

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2018 年 8 月 23 日 (23.08.2018)



(10) 国际公布号
WO 2018/149084 A1

(51) 国际专利分类号:

G06K 9/62 (2006.01)

(21) 国际申请号:

PCT/CN2017/090578

(22) 国际申请日:

2017 年 6 月 28 日 (28.06.2017)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

201710087590.X 2017 年 2 月 17 日 (17.02.2017) CN

(71) 申请人: 平安科技 (深圳) 有限公司 (PING AN

TECHNOLOGY(SHENZHEN)CO.,LTD.) [CN/CN];

中国广东省深圳市福田区八卦岭八卦三路

平安大厦, Guangdong 518000 (CN)。

(72) 发明人: 曾凡刚 (ZENG, Fangang); 中国广东省

深圳市福田区八卦岭八卦三路平安大

厦, Guangdong 518000 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限

公司 (ADVANCE CHINA IP LAW OFFICE); 中国

广东省广州市天河区花城大道 85 号 3901 房, Guangdong 510623 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家

保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区

保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,

(54) Title: METHOD AND DEVICE FOR IDENTIFYING CORRELATION VARIABLE, TERMINAL, AND STORAGE MEDIUM

(54) 发明名称: 相关变量识别方法、装置、终端及存储介质

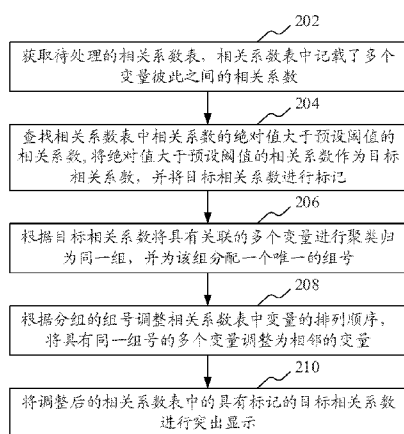


图 2

- 202 Acquire a correlation coefficient table, the correlation coefficient table storing correlation coefficients among multiple variables
- 204 Search, in the correlation coefficient table, for a correlation coefficient having an absolute value greater than a preset threshold, take the correlation coefficient having an absolute value greater than the preset threshold as a target correlation coefficient, and add a marker to the target correlation coefficient
- 206 Perform, according to the target correlation coefficient, clustering to group multiple correlated variables in a same cluster, and assign a unique cluster number to the cluster
- 208 Adjust, according to the assigned cluster number, a sequential order of the variables in the correlation coefficient table, and arrange the multiple variables having the same cluster number to be adjacent to one another
- 210 Highlight the target correlation coefficient with the marker in the adjusted correlation coefficient table

(57) Abstract: A method for identifying a correlation variable comprises: acquiring a correlation coefficient table, the correlation coefficient table storing correlation coefficients among multiple variables; searching, in the correlation coefficient table, for a target correlation coefficient having an absolute value greater than a preset threshold, and adding a marker to the target correlation coefficient; performing, according to the target correlation coefficient, clustering to group multiple correlated variables in a same cluster, and assigning a unique cluster number to the cluster; adjusting, according to the assigned cluster number, a sequential order of the variables in the correlation coefficient table, and arranging the multiple variables having the same cluster number to be adjacent to one another; and highlighting the target correlation coefficient with the marker in the adjusted correlation coefficient table.

(57) 摘要: 一种相关变量识别方法, 包括: 获取相关系数表, 所述相关系数表中记载了多个变量彼此之间的相关系数, 查找所述相关系数表中相关系数的绝对值大于预设阈值的目标相关系数, 并将所述目标相关系数进行标记, 根据所述目标相关系数将相关的多个变量进行聚类归为同一组, 并为该组分配一个唯一的组号, 根据分组的组号调整所述相关系数表中变量的排列顺序, 将具有同一组号的多个变量调整为相邻的变量, 将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

发明名称：相关变量识别方法、装置、终端及存储介质

本申请要求于 2017 年 2 月 17 日提交中国专利局、申请号为 201710087590X、发明名称为“相关变量识别方法和装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本发明涉及计算机处理领域，特别是涉及一种相关变量识别方法、装置、终端及存储介质。

背景技术

10 在数据建模过程中，当变量彼此高度相关时，即相关系数的绝对值很大时，变量会表现出很强的共线性，这时会造成模型失真。所以在建模过程中，一定会处理变量的相关性，传统的处理相关性只能将两个变量相关的数据显示，而三个以上的相关变量则需要人工识别，由于处理的数据往往比较多，通过人工识别无疑会耗时耗力，从而降低了数据建模的速度。

15

发明内容

根据本申请的各种实施例，提供一种相关变量识别方法、装置、终端及存储介质。

一种相关变量识别方法，包括：

20 获取待处理的相关系数表，所述相关系数表中记载了多个变量彼此之间的相关系数；

查找所述相关系数表中相关系数的绝对值大于预设阈值的相关系数，将所述绝对值大于预设阈值的相关系数作为目标相关系数，并将所述目标相关

系数进行标记;

