

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G06N 7/00

H04N 7/16



[12] 发明专利申请公开说明书

[21] 申请号 03813326.1

[43] 公开日 2005 年 9 月 7 日

[11] 公开号 CN 1666227A

[22] 申请日 2003.5.19 [21] 申请号 03813326.1

[30] 优先权

[32] 2002. 6. 10 [33] US [31] 10/165,932

[86] 国际申请 PCT/IB2003/002182 2003.5.19

[87] 国际公布 WO2003/105082 英 2003.12.18

[85] 进入国家阶段日期 2004.12.9

[71] 申请人 皇家飞利浦电子股份有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 A·L·布扎克

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 李亚非 叶恺东

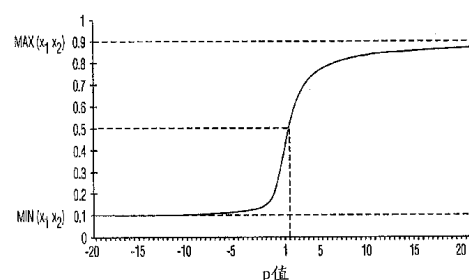
权利要求书 3 页 说明书 8 页 附图 2 页

[54] 发明名称 借助于模糊集连接词所进行的推荐分数的判定融合

[57] 摘要

一种融合推荐分数的方法, 包括以下步骤:

(a) 根据第一信息集来为感兴趣的主题提供一个第一推荐分数; (b) 根据第二信息集为感兴趣的主题提供一个第二推荐分数; (c) 借助于补偿性模糊集连接词来融合第一推荐分数和第二推荐分数; 以及 (d) 基于步骤(c) 的融合来为感兴趣的主题提供最终推荐结果。最终推荐结果可以在显示单元和电视机中的一个设备上输出。用于步骤(c) 中融合的补偿性模糊集连接词可以包括广义平均或伽马模型。在涉及相同主题的同时, 第一和第二推荐分数可以是对于不同人的分数, 其中所述不同的人可以是正在看电视的夫妇。



1. 一种融合推荐分数的方法，包括以下步骤：

(a) 根据第一信息集以及第一方法之一来为感兴趣的主体提供一个第一推荐分数；

5 (b) 根据第二信息集以及第二方法之一来为感兴趣的主体提供一个第二推荐分数；

(c) 借助于补偿性模糊集连接词来融合第一推荐分数和第二推荐分数；以及

10 (d) 基于步骤(c)中的融合来为感兴趣的主体提供最终推荐结果。

2. 根据权利要求1的方法，其中，步骤(b)还包括提供至少一个第三推荐分数，并且步骤(c)包括将所述至少一个第三推荐分数与第一推荐分数以及第二推荐分数相融合。

3. 根据权利要求1的方法，其中，最终推荐结果是在显示单元和电视机中的一个设备上输出的。

4. 根据权利要求1的方法，其中，用于步骤(c)中融合的补偿性模糊集连接词包括广义平均。

5. 根据权利要求4的方法，其中，所述广义平均是根据以下等式确定的：

20
$$g(x_1, x_2, \dots, x_n; p, w_1, w_2, \dots, w_n) = \left(\sum_{i=1}^n w_i x_i^p \right)^{1/p} \quad (1)$$

其中 x_i 是输入， w_i 是加权（重要性因素）， p 是标识与所述输入的并集/交集运算的紧密度的指数。

6. 根据权利要求5的方法，其中， w_i 是由以下等式确定的：

$$\sum_{i=1}^n w_i = 1$$

25 7. 根据权利要求4的方法，其中：

通过改变 p 的值来控制对广义平均的补偿率，这样一来，在 p 值增加的时候，所述运算更接近于并集。

8. 根据权利要求1的方法，其中，用于步骤(c)中融合的补偿性模糊集连接词包括伽马模型。

30 9. 根据权利要求6的方法，其中，所述伽马模型是根据以下等

式确定的：

$$y(x_1, x_2, \dots, x_m) = \left(\prod_{i=1}^m x_i^{\delta_i} \right)^{1-\gamma} \left(1 - \prod_{i=1}^m (1-x_i)^{\delta_i} \right)^{\gamma}, \text{ 其中}$$

