度 $\text{score}(x,y)$ 用共同邻接点数来评估的方式为:

$$\text{score}(x,y)=|\Gamma(x)\cap\Gamma(y)|$$

即共同邻接点数评估的方式为求用户端用户x的好友的集合与参选用户y的好友的集合的交集的用户数(其中,公式中的“|”表示取集合中用户的数量),

交集的用户人数越多，说明参选用户与用户端用户共有的好友数量越多，即连接参选用户与用户端用户的数量越多，相应分值也就越高。

$x$  和  $y$  之间联系的紧密程度  $\text{score}(x,y)$  用 Jaccard 系数来评估的方式为：

$$\text{score}(x,y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}$$

即 Jaccard 系数评估的方式为求用户端用户  $x$  的好友的集合与参选用户  $y$  的好友的集合的交集的用户数除以用户端用户  $x$  的好友的集合与参选用户  $y$  的好友的集合的并集的用户数。这种方式可以防止部分参选用户因好友数量特别多而持续获得高分的情况，除以并集的好友数的方法可以降低此类参选用户的分值。

用户  $x$  和  $y$  之间联系的紧密程度  $\text{score}(x,y)$  用聚类系数来评估的方式为：

$$\text{score}(x,y) = \frac{3 \times (\text{包括顶点 } x, y \text{ 的三角形个数})}{\text{包括顶点 } x, y \text{ 的三连接个数}}$$

其中，“包括顶点  $x, y$  的三角形个数”是指添加一条  $(x, y)$  连接后（即假设用户  $x$  和  $y$  建立好友关系后）会形成的包括顶点  $x, y$  的三角形（即用户  $x, y$  和另一用户  $A$  互为好友关系）个数；同样，“包括顶点  $x, y$  的三连接”也是指添加一条  $(x, y)$  连接后会形成包括顶点  $x, y$  的三连接（即  $x$  或  $y$  与另一用户  $z$  的好友关系形成  $z-x-y$  或  $x-y-z$  这样的非互为好友的关系）个数。“包括顶点  $x, y$  的三角形个数”是一种紧密的聚集关系，“包括顶点  $x, y$  的三连接个数”是一种相对松散的聚集关系，聚类系数描述了聚集关系中紧密的聚集关系相对松散的聚集关系所占的比例大小。另外，要形成包括顶点  $x, y$  的三角形，其在加入  $(x, y)$  连接之前应该是已经存在有  $x-A-y$  这样的连接了，所以用户  $x$  和  $y$  之前是已经存在某种间接的关系了，所以其形成新连接的概率会相对比较大。

用户  $x$  和  $y$  之间联系的紧密程度  $\text{score}(x,y)$  用 Adamic / Adar 系数来评估的方式为：

$$\text{score}(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log |\Gamma(z)|}$$



$z$  是同时为  $x$  和  $y$  的邻接点。 $|\Gamma(z)|$  是  $z$  的度数，即好友个数。该方法其实是为了减小度数较大的共同邻接点的影响，使之更关注于度数较小的共同邻接点。

Katz 方法通过加权求和所有长度的路径数，并通过一个系数使评价价值更关注于短路径（即前面描述的层数越少，路径就越短）：

$$\text{score}(x, y) = \sum_{l=1}^{\infty} \beta^l \cdot |\text{path}_{x,y}^{<l>}|$$

其中， $\text{path}_{x,y}^{<l>}$  表示连接  $x$  和  $y$  之间的所有长度为  $l$  的路径的集合。 $\beta$  是预先确定的一个系数，其取值的大小可以根据网络中用户的数量来确定。当  $\beta$  很小时（例如  $\beta$  为 0.001），该定义的度量标准接近于共同邻接点数的度量标准，因为长度大于等于 3（即层数差大于等于 3）的路径的影响已经很小了。

一个节点  $x$  到  $y$  的相遇时间定义为：在网络中从  $x$  出发随机走动，直到遇到  $y$  时所需的走动步数。我们可以记从  $x$  到  $y$  的相遇时间为  $H_{x,y}$ （ $H_{x,y}$  是根据随机过程的方法计算出的一个值）， $H_{x,y}$  越小表示  $x$  到  $y$  的综合距离（即好友关系间隔层数）越小。但另一方面，因为是随机走动，从  $x$  到  $y$  和从  $y$  到  $x$  的相遇时间并不是对称的，因此可以定义两者之间的互通时间(Commute time)为： $C_{x,y} = H_{x,y} + H_{y,x}$ 。我们直观上可以理解  $H_{x,y}$  和  $C_{x,y}$  都在一定程度上反映两节点之间的相似性，因此可定义为  $\text{score}(x, y)$ ，即

