



(12)发明专利

(10)授权公告号 CN 102841913 B

(45)授权公告日 2017.03.01

(21)申请号 201210147815.3

(51)Int.Cl.

(22)申请日 2012.05.11

G06F 17/30(2006.01)

(65)同一申请的已公布的文献号

申请公布号 CN 102841913 A

(56)对比文件

CN 101577818 A,2009.11.11,

CN 101059815 A,2007.10.24,

(43)申请公布日 2012.12.26

US 2007136274 A1,2007.06.14,

(30)优先权数据

2011-111644 2011.05.18 JP

US 2007073708 A1,2007.03.29,

审查员 谭明敏

(73)专利权人 索尼公司

地址 日本东京都

(72)发明人 金本胜吉 宫寄充弘 萩原丈博

右田隆仁 增田弘之 藤田拓也

森田昌裕

(74)专利代理机构 北京集佳知识产权代理有限

公司 11227

代理人 王萍 陈炜

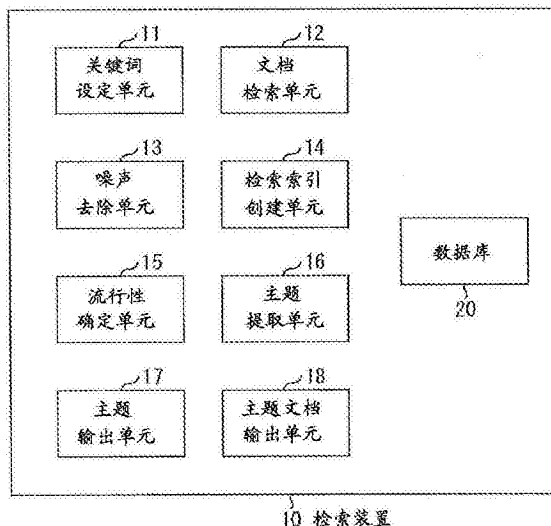
权利要求书1页 说明书10页 附图12页

(54)发明名称

信息处理装置、信息处理方法和程序

(57)摘要

本公开涉及信息处理装置、信息处理方法和程序。根据本公开的信息处理装置包括:获取单元,获取包括测量时段i中的采样值 x_i 的离散系统的时间序列数据;移动偏差计算单元,基于与预定测量时段t之前的预定时段对应的N个采样值 $x_t, x_{t-1}, x_{t-2}, \dots$ 和 x_{t-N+1} 的移动平均值 m_t 计算移动偏差 v_t ;以及评估值计算单元,基于与测量时段t对应的移动偏差 v_t 和与测量时段t-1对应的移动偏差 v_{t-1} ,计算表示测量时段t中的离散系统的时间序列数据的迅速改变的评估值 s_t 。



1. 一种信息处理装置,包括:

获取单元,获取包括测量时段 i 中的采样值 x_i 的离散系统的时间序列数据;

移动偏差计算单元,基于与预定测量时段 t 之前的预定时段对应的 N 个采样值 x_t 、 x_{t-1} 、 x_{t-2} 、 \dots 、和 x_{t-N+1} 的移动平均值 m_t 计算移动偏差 v_t ;以及

评估值计算单元,基于与测量时段 t 对应的移动偏差 v_t 和与测量时段 $t-1$ 对应的移动偏差 v_{t-1} ,计算表示测量时段 t 中的所述离散系统的时间序列数据的迅速改变的评估值 s_t 。

2. 根据权利要求1所述的信息处理装置,

其中所述评估值计算单元计算所述评估值 $s_t = \text{移动偏差 } v_t / \text{移动偏差 } v_{t-1}$ 。

3. 根据权利要求2所述的信息处理装置,

其中所述评估值计算单元对每个测量时段的连续时间序列数据求和并且将所述连续时间序列数据转换为所述离散系统的时间序列数据。

4. 根据权利要求3所述的信息处理装置,

其中所述评估值计算单元将所述测量时段设定为在时间上重叠,并且对每个测量时段的连续时间序列数据求和并且将所述连续时间序列数据转换为所述离散系统的时间序列数据。

5. 一种由信息处理装置执行的信息处理方法,包括:

获取包括测量时段 i 中的采样值 x_i 的离散系统的时间序列数据;

基于与预定测量时段 t 之前的预定时段对应的 N 个采样值 x_t 、 x_{t-1} 、 x_{t-2} 、 \dots 、和 x_{t-N+1} 的移动平均值 m_t 计算移动偏差 v_t ;以及

基于与测量时段 t 对应的移动偏差 v_t 和与测量时段 $t-1$ 对应的移动偏差 v_{t-1} ,计算表示测量时段 t 中的所述离散系统的时间序列数据的迅速改变的评估值 s_t 。

信息处理装置、信息处理方法和程序

技术领域

[0001] 本公开涉及信息处理装置、信息处理方法和程序,具体地,涉及使得能够向用户提供与检索关键词相关联的信息的信息处理装置、信息处理方法和程序。

背景技术

[0002] 传统上,除了网页和博客,互联网上充斥了使用以推特(Twitter)为代表的各种社交网络服务(SNS)的多种信息。此外,从多种信息中提取包括任意关键词的信息的系统是已知的。

[0003] 具体地,在使用现有检索系统的情况下,使用用户任意设定的关键词作为检索条件,可以向用户提供包括检索条件的信息。此外,根据包括检索关键词的信息的新鲜度或者检索频率,可以向用户提供新信息或者频繁检索的信息(例如,参照日本公开专利公布第2009-15407号)。

发明内容

[0004] 如上所述,在相关技术中可以检索包括检索关键词的信息。然而,仍未提出用于提供与检索关键词相关联的信息(可以不包括检索关键词)或者从与检索关键词相关联的信息中提取已成为世界流行主题的信息的技术。

[0005] 鉴于上述情况设计了本公开,并且本公开能够提取已变为世界流行主题的信息。

[0006] 根据本公开的一个实施例,提供了一种信息处理装置,该信息处理装置包括:获取单元,获取包括测量时段 i 中的采样值 x_i 的离散系统的时间序列数据;移动偏差计算单元,基于与预定测量时段 t 之前的预定时段对应的 N 个采样值 $x_t, x_{t-1}, x_{t-2}, \dots$ 和 x_{t-N+1} 的移动平均值 m_t 计算移动偏差 v_t ;以及评估值计算单元,基于与测量时段 t 对应的移动偏差 v_t 和与测量时段 $t-1$ 对应的移动偏差 v_{t-1} ,计算表示测量时段 t 中的离散系统的时间序列数据的迅速改变的评估值 s_t 。

[0007] 评估值计算单元可以计算评估值 $s_t = \text{移动偏差 } v_t / \text{移动偏差 } v_{t-1}$ 。

[0008] 评估值计算单元可以对每个测量时段的连续时间序列数据求和并且将连续时间序列数据转换为离散系统的时间序列数据。

[0009] 评估值计算单元可以将测量时段设定为在时间上重叠,并且对每个测量时段的连续时间序列数据求和并且将连续时间序列数据转换为离散系统的时间序列数据。

[0010] 根据本公开的另一实施例,提供了一种由信息处理装置执行的信息处理方法,该信息处理方法包括:获取包括测量时段 i 中的采样值 x_i 的离散系统的时间序列数据,基于与预定测量时段 t 之前的预定时段对应的 N 个采样值 $x_t, x_{t-1}, x_{t-2}, \dots$ 和 x_{t-N+1} 的移动平均值 m_t 计算移动偏差 v_t ,并且基于与测量时段 t 对应的移动偏差 v_t 和与测量时段 $t-1$ 对应的移动偏差 v_{t-1} ,计算表示测量时段 t 中的离散系统的时间序列数据的迅速改变的评估值 s_t 。

[0011] 根据本公开的另一实施例,提供了一种使计算机用作如下单元的程序:获取单元,获取包括测量时段 i 中的采样值 x_i 的离散系统的时间序列数据;移动偏差计算单元,基于与

预定测量时段 t 之前的预定时段对应的 N 个采样值 $x_t, x_{t-1}, x_{t-2}, \dots$ 和 x_{t-N+1} 的移动平均值 m_t 计算移动偏差 v_t ;以及评估值计算单元,基于与测量时段 t 对应的移动偏差 v_t 和与测量时段 $t-1$ 对应的移动偏差 v_{t-1} ,计算表示测量时段 t 中的离散系统的时间序列数据的迅速改变的评估值 s_t 。

[0012] 根据上面描述的本公开的各实施例,获取包括测量时段 i 中的采样值 x_i 的离散系统的时间序列数据,基于与预定测量时段 t 之前的预定时段对应的 N 个采样值 $x_t, x_{t-1}, x_{t-2}, \dots$ 和 x_{t-N+1} 的移动平均值 m_t 计算移动偏差 v_t ,并且基于与测量时段 t 对应的移动偏差 v_t 和与测量时段 $t-1$ 对应的移动偏差 v_{t-1} ,计算表示测量时段 t 中的离散系统的时间序列数据的迅速改变的评估值 s_t 。

[0013] 根据上面描述的本公开的各实施例,可以提取已变为世界流行主题的信息。

附图说明

[0014] 图1是图示作为本公开的一个实施例的检索装置的配置示例的框图;

[0015] 图2是图示数据库的详细配置的框图;

[0016] 图3是图示检索装置执行的相关联的信息检索处理的流程图;

[0017] 图4是图示噪声去除的示意图;

[0018] 图5是图示主题提取处理的流程图;

[0019] 图6是图示主题候选字符串的示图;