x_i 是输入， δ_i 是加权， γ 是用于标识与所述输入的并集/交集运算的紧密度的补偿程度。

- 5 10. 根据权利要求 9 的方法，其中，所述加权是由以下等式确定的，其中：

$$\sum_{i=0}^m \delta_i = m, 0 \leq \gamma \leq 1;$$

m 是输入的数目。

11. 根据权利要求 8 的方法，还包括通过改变 γ 的值来控制伽马模型的补偿率，这样一来，在 γ 的值增加的时候，所述运算更接近于并集。

12. 根据权利要求 2 的方法，其中，用于步骤 (c) 中融合的补偿性模糊集连接词包括伽马模型。

13. 根据权利要求 10 的方法，其中伽马模型是根据以下等式确定的：

$$\sum_{i=0}^m \delta_i = m, 0 \leq \gamma \leq 1.$$

$$\text{其中, } y(x_1, x_2, \dots, x_m) = \left(\prod_{i=1}^m x_i^{\delta_i} \right)^{1-\gamma} \left(1 - \prod_{i=1}^m (1-x_i)^{\delta_i} \right)^{\gamma},$$

并且其中 x_i 是输入， m 是输入数目。

14. 根据权利要求 1 的方法，其中，第一推荐分数和第二推荐分数包括对于电视节目和电影之一的推荐结果。

15. 根据权利要求 1 的方法，其中，第一推荐分数和第二推荐分数包括对于书籍的推荐结果。

16. 根据权利要求 1 的方法，其中，第一推荐分数和第二推荐分数包括对于音乐的推荐结果。

17. 根据权利要求 2 的方法，其中，第一推荐分数、第二推荐分数以及所述至少一个第三推荐分数包括对于电视节目的推荐结果。

18. 根据权利要求 2 的方法，其中，第一推荐分数、第二推荐分数以及第三推荐分数包括对于书籍和音乐之一的推荐结果。

19. 根据权利要求 1 的方法，其中，步骤 (a) 中的第一推荐分

数是为第一个人提供的，步骤（b）中的第二推荐分数是为第二个人提供的。

20. 根据权利要求 2 的方法，其中，步骤（a）中的第一推荐分数是为第一个人提供的，步骤（b）中的第二推荐分数是为第二个人提供的，第三推荐分数则是为第一个人与第二个人中的一个人提供的。

21. 根据权利要求 2 的方法，其中，步骤（a）中的第一推荐分数是为第一个人提供的，步骤（b）中的第二推荐分数是为第二个人提供的，第三推荐分数则是为第三个人提供的。

借助于模糊集连接词所进行的推荐分数的判定融合

本发明涉及的是推荐诸如电视节目之类的感兴趣的内容的方法。更为特别的是，本发明涉及的是使用模糊集连接词 (fuzzy aggregation connective) 而对来自多个推荐器的电视推荐分数进行判定级融合。

现有技术的电视推荐系统是基于观众的显性喜好或从观看历史中推断的他/她的隐性喜好来产生推荐结果的。

10 举例来说，显性推荐器是以特定用户所关注的电视节目的用户定义为基础的。换言之，用户主动向电视推荐系统提供频道、样式、标题之类的用户喜好。此外还存在隐性推荐器，它是根据用户实际观看或未曾观看的节目来推断用户喜好方面的资料的。在本领域已知的是使用基于观看历史来产生推荐结果的技术，例如显性、隐性贝叶斯网络、隐性判定树以及最近邻居分类器。

目前在本领域中已经提出将隐性贝叶斯网络、隐性判定树和借助投票技术的显性判定相结合（参见 S. Gutta、K. Kurapati 和 D Schaffer 在 2001 年 3 月 27 日提交的名为“Combination of Implicit and Explicit Scores of the Recommender Through Voting”的美国专利申请 09/821,277，并且在这里将所述申请的内容作为背景技术而加以引用）。