$$\text{score}(x, y) = H_{x,y} \text{ 或 } \text{score}(x, y) = C_{x,y}。$$

上述分值的计算方法中，可以单独采用一种，也可以采用多种方法后求平均值，或者赋予每一种计算方法一定的权值再求和。

获得分值之后，在步骤 S120 中，根据参选用户的分值选择候选好友的步骤中可以直接将分值排序后将排在前列的参选用户作为候选好友。由于每次推荐各参选用户获得的分值基本相同，也就是说每次推荐获得的候选用户将会相同（除了上次已经被发起请求的用户加为好友的用户外），这将使得每次推荐没有太多新的变化。为此，在步骤 S120 中，根据参选用户的分值选择候选好友的步

骤中的可以分为如图 2 所示的步骤:

步骤 S122, 根据参选用户的分值大小选取预设数目的候选用户 (例如参选用户的分值的前 100 名作为候选用户);

步骤 S124, 计算候选用户成为好友的概率, 通常分值越高, 概率也就越大;

步骤 S126, 根据候选用户的概率分布随机地选取预定数量的候选好友。

例如, 可以采用如下方式: 假设获得分值排在前 100 名的候选用户, 而只需推荐 10 名候选好友, 则前 20 名中随机推荐 4 名, 前 21-50 名中随机推荐 3 名, 51-100 名中随机推荐 3 名; 另外, 也可采用如下方式: 概率的计算方法为用候选用户的分值除以所有候选用户的分值之和, 用公式表示即为:

$$P(A_i) = \frac{score_i}{\sum_{j=1}^{100} score_j}$$

则  $P(A_1)$ 、 $P(A_2)$ 、 $P(A_3)$ 、...、 $P(A_{100})$  为推荐的概率分布, 即  $A_i$  有  $P(A_i)$  的概率被选取作为被推荐的好友。

再根据以上确定的推荐概率分布随机选择预定数量 (例如 10 名) 作为最后的被推荐好友。简单的实现方法如下:

产生一个  $[0,1]$  之间的随机数  $r$ , 根据所述随机数落入的所述概率分布的范围选取所述候选好友。例如,

若有  $0 \leq r \leq P(A_1)$ , 则选取  $A_1$

若有  $\sum_{j=1}^{k-1} P(A_j) \leq r \leq \sum_{j=1}^k P(A_j)$ , 则选取  $A_k$ , ( $k=2,3,\dots,100$ )

以上方法产生一个随机数就能选择出一个被推荐用户, 用同样的方法进行 10 次, 随机产生 10 个随机数则能选取出 10 个被推荐用户。也可以对  $P(A_i)$  进行简化计算, 对每一个候选潜在好友取相同的概率, 即取  $P(A_i)=1/100$ , 此时以上方法相当于在 100 个候选潜在好友中按相同的概率直接随机选择 10 个作为最后的被推荐好友。

此外，还提供了一种用户推荐系统 200，如图 3 所示，包括：推荐管理模块 210、预测模块 220、存取模块 230 及个人信息模块 240。

推荐管理模块 210 与用户端的推荐显示模块 202 相连，用于接收推荐显示模块 202 发出的推荐请求，并将获得的候选好友的结果发送给推荐显示模块 202，供推荐显示模块 202 显示给用户。

存取模块 230 用于读取好友关系存储模块 204 中存储的数据库中用户与用户之间的好友关系网络信息。

预测模块 220 与所述推荐管理模块 210 及存取模块 230 相连，用于根据所述推荐请求及好友关系网络信息预测候选好友并将所述候选好友通过所述推荐管理模块 210 返回所述用户端。预测模块 220 的预测方式如步骤 S120 所述，不再赘述。

个人信息模块 240 用于在预测模块确定候选好友后，从个人信息存储模块 206 中获得推荐好友的个人信息，供发起请求的用户参考。

用户端与用户推荐系统 200 可以都位于客户端中，用户推荐系统 200 通过网络在服务器端获得好友关系及用户的个人信息。在优选的实施方式中，用户端位于客户端中，用户推荐系统 200 位于服务器端中。