[0020] 图7是图示变为检索装置的用户接口的屏幕的显示示例的示图;

[0021] 图8是图示变为检索装置的用户接口的屏幕的显示示例的示图;

[0022] 图9A和9B是图示频率测量时段的示图;

[0023] 图10是图示频率演变的示例的示图;

[0024] 图11是图示与图10对应的频率的移动平均值和移动方差的示图;

[0025] 图12是图示与图10对应的评估值的示图;

[0026] 图13是图示图10至12的统一状态的示图;以及

[0027] 图14是图示计算机的配置示例的框图。

具体实施方式

[0028] 下面将参照附图详细描述本公开的优选实施例。

[0029] 根据本公开的一个实施例,提供了一种信息处理装置,该信息处理装置包括:获取单元,获取包括测量时段 i 中的采样值 x_i 的离散系统的时间序列数据;移动偏差计算单元,基于与预定测量时段 t 之前的预定时段对应的 N 个采样值 $x_t, x_{t-1}, x_{t-2}, \dots$ 和 x_{t-N+1} 的移动平均值 m_t 计算移动偏差 v_t ;以及评估值计算单元,基于与测量时段 t 对应的移动偏差 v_t 和与测量时段 $t-1$ 对应的移动偏差 v_{t-1} ,计算表示测量时段 t 中的离散系统的时间序列数据的迅速改变的评估值 s_t 。

[0030] 根据本公开的另一实施例,提供了一种由信息处理装置执行的信息处理方法,该信息处理方法包括:获取包括测量时段 i 中的采样值 x_i 的离散系统的时间序列数据,基于与预定测量时段 t 之前的预定时段对应的 N 个采样值 $x_t, x_{t-1}, x_{t-2}, \dots$ 和 x_{t-N+1} 的移动平均值 m_t 计算移动偏差 v_t ,并且基于与测量时段 t 对应的移动偏差 v_t 和与测量时段 $t-1$ 对应的移动偏

差 v_{t-1} ,计算表示测量时段 t 中的离散系统的时间序列数据的迅速改变的评估值 s_t 。

[0031] <1. 实施例>

[0032] 首先,将描述与应用了根据本公开的信息处理装置的实施例对应的检索装置的概况。检索装置将互联网或者内联网上出现的各种文档设定为检索对象,检索包括检索关键词的文档,并且提取检索文档中共同包括的字符串(以下称为同现关键词或者主题)。检索装置提供在包括检索关键词和同现关键词的互联网上的文档中在预定时间点已变为世界流行主题(时髦主题)的信息,作为与检索信息相关联的信息。

[0033] 例如,检索装置将互联网上出现的推特的发布信息(tweets)(显示推特用户张贴(输入)的140个或更少的字符的短句)设定为检索对象,检索包括检索关键词的发布信息,并且提取检索的发布信息中共同包括的同现关键词。检索装置针对每个提取的同现关键词计算表示流行性的评估值,显示评估值列表使得同现关键词被用户选择,并且向用户提供包括所选择的同现关键词和检索关键词的发布信息。由此,可以向用户提供关于已变为世界流行主题的信息的发布信息。

[0034] 例如,如果检索关键词被设定为“浅草寺(Sensoji Temple)”,则提取“台东区(Taito Ward)”、“护国寺(Gokokuji)”、“地震”、“在浅草”、以及“交叉点”作为同现关键词。如果用户从提取的同现关键词中选择“地震”,则向用户提供包括所选同现关键词“地震”和检索关键词“浅草寺”的发布信息。

[0035] 检索关键词可以由用户输入,或者可以基于用户的操作历史自动设定。例如,用户创建的文档中频繁出现的字符串、用户创建的播放列表中包括的艺术家名称或者歌曲标题、以及用户观看的电视节目中频繁出现的明星名称可以被提取,并且可以被设定为检索关键词。

[0036] 作为与检索关键词的对比,可以设定一个或者多个对比关键词。与检索关键词相似,对比关键词可以由用户输入,或者可以基于用户的操作历史自动设定。当自动设定对比关键词时,可以基于设定的检索关键词确定对比关键词。例如,当检索关键词是艺术家名称时,可以从互联网上的信息中检索来自同一国家的另一个艺术家,并且该另一个艺术家的艺术家名称可以被确定为对比关键词。

[0037] 例如,当AAA被设定为检索关键词而BBB被设定为对比关键词时,从包括检索关键词AAA的多个发布信息中提取同现关键词。然而,在包括对比关键词BBB的多个发布信息中出现频率高的关键词被排除在外。

[0038] 可以将多个字符串设定为检索关键词和对比关键词,并且可以执行“与(AND)”检索。

[0039] 在本公开中,下面将描述其中推特的每个发布信息被设定为检索对象的情况的示例。然而,作为实施例的检索装置的检索对象不限于发布信息。

[0040] 检索对象文档和检索关键词不限于由诸如日语和英语的自然语言表示的检索对象文档和检索关键词,只要检索对象文档和检索关键词可以由字符串或者符号串表示。例如,DNA信息、音素、乐谱信息、通过以一维排列方式表示被量化的并且包括在符号串中的真实数值获得的数据、以及通过以一维排列方式表示通过以多维排列方式表示被量化的并且包括在符号串中的真实数值获得的数据而获得的数据,可以被设定为检索对象文档和检索关键词。

[0041] [检索装置的配置示例]

[0042] 图1图示了作为实施例的检索装置中包括的功能块的配置示例。检索装置10包括：关键词设定单元11、文档检索单元12、噪声去除单元13、检索索引创建单元14、流行性确定单元15、主题提取单元16、主题输出单元17、主题文档输出单元18以及数据库20。图2图示了数据库(DB)20的详细配置。数据库20包括检索文档存储数据库(DB)21、文档检索索引数据库(DB)22、以及主题存储数据库(DB)23。

[0043] 关键词设定单元11将用户输入的字符串设定为检索关键词。关键词设定单元11将用户输入的字符串设定为对比关键词。关键词设定单元11可以自动设定检索关键词和对比关键词至少之一。

[0044] 文档检索单元12将互联网上出现的推特的每个发布信息设定为检索对象，并且检索包括检索关键词的发布信息。文档检索单元12将互联网上出现的推特的每个发布信息设定为检索对象，并且检索包括对比关键词的发布信息。可以将被设定为检索对象的每个发布信息的提及日期和时间的时段限定为从当前时间到一个月之前。作为文档检索单元12的检索结果获得的发布信息与检索关键词或者对比关键词相关联，并且关联结果被存储在数据库20的检索文档存储数据库21中。

[0045] 噪声去除单元13从作为检索结果获得的发布信息中去除不会变为同现关键词的字符串(以下称为噪声)。这将参照图4具体描述。

[0046] 检索索引创建单元14基于关于作为检索结果获得的存储在检索文档存储数据库20中的发布信息的后缀数组，创建检索索引。所创建的检索索引存储在数据库20的文档检索索引数据库22中。通过创建检索索引，可以高速计算当在每个发布信息中提取同现关键词时所需的主题(同现关键词)候选字符串的出现频率DF(文档频率)。

[0047] 当自动设定检索关键词或者对比关键词时，流行性确定单元15确定检索关键词或者对比关键词的候选者的流行性。流行性确定单元15确定所提取的同现关键词(主题)的流行性。

[0048] 主题提取单元16从噪声被去除的检索结果的每个发布信息中提取同现关键词(主题)。所提取的同现关键词(主题)存储在数据库20的主题存储数据库23中。

[0049] 主题输出单元17输出提取的同现关键词(主题)。主题输出单元17可以具有用于基于所提取的同现关键词(主题)自动创建发布信息并且在推特上张贴该发布信息的机器人(bot)创建功能。

[0050] 主题文档输出单元18从检索文档存储数据库21获取包括所提取的同现关键词(主题)的发布信息，并且输出这些发布信息。

[0051] [描述操作]

[0052] 接着将描述检索装置10的操作。图3是图示检索装置10执行的相关联的信息检索处理的流程图。

[0053] 在步骤S1中，关键词设定单元11将用户输入的字符串设定为检索关键词。用户创建的文档中频繁出现的字符串、用户创建的播放列表中包括的艺术家名称或者歌曲标题、以及用户观看的电视节目中频繁出现的明星名称都可以被提取并且设定为检索关键词。在这种情况下，可以针对所提取的艺术家名称计算下面描述的流行性的评估值，并且可以采用评估值是预定阈值或更大的艺术家名称作为检索关键词。

[0054] 在步骤S1中,关键词设定单元11将用户输入的字符串或者自动确定的字符串设定为对比关键词。可以省略对比关键词的设定。

[0055] 在步骤S2中,文档检索单元12将互联网上出现的推特的每个发布信息设定为检索对象并且检索包括检索关键词的发布信息。检索结果的发布信息与检索关键词相关联并且关联结果存储在检索文档存储数据库21中。当对比关键词被设定时,文档检索单元12将互联网上出现的推特的每个发布信息设定为检索对象并且检索包括该对比关键词的发布信息。检索结果的发布信息与对比关键词相关联并且关联结果存储在检索文档存储数据库21中。

[0056] 在步骤S3中,噪声去除单元13从作为检索结果获得的发布信息中去除不会变为同现关键词的噪声。

[0057] 图4图示了作为检索结果的示例的发布信息。在图4中,下划线字符串被噪声去除单元13作为噪声去除。就是说,当检索对象是发布信息时,意指锐推(retweet)的“RT”,表示答复方的目的地的“@用户名”(在图4中用户名是tokyotralver),表示统一资源定位符(URL)的“http://...”,以及表示无用信息的“#...”被去除。