根据所述目标相关系数将相关的多个变量进行聚类归为同一组, 并为该组分配一个唯一的组号;

根据分组的组号调整所述相关系数表中变量的排列顺序, 将具有同一组号的多个变量调整为相邻的变量; 及

将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

一种相关变量识别装置, 包括:

获取模块, 用于获取待处理的相关系数表, 所述相关系数表中记载了多个变量彼此之间的相关系数;

查找模块, 用于查找所述相关系数表中相关系数的绝对值大于预设阈值的相关系数, 将所述绝对值大于预设阈值的相关系数作为目标相关系数, 并将所述目标相关系数进行标记;

聚类模块, 用于根据所述目标相关系数将相关的多个变量进行聚类归为同一组, 并为该组分配一个唯一的组号;

调整模块, 用于根据分组的组号调整所述相关系数表中变量的排列顺序, 将具有同一组号的多个变量调整为相邻的变量; 及

显示模块, 用于将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

一种终端, 包括存储器和处理器, 所述存储器中存储有计算机可读指令, 所述计算机可读指令被所述处理器执行时, 使得所述处理器执行以下步骤:

获取待处理的相关系数表, 所述相关系数表中记载了多个变量彼此之间的相关系数;

查找所述相关系数表中相关系数的绝对值大于预设阈值的相关系数, 将所述绝对值大于预设阈值的相关系数作为目标相关系数, 并将所述目标相关系数进行标记;

根据所述目标相关系数将相关的多个变量进行聚类归为同一组, 并为该

组分配一个唯一的组号；

根据分组的组号调整所述相关系数表中变量的排列顺序，将具有同一组号的多个变量调整为相邻的变量；及

将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取待处理的相关系数表，所述相关系数表中记载了多个变量彼此之间的相关系数；

查找所述相关系数表中相关系数的绝对值大于预设阈值的相关系数，将所述绝对值大于预设阈值的相关系数作为目标相关系数，并将所述目标相关系数进行标记；

根据所述目标相关系数将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号；

根据分组的组号调整所述相关系数表中变量的排列顺序，将具有同一组号的多个变量调整为相邻的变量；及

将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

本发明的一个或多个实施例的细节在下面的附图和描述中提出。本发明的其它特征、目的和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为一个实施例中终端的内部结构框图；

图 2 为一个实施例中相关变量识别方法流程图；

图 3A 为传统方法的部分识别结果的示意图；

图 3B 为一个实施例中部分识别结果的示意图；

5 图 4 为一个实施例中根据目标相关系数将相关的多个变量聚类为同一组的方法流程图；

图 5 为一个实施例中若目标变量未被分组则为该目标变量分配组号的方法流程图；

图 6 为另一个实施例中相关变量识别方法流程图；

10 图 7 为一个实施例中相关变量识别装置的结构框图；

图 8 为一个实施例中聚类模块的结构框图；

图 9 为另一个实施例中相关变量识别装置的结构框图。

具体实施方式

15 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

如图 1 所示，在一个实施例中，终端 102 的内部结构如图 1 所示，包括通过系统总线连接的处理器、内存储器、非易失性存储介质、网络接口、显示屏和输入装置。其中，终端 102 的非易失性存储介质存储有操作系统和计算机可读指令，该计算机可读指令可被处理器执行以实现适用于终端 102 的一种相关变量识别方法。该处理器用于提供计算和控制能力，支撑整个终端的运行。终端中的内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。网络接口用于连接到网络进行通信。终端 102 的显示屏可以是液晶显示屏或者电子墨水显示屏等，输入装置可以是显示屏上覆盖的触摸层，也可以是电子设备外壳上设置的按键、轨迹球或触控板，也可以是外接的键盘、触控板或鼠标等。该终端可以是平板电脑、笔记本电脑、台

20

25

式计算机等。本领域技术人员可以理解，图 1 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的终端的限定，具体的终端可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

5 如图 2 所示，在一个实施例中，提出了一种相关变量识别方法，该方法可应用于如图 1 所示的终端中，其包括以下步骤：

步骤 202，获取待处理的相关系数表，相关系数表中记载了多个变量彼此之间的相关系数。

在本实施例中，在终端中进行数据建模的过程中往往会涉及到很多个变
10 量，当变量彼此之间的相关性比较高时，即变量之间的相关系数的绝对值比较大时，变量之间会表现出很强的共线性，容易造成模型失真。所以为了避免模型失真，需要识别出变量相关性比较高的变量进行相应的处理。首先，获取待处理的相关系数表，其中，相关系数表是预先通过计算多个变量两两之间的相关性得到的，里面记载了两两变量之间的相关系数，相关系数的绝
15 对值代表了两个变量之间的相关性，绝对值越大，说明两个变量之间的相关性越大。具体的，为了识别出相关性比较高的变量，首先，终端根据用户的指令获取待处理的相关系数表，相关系数表中记载了多个变量两两之间的相关系数，表 1 为一个实施例中相关系数表的示意图：