上述用户推荐方法及系统中，利用用户之间的好友关系进行预测，不依赖于用户的个人信息，并且能够通过好友关系扩展现实中的社交关系交际面。

如果将即时通讯中的好友之间是否有现实关系为基础来划分好友关系的话，可以把好友关系分为实关系和虚关系。实关系是指有现实中的社交关系为基础的好友关系；虚关系是指无现实中的社交关系为基础的好友关系，是一种纯粹的网友关系。实关系和虚关系在即时通讯中通常具有不同的表现形式，对用户的影响也不一样。实关系中的两个用户之间不一定会发生很频繁的通讯，但他们之间的通讯具有持久性，用户通常也会比较珍惜这种关系的存在；虚关系中的两方可能在某一时间段内有较频繁的通讯，但他们之间的联系不具有持久性，这种关系的存在与否通常对用户的影响不会很大。

假设用户的好友关系中占主导地位的是虚关系，则用户在注册一个即时通

讯账户后随便添加几个好友，在经过一段时间的聊天后很可能会放弃这个账户，甚至放弃这个即时通讯工具。而如果用户的好友关系中占主导的是实关系，则一个用户放弃使用一个账户必须付出“失去一种跟朋友沟通渠道”的代价。由此使得用户能稳定地使用即时通讯产品。很多即时通讯账户新注册后不能持续活跃，或者很快流失，其很大原因就是因为没有以实关系为基础。因此，实关系对维持用户关系网的稳定是有很大大作用的，能够持久地吸引用户使用即时通讯服务。

事实上，上述用户推荐方法及系统中预测获得的候选好友大多数都是与发起请求的用户具备实关系的用户或者能够与发起请求的用户发展成实关系的用户。如图4所示，假设发起请求的用户A与用户B、C、D是同学关系，但用户D与用户A在网络中还不是好友关系，同时用户B和用户C分别与用户E和用户F是虚关系的好友关系。通常，由于用户E、F是虚关系，其在网络中较难与用户A再有别的网络关系；而用户D与用户A具备实关系的基础（同学关系），可能还通过别的同学建立网络好友关系网络，因此，通过上述用户推荐方法和系统，用户D被推荐作为用户A的候选好友的可能性大，而用户E、F则相对较小。

上述用户推荐方法和系统基于链预测的用户推荐技术能实现以实关系来扩充用户关系网中的好友关系。一方面能满足用户扩大好友圈的需求；另一方面能起到稳定即时通讯服务中的好友关系网的作用。

扩大好友圈：好友关系在一定程度上代表了用户的真实社交关系。每一用户都在其一定的社交圈活动。在朋友关系网络中，朋友的朋友很可能也会成为自己的朋友，使得社交圈不断地扩大。每一个人，通常对相同领域、相同地域、相同爱好的朋友感兴趣。“相同领域”、“相同地域”、“相同爱好”就代表了一个人工作、专业上的社交圈、家庭、生活所在地的社交圈、以及兴趣活动等的社交圈。这些社交圈内的好友关系都是以实关系为基础的。基于链预测技术的用户推荐通过现有好友关系网预测用户的潜在好友关系，通过上面的例子可以看出，这种预测结果所得到的潜在好友关系也有很高的概率属于实关系。因此，能通过上述用户推荐方法和系统让用户以实关系来扩大反映其社交圈的好友

圈，让用户能有更大的机会方便地找到自己现实生活中的好友。

稳定好友关系网：从网络拓扑结构来看，好友关系网是一个无标度网络，它具有无标度特性。无标度特性说明网络中存在有大量的低度节点（即好友个数很少的用户）。度较低的用户在社会关系网中的重要性就会下降，则该即时通讯工具在用户的社交关系中的重要性就会较低，该用户的稳定性也会较低，其放弃使用该即时通讯工具的概率就会较大。要使用户稳定地使用即时通讯工具就必须充实其好友关系中的实关系。基于链预测技术的用户推荐方法及系统能够实现用实关系来扩大用户的好友圈，从而将用户的社交圈较真实地“复制”到网络中的好友关系网中，更大程度地扩大用户之间的交际广度，最终达到稳定好友关系网的作用。

以上所述实施例仅表达了本发明的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干变形和改进，这些都属于本发明的保护范围。因此，本发明专利的保护范围应以所附权利要求为准。