[0058] 回到图3,在步骤S4中,检索索引创建单元14基于后缀数组针对作为检索结果获得的被存储在检索文档存储数据库20中的发布信息创建检索索引。所创建的检索索引被存储在文档检索索引数据库22中。

[0059] 在步骤S5中,主题提取单元16执行主题提取处理,用于从噪声被去除的检索结果的每个发布信息中提取同现关键词(主题)。所提取的同现关键词(主题)被存储在数据库20的主题存储数据库23中。

[0060] 图5是具体图示主题提取处理的流程图。

[0061] 在步骤S11中,主题提取单元16在噪声被去除的检索结果的发布信息组中出现的所有部分字符串中,提取仅作为其他部分字符串的一部分出现的部分字符串之外的字符串组。这相当于在出现频率DF不变的范围中提取最长的部分字符串组。可以使用基于后缀数组的检索索引来高速执行该处理。

[0062] 根据下述字符类型的符合规则的字符串被排除在主题候选字符串之外,并且提取其余字符串作为主题候选字符串。

[0063] [采用的字符类型]

[0064] 空格(空白)、半角英文字符、罗马字符扩展、平假名、片假名、全角符号、长音符号、半角符号、控制字符、无效字符、日本汉字、半角数字、标点符号、朝鲜语字符、泰语字符、阿拉伯字符、希伯来语字符、西里尔字符以及希腊字符被采用作为字符类型。

[0065] [从主题候选字符串中排除标志(token)的规则]

[0066] 当标志之前的字符(前一标志的最后字符)是长音符号时,该标志不被指定为主题候选字符串。

[0067] 当标志的第一字符是空格时,该标志不被指定为主题候选字符串。

[0068] 当标志的第一字符是全角符号时,该标志不被指定为主题候选字符串。

[0069] 当标志的第一字符是长音符号时,该标志不被指定为主题候选字符串。

[0070] 当标志的第一字符是半角符号时,该标志不被指定为主题候选字符串。

[0071] 当标志的第一字符是控制字符和无效字符时,该标志不被指定为主题候选字符

串。

[0072] 当标志的第一字符是标点符号时,该标志不被指定为主题候选字符串。

[0073] 当标志之后的字符(下一标志的第一字符)是长音符号时,该标志不被指定为主题候选字符串。

[0074] 当标志的最后字符是空格时,该标志不被指定为主题候选字符串。

[0075] 当标志的最后字符是全角符号时,该标志不被指定为主题候选字符串。

[0076] 当标志的最后字符是半角符号时,该标志不被指定为主题候选字符串。

[0077] 当标志的最后字符是控制字符和无效字符时,该标志不被指定为主题候选字符串。

[0078] 当标志的最后字符是标点符号时,该标志不被指定为主题候选字符串。

[0079] 当标志之前的字符(前一标志的最后字符)和标志的第一字符或者标志之后的字符(下一标志的第一字符)和标志的最后字符均为半角英文字符和罗马字符扩展时,该标志不被指定为主题候选字符串。

[0080] 当标志之前的字符(前一标志的最后字符)和标志的第一字符或者标志之后的字符(下一标志的第一字符)和标志的最后字符均是片假名时,该标志不被指定为主题候选字符串。

[0081] 当标志之前的字符(前一标志的最后字符)和标志的第一字符或者标志之后的字符(下一标志的第一字符)和标志的最后字符均是半角数字时,该标志不被指定为主题候选字符串。

[0082] 当标志之前的字符(前一标志的最后字符)和标志的第一字符或者标志之后的字符(下一标志的第一字符)和标志的最后字符均是朝鲜语字符时,该标志不被指定为主题候选字符串。

[0083] 当标志之前的字符(前一标志的最后字符)和标志的第一字符或者标志之后的字符(下一标志的第一字符)和标志的最后字符均是西里尔字符时,该标志不被指定为主题候选字符串。

[0084] 例如,如图6所示,当噪声被去除的发布信息是“People who stock upon chocolate raise your hands(囤积巧克力的人,举手)”时,首先在检索结果的发布信息组中的所有部分字符串中提取仅作为其他部分字符串的一部分出现的部分字符串之外的字符串组。例如,当“cho”、“chocolate,”和“chocolate”的出现频率DF分别是10、10和4时,提取“chocolate”。然而,不提取“cho”。然后,通过应用从主题候选字符串中排除标志的规则,提取主题候选字符串。

[0085] 这样,主题提取单元16可以基于出现频率DF的变化点以及字符类型的差异,而非根据检索对象文档的语言,来提取主题候选字符串。然而,主题提取单元16可以基于文档语言的特性使用语形学分析来提取主题候选字符串。

[0086] 当提取相似的字符串作为主题候选字符串时,可以收集相似的字符串作为一个字符串。在该情况下,相似意味着字符串的相似度高,并且出现文档的相似度高。

[0087] 在步骤S12中,主题提取单元16使用存储在文档检索索引数据库22中的检索索引,来计算噪声被去除的检索结果的发布信息中的每个主题字符串的出现频率DF。

[0088] 在步骤S13中,主题提取单元16采用其中出现频率DF满足预定条件的主题候选字

字符串作为主题(同现关键词)。就是说,当检索关键词和对比关键词两者被设定时,主题提取单元16采用如下主题候选字符串作为主题:其中通过将使用检索关键词的检索结果的发布信息的出现频率DF除以使用对比关键词的检索结果的发布信息的出现频率DF而获得的值是预定阈值或更大。当仅检索关键词被设定时,主题提取单元16采用如下主题候选字符串作为主题:其中使用检索关键词的检索结果的发布信息的出现频率DF是预定阈值或更大。

[0089] 当确定是否采用主题候选字符串作为主题时,不同于使用上述出现频率DF,可以计算并且使用信息增益、交互信息、二元正态分离、倍数变化、以及相关系数。可以执行诸如卡方测试的测试以测量主题的特异性。

[0090] 在如上所述提取主题之后,主题提取处理结束并且处理返回到图3的步骤S6。

[0091] 在步骤S6中,流行性确定单元15针对在步骤S5中提取的每个同现关键词(主题)计算流行性的评估值。下文将参照附图9A至13描述计算方法。

[0092] 在步骤S7中,主题输出单元17向用户提供所提取的同现关键词(主题)及其流行性的评估值。在步骤S8中,当检索装置自动设定主题时,主题输出单元17可以不向用户提供所提取的同现关键词(主题)及其流行性的评估值。

[0093] 如果用户选择所提供的同现关键词(主题)或者由检索装置自动选择其中流行性的评估值是阈值或更大的同现关键词,则在步骤S8中,主题文档输出单元18从检索文档存储数据库21获取包括所提取的同现关键词(主题)和检索关键词的发布信息,并且将这些发布信息作为与检索关键词相关联的信息提供给用户。当获取的多个发布信息彼此相似时,可以收集多个发布信息作为一个发布信息并且可以将该发布信息提供给用户。以这种方式,作为相关联的信息检索处理而执行的一系列操作结束。

[0094] [用作用户接口的屏幕显示示例]

[0095] 图7图示了用作检索装置10的用户接口的屏幕的显示示例。屏幕50设置有检索关键词输入栏51、获取发布信息按钮52、从发布信息获取主题词按钮53、显示发布信息按钮54、主题显示栏55、评估值显示栏56、和发布信息显示栏57。

[0096] 用户可以将检索关键词输入到检索关键词输入栏51。如果用户操作获取发布信息按钮52,则从互联网上出现的推特的发布信息中检索包括检索关键词的发布信息。

[0097] 如果用户操作从发布信息获取主题词按钮53,则从检索结果的发布信息中提取同现关键词(主题)并且使同现关键词和流行性的评估值显示在主题显示栏55中。如果用户选择显示在主题显示栏55上的同现关键词(主题),则关于所选择的同现关键词(主题)的流行性的评估值的时间演变显示在评估值显示栏56上。

[0098] 如果在其中选择了同现关键词(主题)的状态下用户操作显示发布信息按钮54,则包括检索关键词和所选择的同现关键词(主题)的发布信息显示在发布信息显示栏57中。

[0099] 例如,如图7所示,如果用户将“浅草寺(Sensoji Temple)”作为检索关键词输入到检索关键词输入栏51并且操作获取发布信息按钮52,则包括检索关键词“浅草寺(Sensoji Temple)”的发布信息被检索。在该情况下,如果用户操作从发布信息获得主题词按钮53,则同现关键词(主题)“台东区(Taito Ward)”、“护国寺(Gokokuji)”、“地震”、“地震灾难发生时间:下午2:46”、“在浅草”、和“交叉点”以及流行性的评估值显示在主题显示栏55中。

[0100] 如果用户从显示在主题显示栏55上的同现关键词(主题)中选择“台东区(Taito Ward)”,则所选择的同现关键词(主题)的流行性的评估值的时间演变显示在评估值显示栏

56中。

[0101] 如果在其中选择“台东区(Taito Ward)”作为同现关键词(主题)的状态下用户操作显示发布信息按钮54,则包括检索关键词“浅草寺(Sensoji Temple)”和所选择的同现关键词(主题)“台东区(Taito Ward)”的发布信息显示在发布信息显示栏57中。然而,在图7中,在发布信息显示栏57中发布信息的句子由*(星号)代替。