25 推荐系统可以根据节目或表演的元数据来对所述节目或表演的内容或是内容描述进行分析并产生推荐分数。推荐分数是对用户所做内容评价的估计，它可以用于编译推荐列表，也可以用于自动记录之类的操作。目前已经注意到，不同推荐工具为同一数据集提供的推荐结果通常会稍有不同，其中举例来说，所述数据集可以是给定周的可用节目列表。所生成的推荐结果中的差别则是由于不同推荐工具使用了不同信息造成的，并且这些信息通常是补偿性的。举例来说，从指定用户那里获取的显性信息与那些从用户观看历史中确定的隐性信息是存在很大差别的。

30 此外，不同的推荐机制通常会存在固有的偏向，这将会影响到最终的推荐结果。通过组合来自不同推荐器的推荐分数，可以增强相对

于用户做出的推荐结果。因此，在本领域中需要对来自多个不同电视推荐器的推荐结果进行融合，以使所述推荐器将所述建议基于更类似于人的判定，从而增强为用户提出的推荐结果。此外，在本领域中还需要提出特定的模糊集连接词，以便将其作为一种组合若干推荐结果的方法来执行推荐结果融合，到目前为止，这种处理在本领域还是未知的。

本发明公开的是一种此前在本领域中未知的、用于融合电视推荐器的方法和系统。在本发明中使用了多个模糊集连接词来执行推荐器的融合，从而提供一种增强的效能，以便产生诸如电视节目、所要购买/阅读的书籍、所要观看的电影等内容的最终推荐结果。

根据本发明，在这里使用了补偿性模糊集连接词而对来自单个推荐引擎的推荐结果进行融合。由于补偿性模糊集的连接词的数学属性是对个人倾向进行模拟，从而通过强调一个方面的某些属性来补偿另一个方面的属性的欠缺，因此使用补偿性模糊集连接词来模拟人的判定过程将会产生非常好的结果。

正如更详细描述的那样，本发明是使用模糊集连接词来执行一系列推荐处理的，由此提供了一种此前在本领域中未知并且更为灵活的执行推荐结果融合的方法。这些连接词在不同推荐结果的并集与交集之间提供一个位置。此外，由于投票方案只能执行 n 中选 1、 n 中选 2、 n 中选 k 等类型的功能，因此与使用投票方案相比，本发明提供了更大的灵活性。经由一个简单加权平均来使用模糊集连接词，从而执行一系列推荐处理，这其中的一个优点在于：它们可以对其输出推荐结果之间的不同补偿等级进行模拟，这种模拟是不能通过简单加权平均实现的。

图 1 是描述根据本发明一个实施例并且用作模糊连接词的广义平均 (generalized mean) 的例示。

图 2 是描述根据本发明一个实施例并且用作模糊连接词的伽马模型的例示。

图 3 是根据本发明的基本方法的流程图。

关于观众选择观看哪些节目以及不选择观看哪些节目的判定是一个由人做出的判定。在模拟人的判定过程(并融合这种判定)的时候，较为有利的是使用那些与人的判定相类似的方法。目前已经证

实, 补偿性模糊集连接词能够很好地模拟人的判定过程。由于人们通过强调某一方面的某些属性来补偿另一方面的属性欠缺的倾向, 因此这些连接词可以产生非常好的结果。并且通过模拟人的判定过程, 可以得到更精确的推荐分数。

- 5 应该理解的是, 在这里存在很多融合推荐结果的理由。其中一种情况是存在若干个推荐器用于指定主题。举例来说, 这个主题可以是TV。各个推荐器都使用了不同先验数据来提出推荐结果。举例来说, 第一推荐器是使其推荐结果基于用户提交调查表时陈述的显性兴趣的显性推荐器。第二推荐器是通过使用观众观看历史来计算推荐结果的TV推荐器。
- 10

另一种情况则是第一和第二推荐器全都基于观众的观看历史来进行推荐。然而, 它们是使用不同方法来进行推荐的(举例来说, 第一推荐器使用的是神经网络, 第二推荐器使用的是贝叶斯定理引擎)。