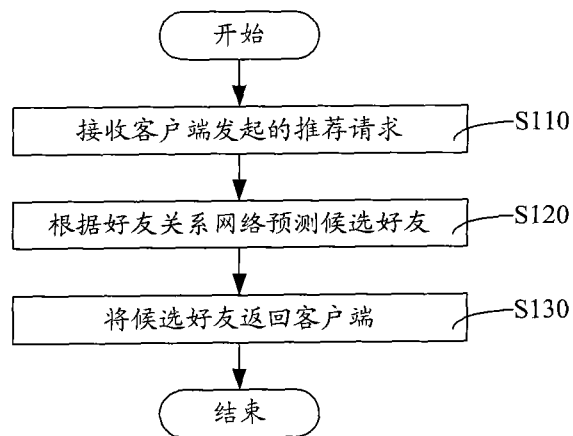


图 1

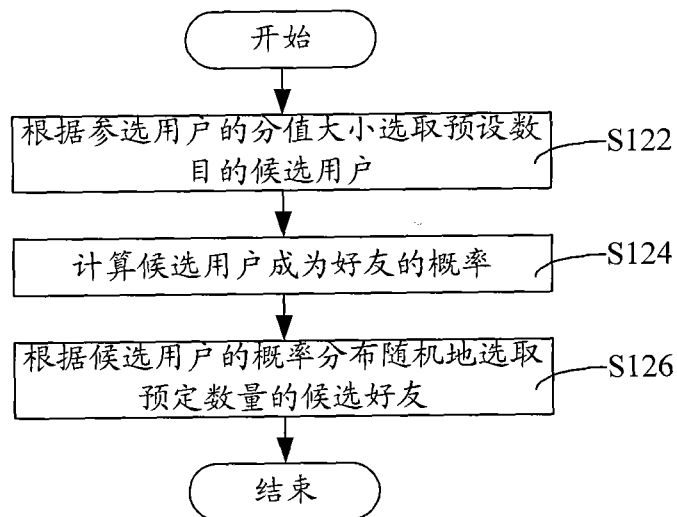


图 2

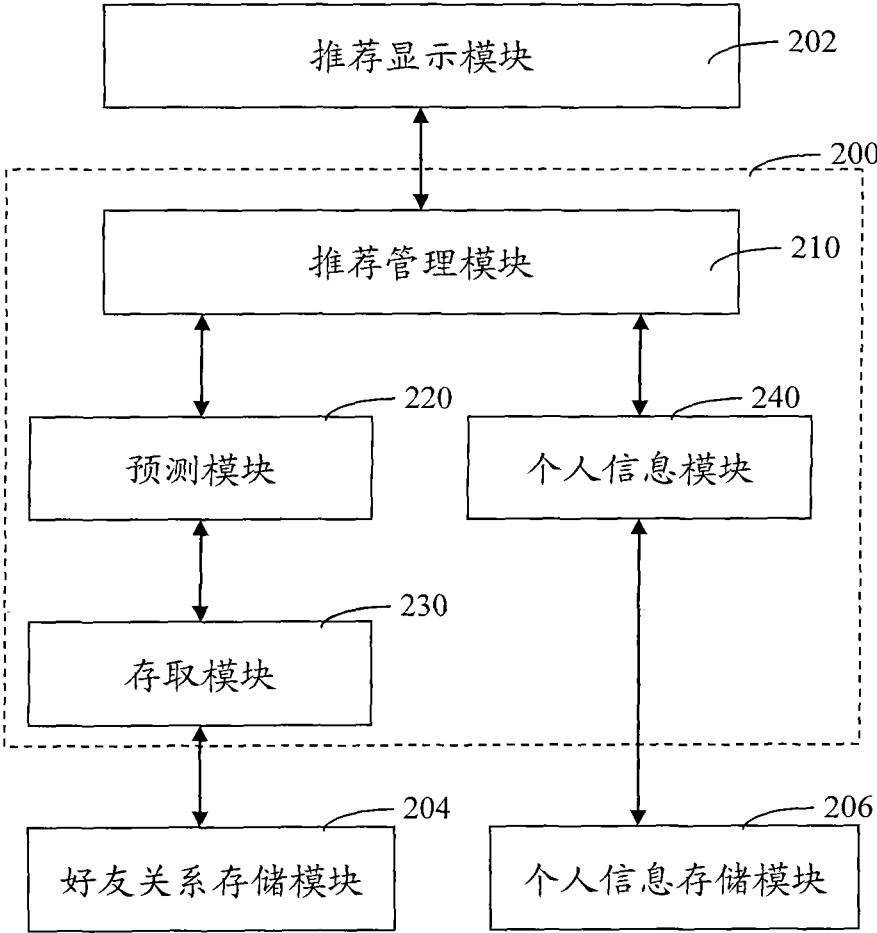


图 3

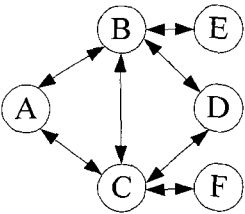


图 4