[0102] 例如,如图8所示,如果用户在检索关键词输入栏51上输入“蔬菜(vegetables)”作为检索关键词,并且操作获取发布信息按钮52,则包括检索关键词“蔬菜(vegetables)”的发布信息被检索。在这种情况下,如果用户操作从发布信息获取主题词按钮53,则同现关键词(主题)“儿童(child)”、“儿童的(of child)”、“制成饮料(made to drink)”、“饮用(drunk)”、“喂养(fed)”、“有运输限制的(of shipment limitation)”、以及“消费者的(of consumer)”以及流行性的评估值显示在主题显示栏55上。

[0103] 如果用户从主题显示栏55上显示的同现关键词(主题)中选择“有运输限制的(of shipment limitation)”,则关于所选择的同现关键词(主题)的流行性的评估值的时间演变显示在评估值显示栏56上。

[0104] 如果在其中选择“有运输限制的(of shipment limitation)”作为同现关键词(主题)的状态下用户操作显示发布信息按钮54,则包括检索关键词“蔬菜(vegetables)”和所选择的同现关键词(主题)“有运输限制的(of shipment limitation)”的发布信息显示在发布信息显示栏57上。然而,在图8中,在发布信息显示栏57中,发布信息的句子由*(星号)代替。

[0105] 如上所述,检索装置10可以针对每个主题收集包括用户感兴趣的主题的发布信息并且可以将这些发布信息提供给用户。如果检索关键词被自动设定,则检索装置10可以针对每个主题收集包括用户感兴趣的估计主题的发布信息并且可以将这些发布信息提供给用户。

[0106] [计算流行性的评估值的方法]

[0107] 接着,将描述在相关联的信息检索处理的步骤S6中计算同现关键词的流行性的评估值的方法。

[0108] 首先,基于其中出现同现关键词的发布信息的张贴日期和时间,将检索结果的发布信息中的同现关键词的出现频率DF转换为离散系统的时间序列数据。具体地,将同现关键词的出现频率DF转换为预定测量时段(例如,24小时)中的频率。

[0109] 图9A和9B图示了设定频率测量时段的方法。就是说,如图9A所示,频率测量时段未被设定为在时间轴T上重叠,并且如图9B所示,频率测量时段可以被设定为在时间轴T上重叠。

[0110] 当频率测量时段被设定为不在时间轴T上重叠时,每个测量间隔中的频率和变为出现频率DF。当频率测量时段被设定为在时间轴T上重叠时,可以在短时段中获取多个频率的样本。

[0111] 当特定测量时段t中的频率被设定为 x_t 时,使用测量时段t之前的N个测量时段t、t-1、t-2、...、和t-N+1中的频率 $x_t, x_{t-1}, x_{t-2}, \dots$ 、和 x_{t-N+1} 计算测量时段t中的流行性的评估值 S_t 。

[0112] 具体地,依次计算移动平均值 m_t 、移动偏差 v_t 和评估值 S_t 。

[0113] 移动平均值 $m_t = (\sum x_i) / N \cdots (1)$

[0114] 移动偏差 $v_t = \sqrt{((\sum (m_t - x_i)) / N) \cdots (2)}$

[0115] 评估值 $s_t = v_t / v_{t-1} \cdots (3)$

[0116] Σ 意指与 $i=t$ 至 $i=t-N+1$ 对应的 N 个值的和。

[0117] 例如,当用作离散系统的时间序列数据的频率 x_t 如图10所示演变时,移动平均值 m_t 如图11中的粗线所示演变,而移动偏差 v_t 基于图11中的粗线如细线所示以带的形式演变。同时,评估值 s_t 如图12所示演变。图13图示了图10和12的重叠状态。

[0118] 如从图13所见,当频率 x_t 迅速改变时,评估值 s_t 增加。因此,如果针对同现关键词计算评估值 s_t ,则当确定关键词是否已变为(正趋向于)世界流行主题时,评估值可以被用作索引。

[0119] 当测量时段 t 短时,评估值 s_t 表示短期流行性趋势,并且当测量时段 t 长时,评估值 s_t 表示长期流行性趋势。因此,可以计算测量时段 t 短时(例如,1天=24小时)的评估值 $s_{t(1天)}$ 和当测量时段长时(例如,1个月=30天)的评估值 $s_{t(30天)}$,并且可以计算评估值的加权平均值作为最终评估值。所计算的最终评估值可以用作索引,用于表示关键词是否已变为(正趋向于)世界流行主题的短期流行性趋势和长期流行性趋势。

[0120] [评估值的其他使用目的]

[0121] 除了确定同现关键词的流行性之外,还可以以各种方式使用评估值 s_t 。

[0122] 例如,如果预定时段中的各种产品中的每个的销售量被设定为频率 x_t 并且计算评估值 s_t ,则评估值 s_t 可以用作用于确定拳头产品的索引。

[0123] 如果检索关键词的检索次数被设定为频率 x_t 并且计算评估值 s_t ,则评估值 s_t 可以用作用于确定已变为世界流行主题的关键词的索引。

[0124] 上述系列处理可以由硬件或软件实现。当系列处理由软件实现时,形成软件的程序从程序记录介质安装到嵌入专用硬件的计算机中以及可以安装各种程序并且可以执行各种功能的通用计算机中。

[0125] 图14是图示了通过程序执行系列处理的计算机的硬件配置示例的框图。

[0126] 在计算机100中,中央处理单元(CPU)101、只读存储器(ROM)以及随机存取存储器(RAM)103通过总线104互相连接。

[0127] 输入/输出接口105连接到总线104。输入单元106,包括键盘、鼠标和麦克风;输出单元107,包括显示器和扬声器;存储单元108,使用硬盘或者非易失性存储器配置;通信单元109,使用网络接口配置;以及驱动器110,驱动诸如磁盘、光盘、磁光盘或者半导体存储器的可移除介质111连接到输入/输出接口105。

[0128] 在如上所述配置的计算机100中,CPU 101通过输入/输出接口105和总线104将存储在存储单元108中的程序加载到RAM 103中,并且执行这些程序,并且执行系列处理。

[0129] 由计算机执行的程序可以根据本公开中描述的顺序按时间序列处理,并且可以并行地处理或者在执行调用时在所需的定时处理。

[0130] 一个计算机可以处理程序并且多个计算机可以对程序执行分布式处理。程序可以被传送到远程计算机并且可以被执行。

[0131] 本公开的实施例不限于以上示例,并且在在不偏离本公开的精神和范围的情况下,可以进行各种改变。

[0132] 本公开包含与2011年5月18日提交日本专利局的日本在先专利申请JP 2011-111644中公开的主题内容相关的主题内容,其整体内容通过引用合并于此。