- 15 第三种情况可以是这样一种情况, 其中所述推荐器是为不同人员开发的TV推荐器。每一个推荐器都以某个人的喜好为基础。在这些
- 人希望一起观看的时候, 需要一个最终推荐结果。这个推荐结果是通过融合来自个人的推荐结果而被获取的。

- 在本发明中将特定电视推荐器的判定定义为推荐器预测的观众喜欢观看或者不喜欢观看任何指定电视节目的程度。
- 20

- 然后, 所述判定由模糊集连接词组合在一起。特别地, 根据本发明, 那些经过选择而对推荐器进行融合的模糊集连接词是补偿性的。补偿性的模糊集连接词的实例是本领域普通技术人员所了解的广义平均和伽马模型。在H-J. Zimmermann与P. Zysno发表于Fuzzy Sets and Systems 4第37~51页(1980)的论文“Latent Connectives in Human Decision making”中描述了伽马模型, 并且在H. Dyckhoff与W. Pedrycz发表于Fuzzy Sets and Systems 14, 第143~154页(1984)的论文“Generalized Mean as Model of Compensative Connectives”中描述了广义平均, 在此将所有这些文献作为背景材料加以引用。
- 25
- 30

广义平均和伽马模型连接词具有如下优点, 它们在用交集运算符表征的无补偿极端以及用并集运算符表征的全补偿极端中间提供了

一个位置。在第一种情况中，如果不同来源（推荐器）之间没有补偿，则从各个来源（推荐器）感知判定空间的不同特征。一般来说，在基于若干准则的推荐中（例如电视节目推荐），期望提供一定的补偿量，并且由此补偿性连接词将会很好地描述所述融合处理。

- 5 举例来说，在进行电视节目推荐的时候，某个人想要采用无补偿（用交集运算符表征）极端以及用并集运算符表征的全补偿极端这两个极端之间的某个位置。无补偿意味着信息是补偿性的，全补偿则意味着信息是冗余的。

- 10 当不同信息源（推荐器）之间没有补偿时，判定空间的不同特性是从各个信息源感知的。在基于若干准则的推荐中，期望提供一定的补偿量，并且由此补偿性连接词很好地描述了所述融合过程。

- 模糊集理论是一种用于证物的判定融合/聚集的方法。举例来说，除了交集和并集之外，在这里还可以使用若干连接词来进行聚集。在传统的集合论中，只有并集和交集才可用于进行聚合，但在模糊逻辑中，补偿性连接词具有这样一种属性，即是一个准则的更高满意度可以在某种程度上补偿另一个准则的较低满意度。个人选择的特定连接词不但取决于属性和相关重要性或是准则，而且还取决于所述判定处理所强加的要求。这种要求可以是所有准则全都得到满足，也可以是满足其中任何一个准则。在第一种情况下应该使用一个交集连接词，而在第二种情况下则应该使用并集连接词。在下文中更全面地描述了一种推荐器方法和系统，其中使用了模糊集和融合来为一个或多个用户提供更精确的最终推荐结果。本领域普通技术人员应该理解，在这里可以使用任何补偿性运算符来进行推荐器的融合。此外，特定连接词的选择取决于指定应用所采用的判定策略。以下分别对广义平均和伽马模式进行论述。
- 25

广义平均

广义平均是由以下等式定义的：

$$g(x_1, x_2, \dots, x_n; p, w_1, w_2, \dots, w_n) = \left(\sum_{i=1}^n w_i x_i^p \right)^{1/p} \quad (1)$$

- 30 其中 x_i 是输入， w_i 是加权（重要性因素）， p 是指示与并集/交集运算的紧密度的指数。 p 越小，操作越接近于交集运算。 p 越大，所述

操作则越接近于并集运算。

w_i 可以是用于不同准则的相关重要性因数，其中

$$\sum_{i=1}^n w_i = 1 \quad (2)$$

图 1 是广义平均的例示，在根据本发明的推荐系统中可以将其用作一种补偿性的模糊集连接词。广义平均连接词的行为对应于 $x_1=0.1$ 和 $x_2=0.9$ 的聚集。此外，广义平均的诱人之处在于：

$$\min(a, b) \leq \text{mean}(a, b) \leq \max(a, b);$$

平均值是随着 p 的增加而增大的；通过在负无穷与正无穷之间改变 p 的值，个人可以获取最小值与最大值之间的所有值。在极端情况下，广义平均运算符可被用作交集或并集。此外，通过改变 p ，可以对广义平均的补偿率进行控制。