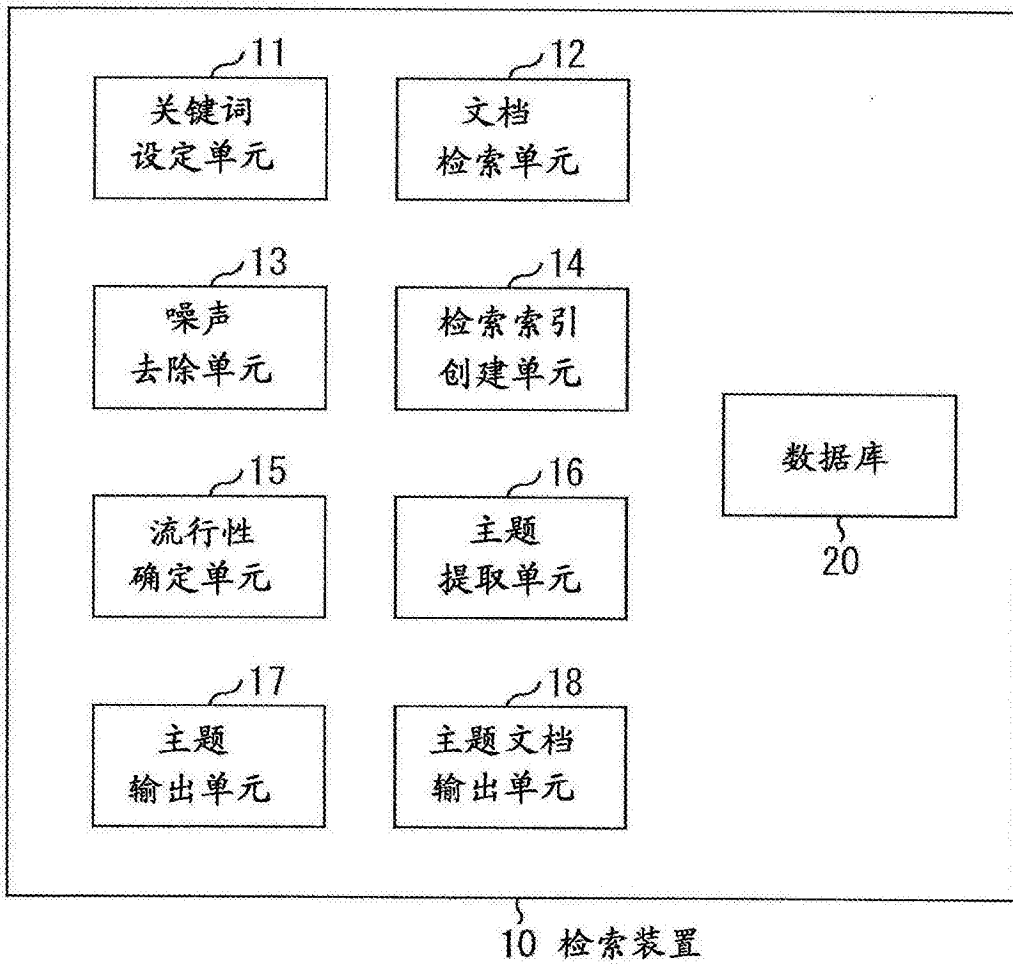


图1

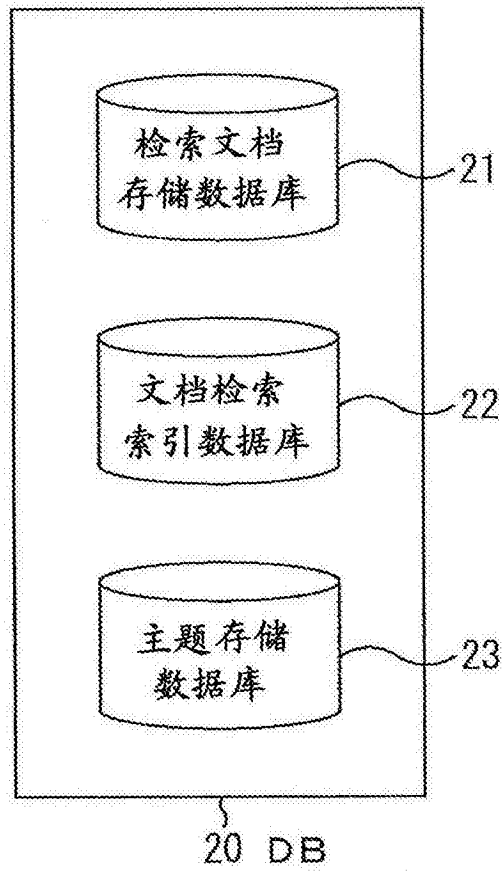


图2

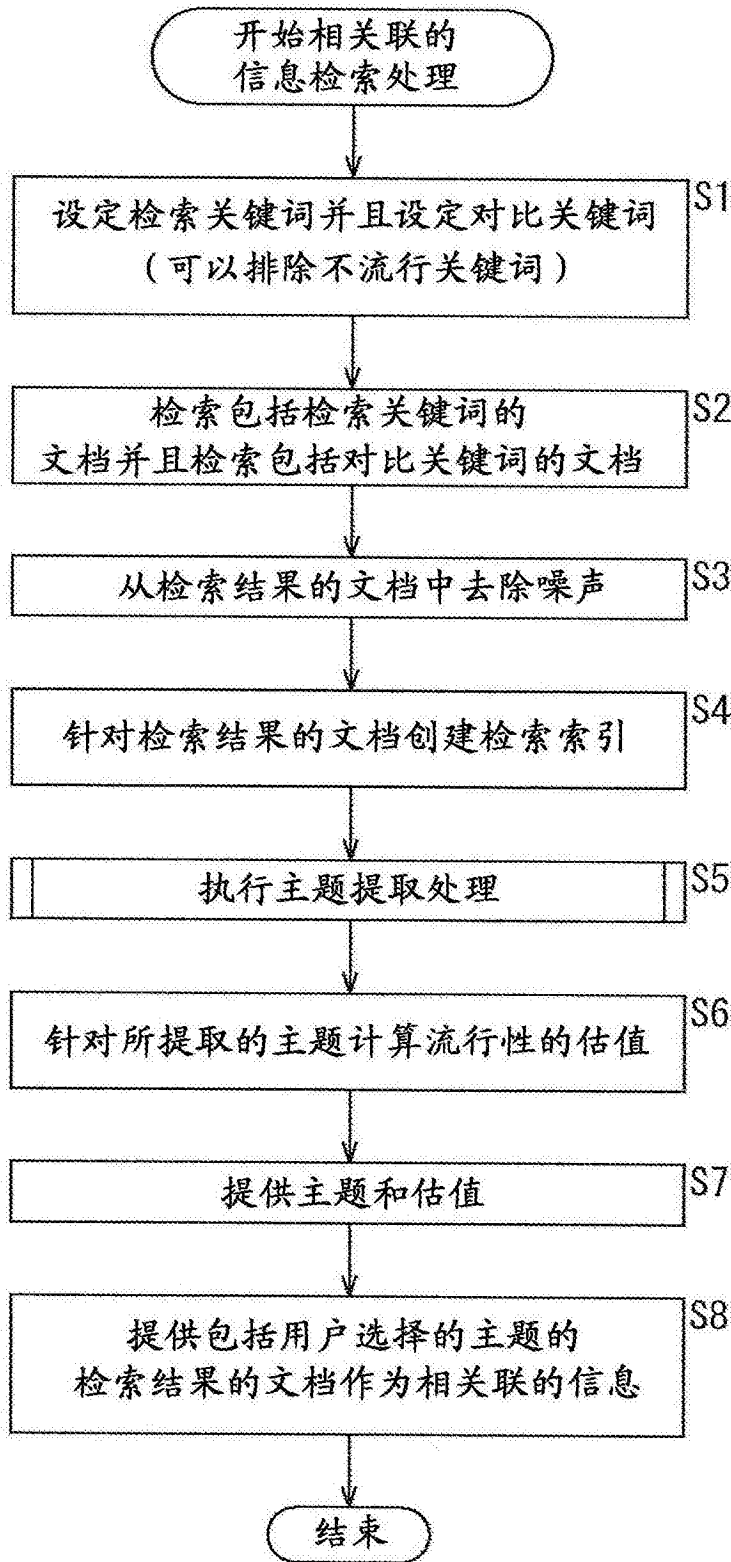


图3

RT @tokyotraveler: 我现在在在台区浅草的浅草寺。人好多!
http://****.****/****

图4

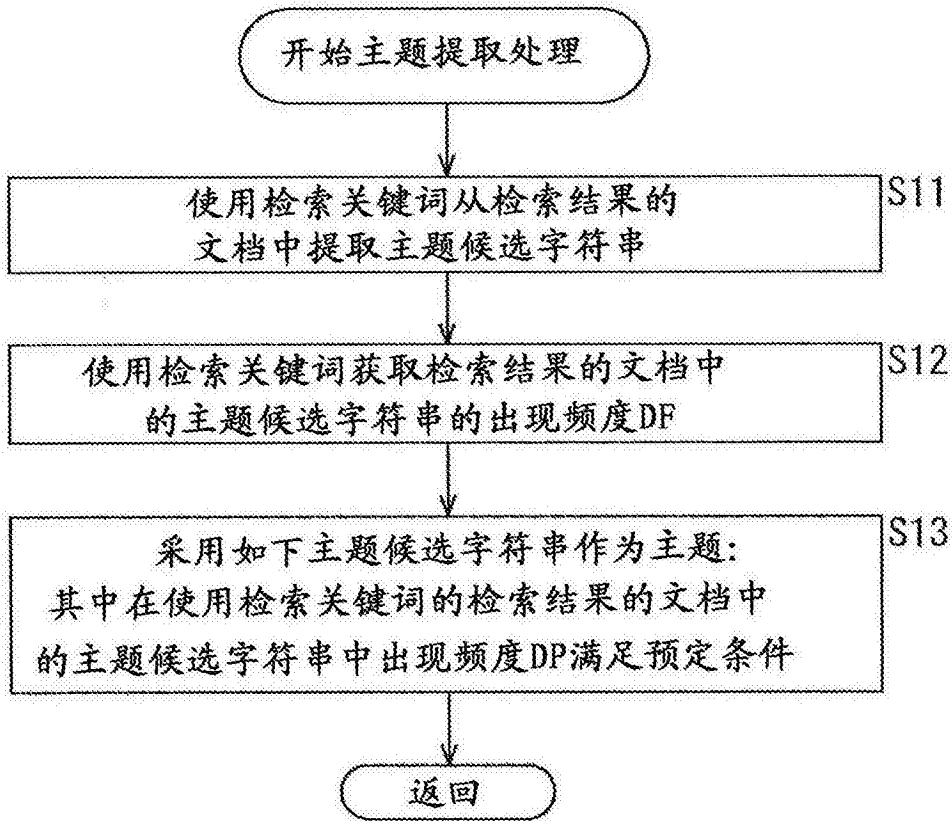


图5



图6

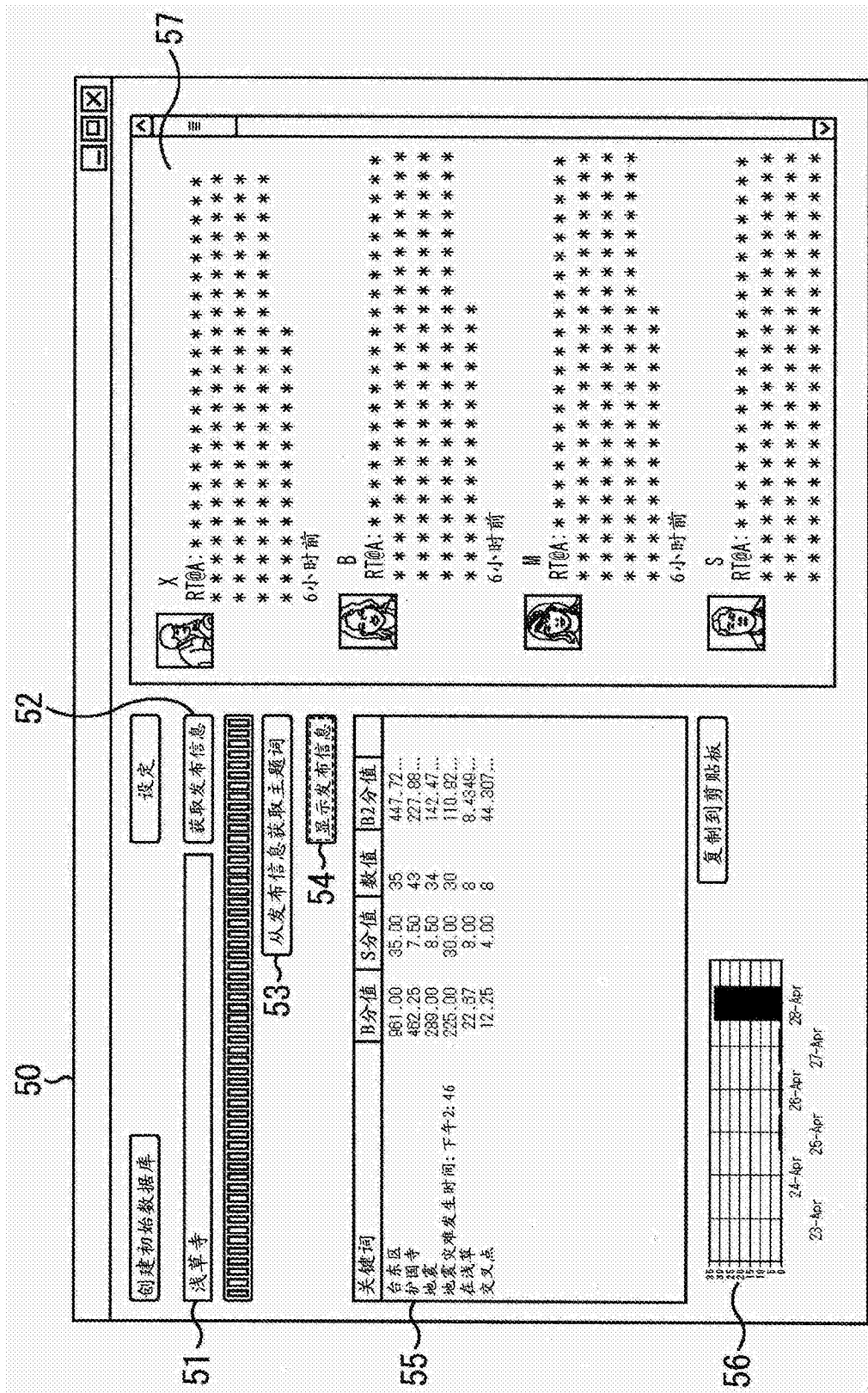


图7