伽马模型

如图 2 所示，在某些情况下，与其他模型相比，伽马模型提供了与人做出的判定更为接近的匹配。所述伽马模型是如下定义的：

$$y(x_1, x_2, \dots, x_m) = \left(\prod_{i=1}^m x_i^{\delta_i} \right)^{1-\gamma} \left(1 - \prod_{i=1}^m (1-x_i)^{\delta_i} \right)^{\gamma} \quad (3)$$

其中：

$$\sum_{i=1}^m \delta_i = m, 0 \leq \gamma \leq 1. \quad (4)$$

并且其中：

x_i 是输入， m 是输入数目， γ 是补偿度。对 $\gamma=0$ 来说，所述伽马模型将会变成一个交集；对 $\gamma=1$ 来说，所述伽马模型将会变成一个并集。对介于 0 与 1 之间的值来说，所述伽马模型是一个处于交集与并集之间的补偿连接词。 γ 越接近 0，则执行越“类似交集”的运算； γ 越接近 1，则执行越类似并集的操作。

伽马模型是分别以交集和并集的代数表示所著称的乘积与代数和的一个凸组合。在等式 (3) 中，将要聚集的输入 x_i 来自间隔 $(0, 1)$ ， δ_i 是与 x_i 相关联的加权， γ 是对并集与交集部分之间的补偿度进行控制的参数。

为了方便起见,图2中的伽马模型的交集和并集部分分别是由 y_1 和 y_2 表示的:

$$y_1 = y(x_1, x_2, \dots, x_m) = \left(\prod_{i=1}^m x_i^{\delta_i} \right)^{1-\gamma}, \quad \text{和}$$

$$y_2 = \left(1 - \prod_{i=1}^m (1 - x_i)^{\delta_i} \right)^{\gamma}$$

5 实例 1

我们先对以下表示一个家庭如何决策的实例进行查看。

在这个家庭中有一对夫妇,其中每个人都具有一个用于电视节目的单独的推荐系统。在星期一晚间的电视中,有一个频道正在播放“老友记(Friends)”,同时还有一个频道正在上演帕瓦罗蒂的歌剧。那么系统更倾向于推荐观看哪一个节目呢?以下的表1给出了某种洞悉。

表 1:

	丈夫	妻子	W1	W2	P	加权平均	广义平均
帕瓦罗蒂	0.1	0.99	0.5	0.5	2	0.545	0.704
老友记	0.49	0.6	0.5	0.5	2	0.545	0.548

上面的表1显示了丈夫和妻子的推荐分数。对于这两个推荐结果的加权平均都是相同的0.545。

相应地,即使加权平均相等,系统也必须从“帕瓦罗蒂”与“老友记”之间选出更推荐观看的节目。然而在这里,广义平均分数(对于 $p=2$)是存在差别的:对帕瓦罗蒂来说,广义平均分数是0.704(因为其中一个人非常喜欢它(妻子具有大小为0.99的分数),以至于不能同意不观看所述节目),而老友记的广义平均分数仅为0.548(因为此二人对这个节目都很“感兴趣”,例如夫妇二人分别具有0.49和0.6的分数),但是没有人对此节目有极强的意见。因此,某个人越想观看节目,那么广义平均的值也就越大(当 $P=2$ 时)。这样一来,根据本发明,与老友记相比,帕瓦罗蒂将会得到更多的推荐。在表2中例示了另一个家庭如何就所要观看的电视节目做出判定的实例。

表 2:

	丈夫	妻子	W1	W2	p	加权平均	广义平均
帕瓦罗蒂	0.1	0.99	0.5	0.5	-0.5	0.545	0.230
老友记	0.49	0.6	0.5	0.5	-0.5	0.545	0.541

对丈夫和妻子来说，帕瓦罗蒂和老友记的分数与表 1 的先前所示是相同的，并且每一个节目的加权平均也是相同的。然而，这个家庭喜欢多数人意见形式的判定，也就是说，它们使用这种判定来选择观看没有被任何人低评的节目，在这种情况下，用于广义平均的指数 p 是 0.5，这意味着这个判定是一个更多以交集为基础的操作。对帕瓦罗蒂来说，广义平均结果仅为 0.23，而对老友记来说，广义平均结果是相对较高的 0.541。因此更倾向于向这对夫妇推荐老友记。

图 3 是本发明的方法的概述的流程图。

在步骤 305，根据 TV 观看历史之类的主题信息而为感兴趣的主体提供了第一推荐分数；可替换地，第一推荐分数可以对应于第一个人（类似于实例中的丈夫和妻子）。此外，所述感兴趣的主体也可以是电视、电影、音乐、书籍、餐厅等等。

在步骤 310 中，根据不同的信息集（例如电影进行历史）而为感兴趣的同一主题提供了一个第二推荐分数。可替换地，第二推荐分数也可以使用相同的信息集，但是使用不同的推荐引擎；此外，第二推荐分数也可以对应于第二个人（与先前实例一样）。

在步骤 315 中，通过使用模糊集连接词来融合第一和第二推荐分数。其中举例来说，模糊集连接词的类型可以是广义平均或伽马模型。

最后，在步骤 320 中提供来自步骤 315 的融合的最终推荐结果。因此，通过使用模糊集连接词来进行融合，可以提供精度极大提升的推荐结果，这是因为如先前实例所述，在这里还可以包含其他因素，例如不能分解成带有任何精度的投票方案的两个人观看电视的喜好度，例如，即使对这两个人来说，加权平均可能表示的是相同的推荐分数，也希望就某个没有人会不喜欢的节目而达成一致。通过对这些因素进行量化并且通过使用模糊集连接词来融合第一和第二推荐分数，从而提供最终推荐结果，这种处理到目前为止仍然是未知的，该处理更准确地描述了由人做出的判定。

- 应该指出的是，在不脱离本发明的实质和所附权利要求范围的情况下，可以进行不同的修改。举例来说，模糊集融合的内容可以不同于上文提到的内容，其中包括但不限于运动、消费品购买（例如衣服、电器、珠宝、耐用和非耐用商品）。在不脱离请求保护的发明的实质和范围的情况下，可以对执行广义平均或伽马模型的实际方法进行较小的改变。
- 5