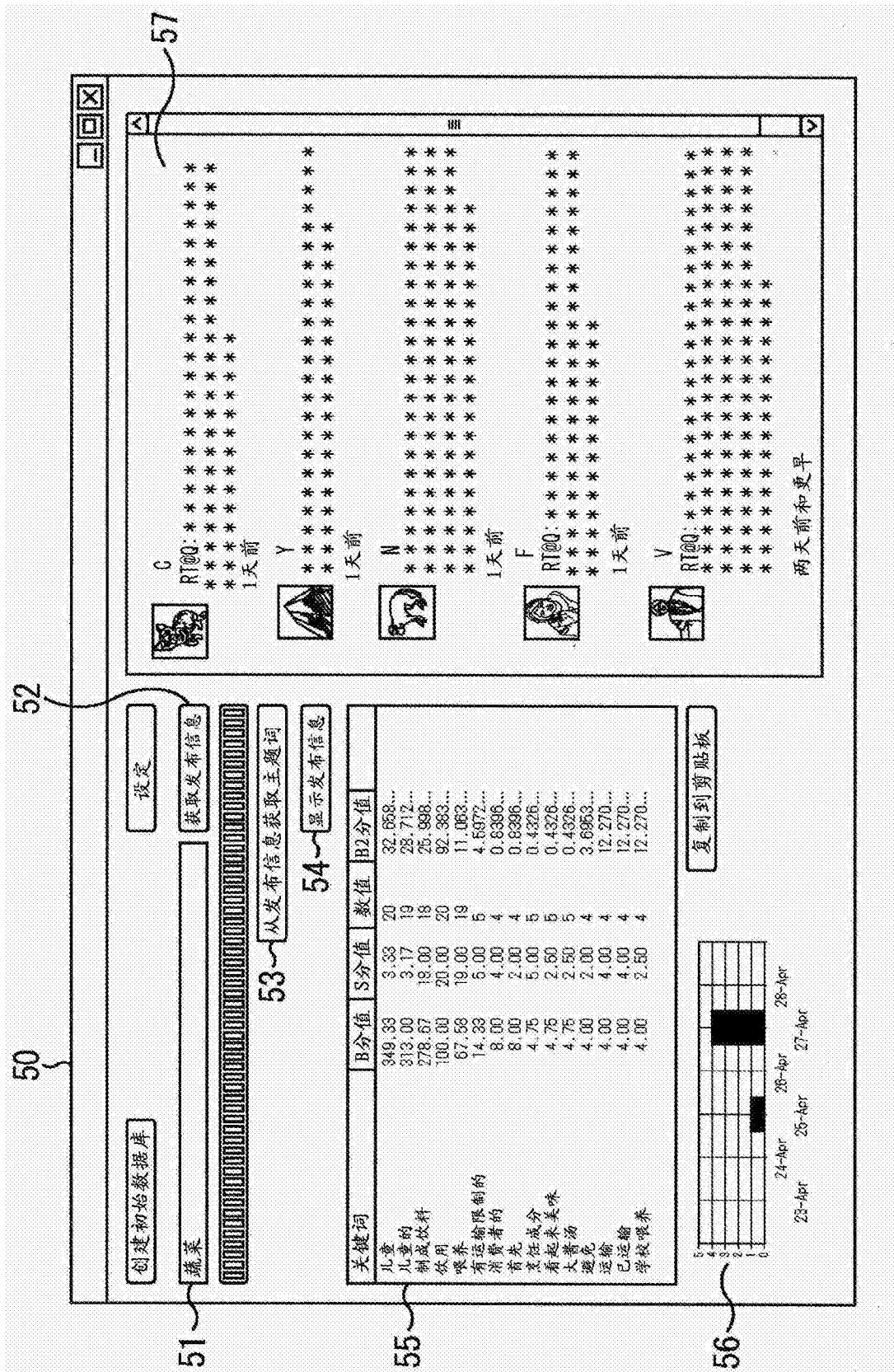


图8

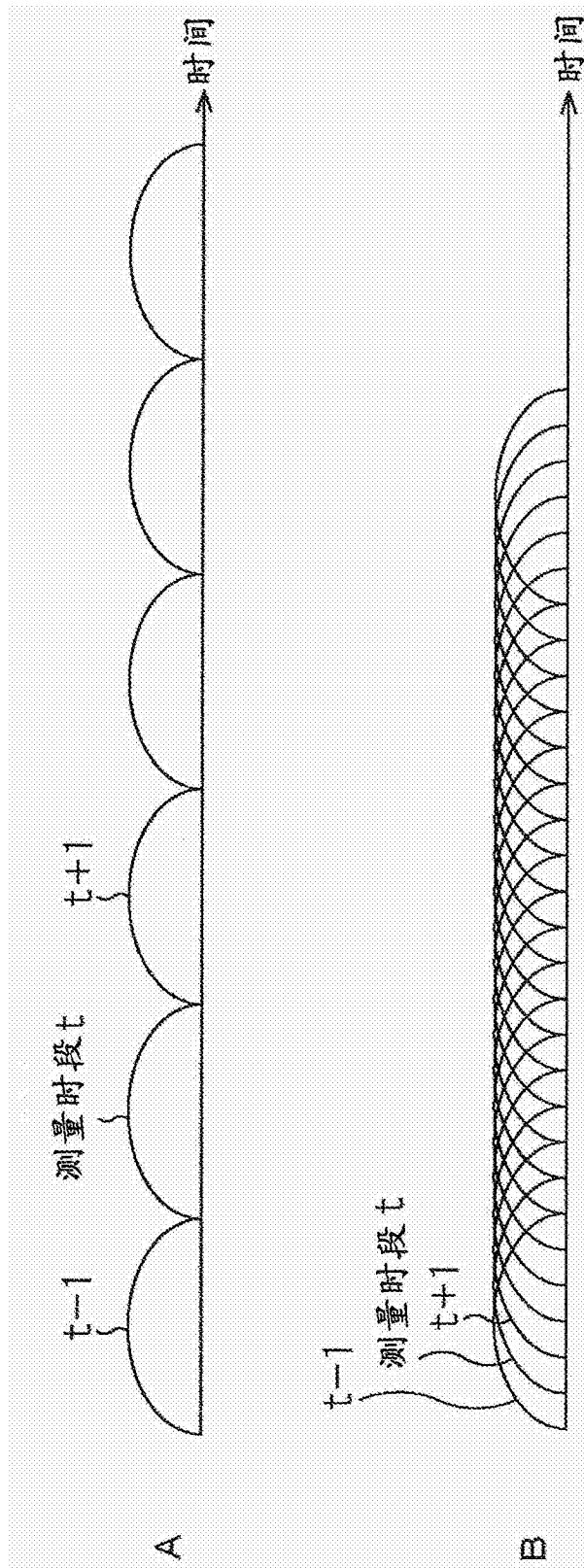


图9

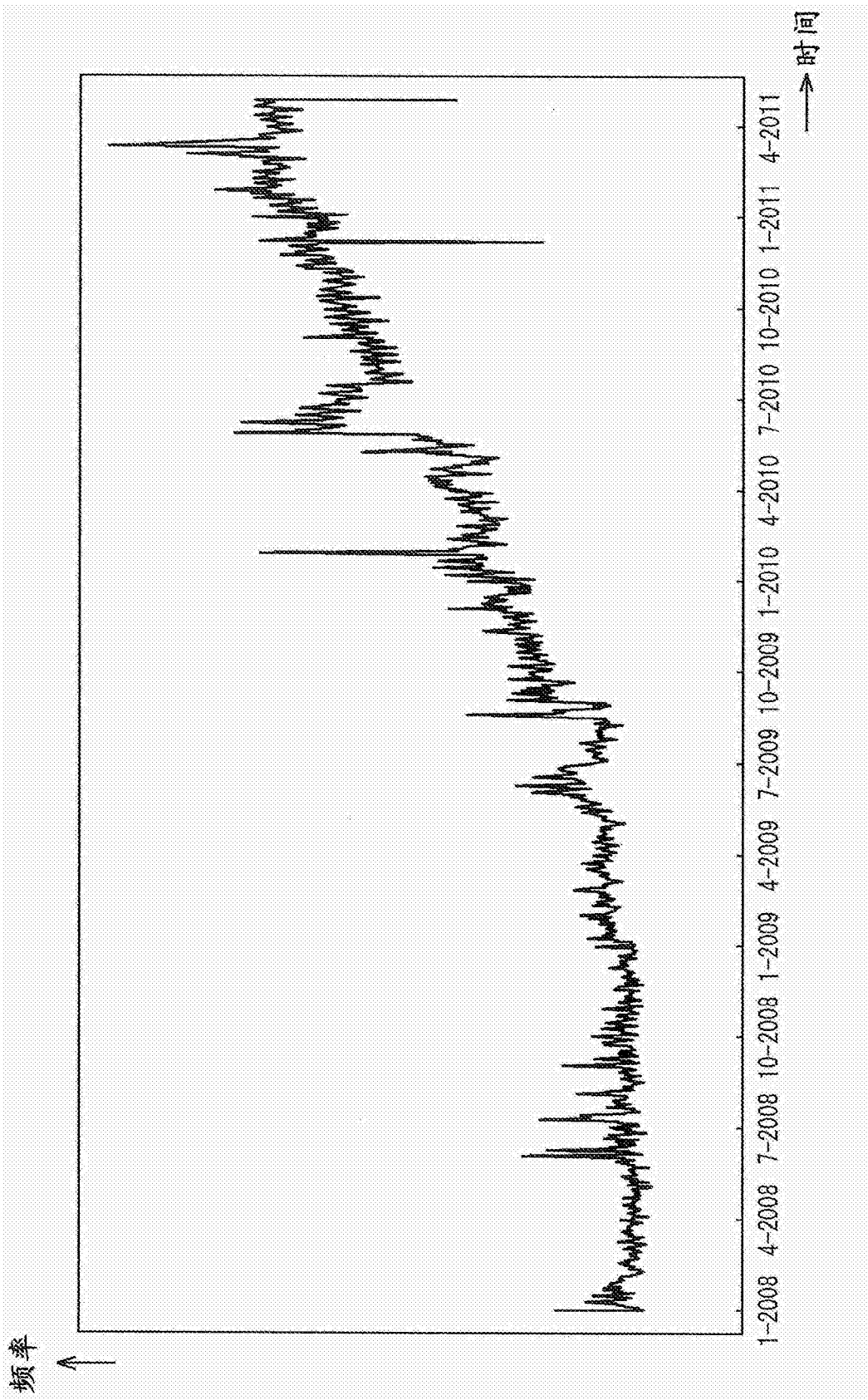


图10

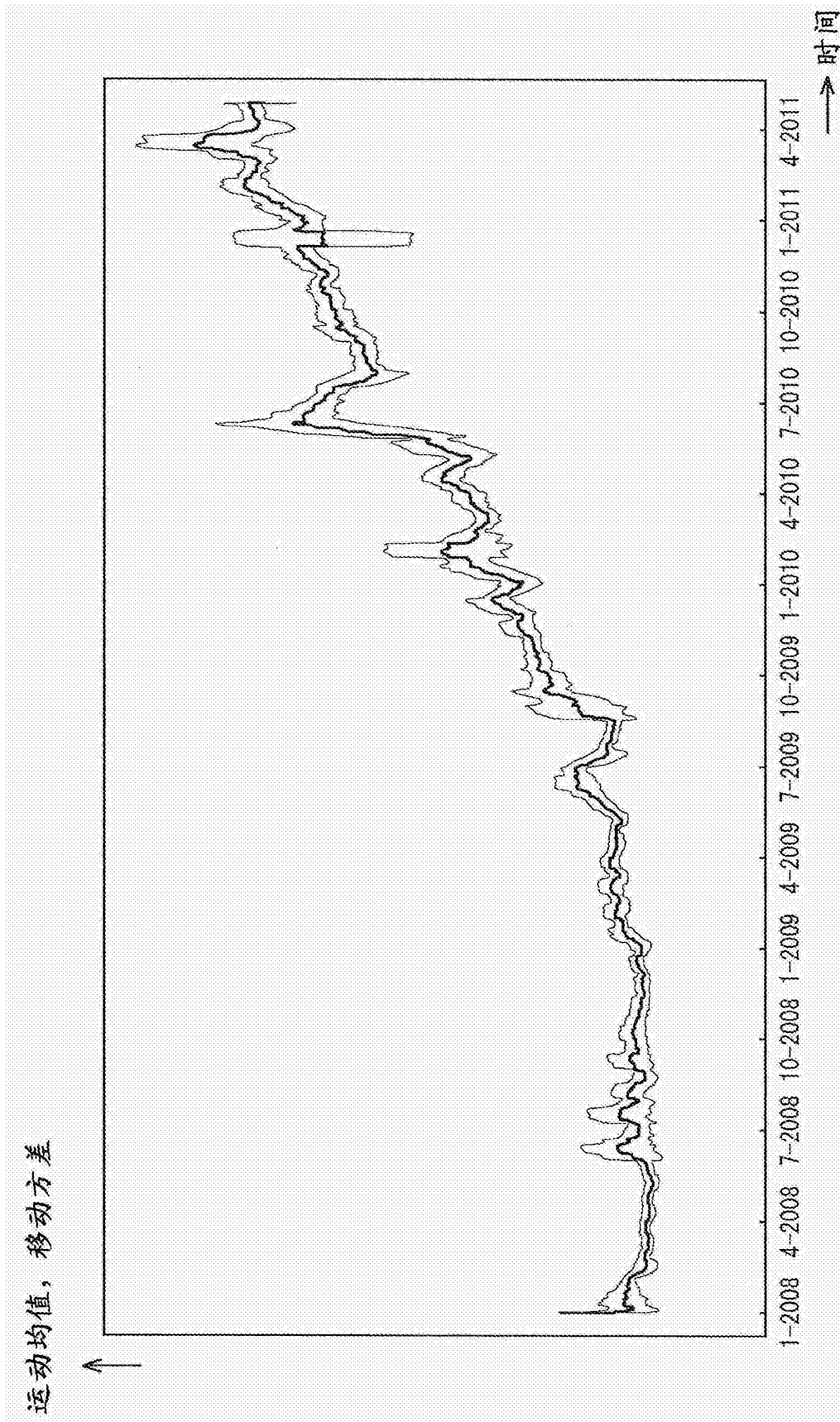