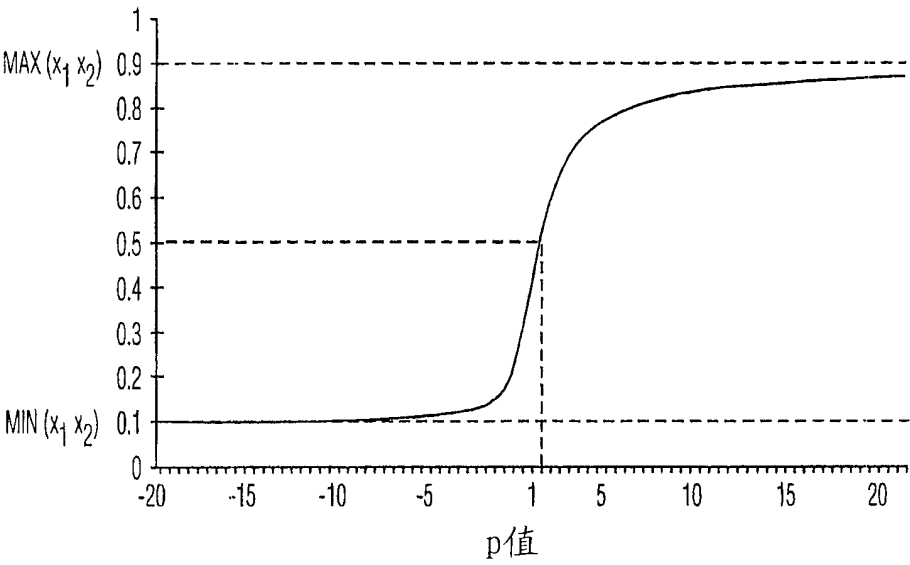


图 1

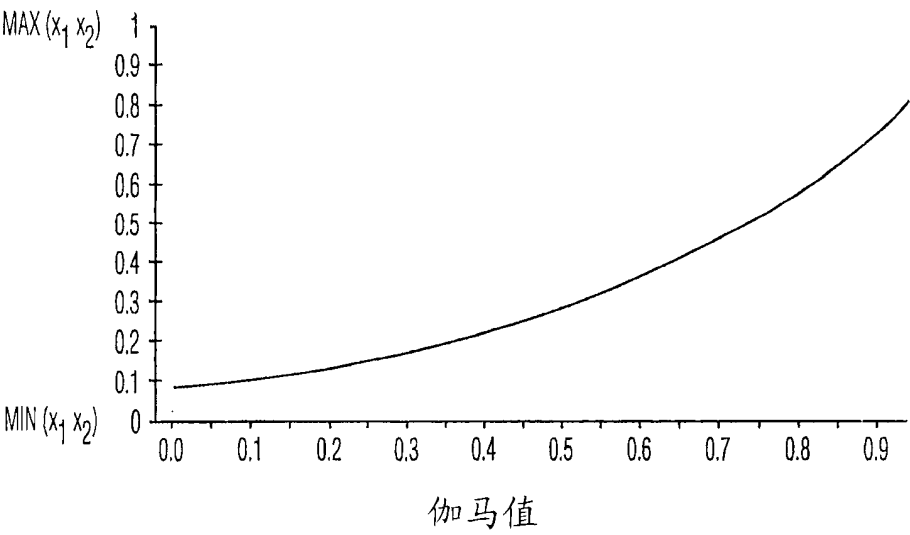


图 2

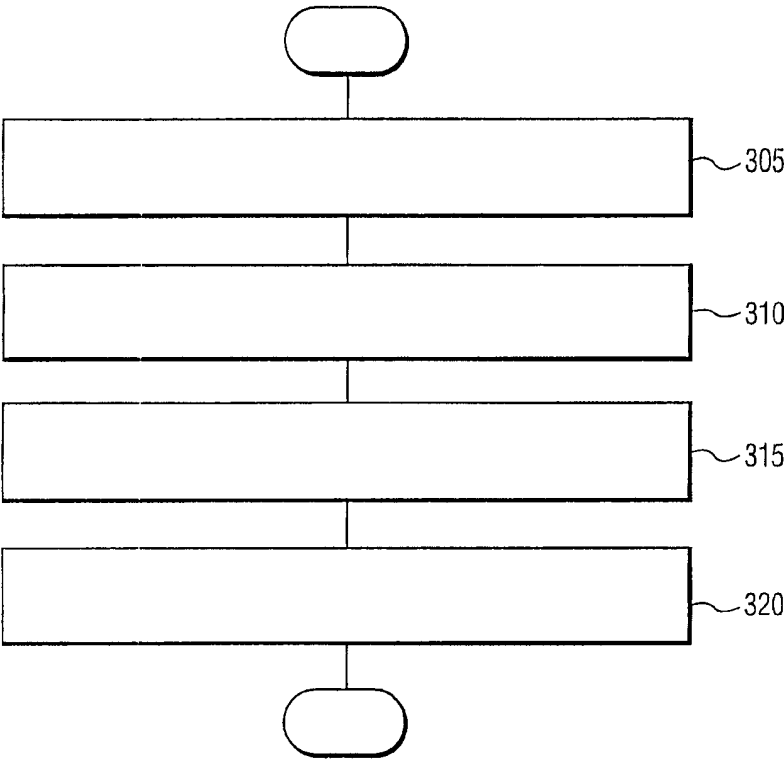


图 3