图11

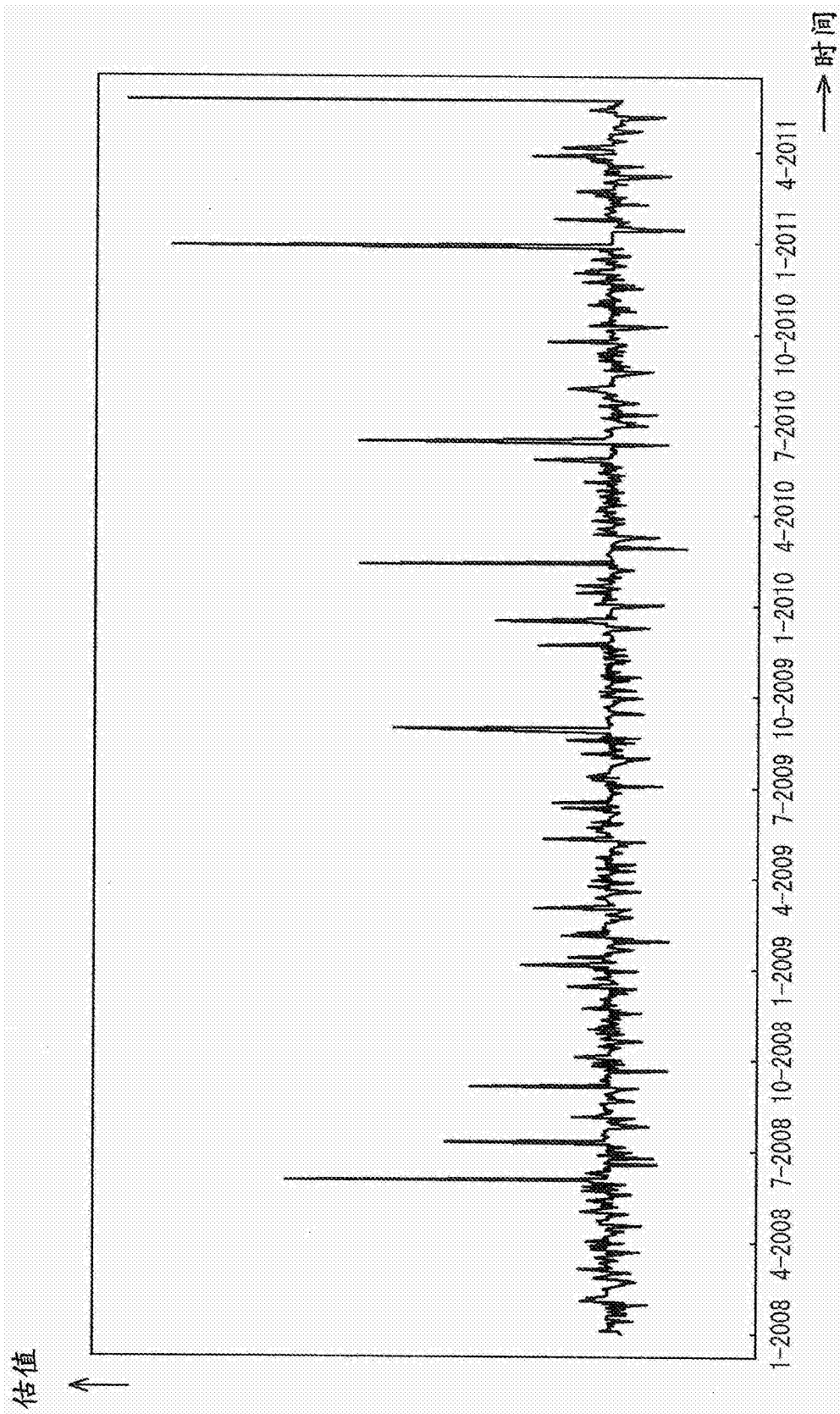


图12

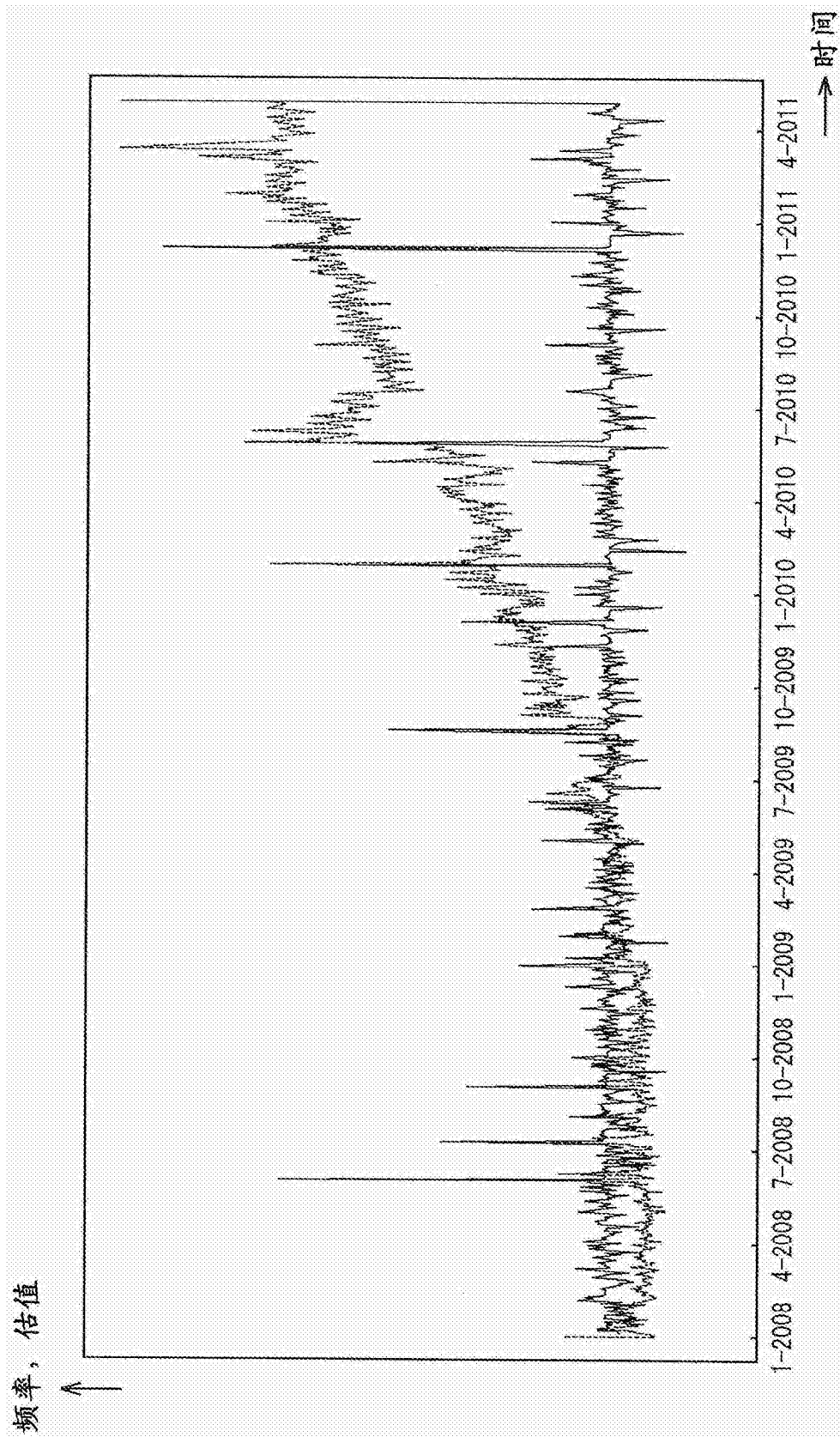


图13

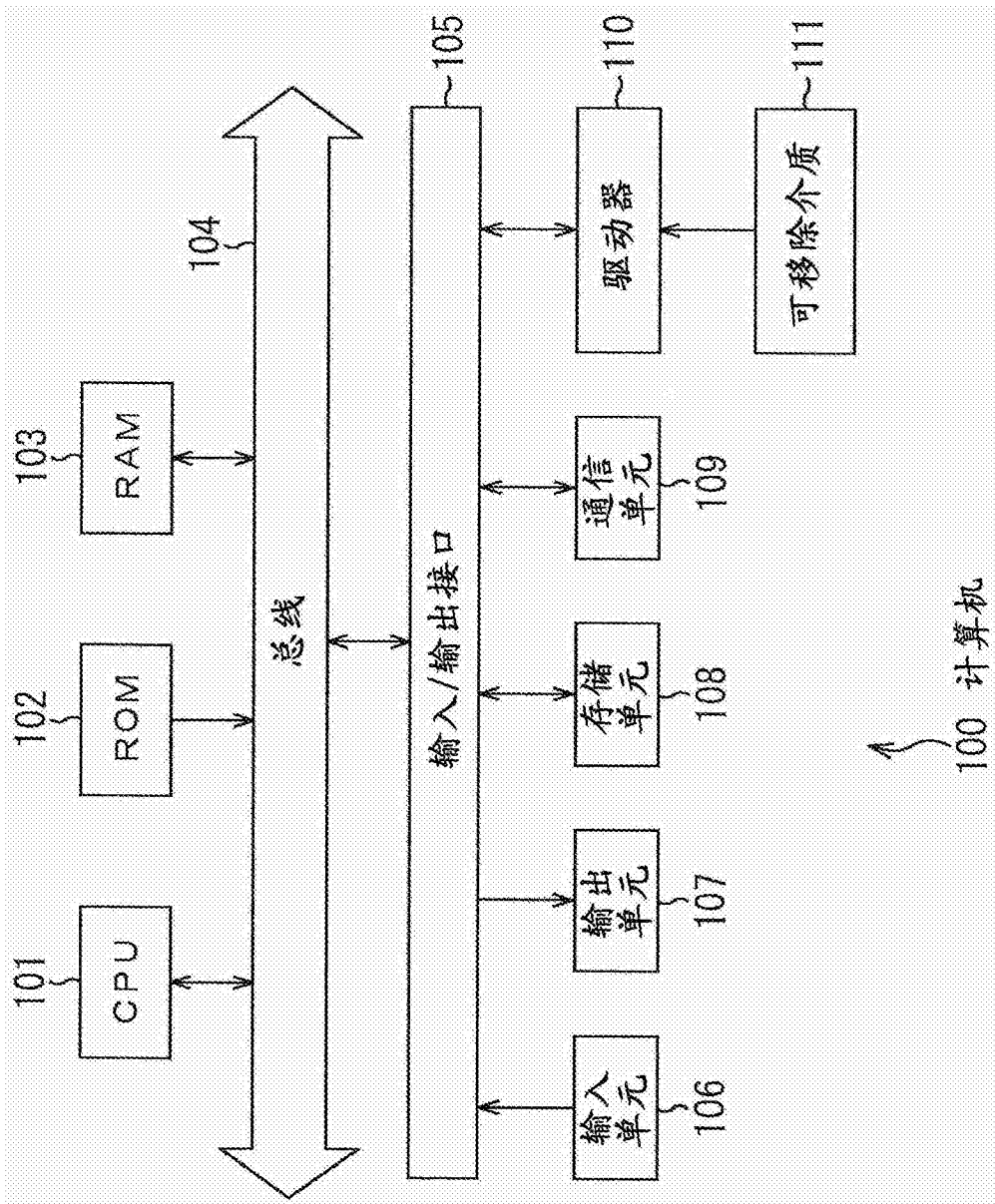


图14