表 1

变量名称	A1	A2	A3	...	An
A1	1	0.007482	0.027993	...	0.684049
A2	0.007482	1	0.835227	...	0.472902
A3	0.027993	0.835227	1	...	-0.616960
...	...	...	...	1	...
An	0.684049	0.472902	-0.616960	...	1

20 其中，A1，A2，A3，... An 表示的是各个变量，任两个变量都对应一个相关系数，比如，A1 和 A2 对应的相关系数为 0.007482（即 A1 行和 A2 列对

应的值或者 A2 行和 A1 列对应的值)。而每个相关系数的绝对值的大小代表了两个变量彼此之间的相关性,绝对值越大,说明两个变量之间的相关性越高。

步骤 204,查找相关系数表中相关系数的绝对值大于预设阈值的相关系数,将绝对值大于预设阈值的相关系数作为目标相关系数,并将目标相关系数进行标记。

在本实施例中,在终端中预先设置相关系数的阈值,两个变量之间的相关系数只有大于该预设阈值才认为两个变量相关,否则,说明该两个变量不相关。查找相关系数表中相关系数的绝对值大于预设阈值的相关系数,将找到的相关系数作为目标相关系数,并将这些目标相关系数进行标记。具体地,比如,预设阈值设为 0.75,凡是相关系数的绝对值大于该预设阈值的就说明两个变量相关。遍历整个相关系数表,找到所有相关系数的绝对值大于该预设阈值(>0.75)的相关系数,并将这些相关系数作为目标相关系数,然后将这些目标相关系数进行标记,便于后续进行对应的处理。

步骤 206,根据目标相关系数将具有相关性的多个变量进行聚类归为同一组,并为该组分配一个唯一的组号。

在本实施例中,相关系数表中的相关系数代表的是两个变量的相关性,只有相关性大于预设阈值,两个变量才认为是相关的,否则就是不相关的。故,两个变量相关定义为:如果相关系数的绝对值大于预设阈值,则对应的两个变量相关;三个变量相关定义为:如果 A 和 B 的相关系数绝对值大于预设阈值, B 和 C 的相关系数绝对值大于预设阈值,则 A、B 和 C 相关。也就是说,三个变量相关并不要求彼此都相关,只需要两两相关的变量具有一个共同的变量即可。即当 A 和 B 相关, B 和 C 相关,两者有共同的变量 B 就可以将三者关联起来,此时不管 A 和 C 的相关系数绝对值是否大于预设阈值, A 和 B 和 C 都相关,依次类推,可以得到三个以上的变量是否相关。具体地,由于目标相关系数是指绝对值大于预设阈值的相关系数,所以与目标相关系数对应的两个变量必然是相关的。两个相关的变量称为一对相关变量,如果

两对相关变量具有相同的变量，那么该两对中包括的变量都相关，进一步的，如果有其他变量与这两对变量中的任一变量相关，那么该其他变量也与这三个变量相关，依次类推。比如，A 和 B 相关，B 和 C 相关、C 和 D 相关，D 和 E 相关，那么 A、B、C、D 和 E 这多个变量相关。也就是说，对于相关的几个变量，只要其他变量与该相关的几个变量中的任何一个相关，则该其他变量与这几个变量相关。在本实施例中，终端通过聚类将相关的多个变量归为同一组，即将多个相关的变量分为一组，并为该组分配一个组号，也就是说，具有相同组号的多个变量相关。

步骤 208，根据分组的组号调整相关系数表中变量的排列顺序，将具有同一组号的多个变量调整为相邻的变量。

在本实施例中，终端将相关的多个变量的组号设置为相同后，将具有同一组号的多个变量调整为相邻的变量，即根据分组的组号重新调整相关系数表中各个变量的排列顺序。这样，具有关联的多个变量就聚集在了一起，便于后续可以快速识别相关变量。

步骤 210，将调整后的相关系数表中的具有标记的目标相关系数进行突出显示。

在本实施例中，将具有同一组号的多个变量调整为相邻的变量后，将调整后的相关系数表中的具有标记的目标相关系数进行突出显示，图 3B 为一个实施例中部分识别结果的示意图（由于变量数目往往很多，图 3B 只展示了部分识别结果），其中，加灰色底纹的为目标相关系数，即设置大于 0.75 的相关系数为目标相关系数。多个目标相关系数聚集在一起所对应的多个变量为相关变量。为了将相关的变量聚集在一起，所以打破了原来的顺序排列，这样从图中就可以明显的看出多个变量之间的相关性，从而实现了快速识别多个相关变量，有利于提高建模过程中的速度。传统的只能识别两个变量相关的数据进行显示，若要识别三个以上的变量，则是通过将列表中没有目标相关系数的行和列隐藏，如 3A 所示，然后通过人工识别的方法来找到三个以上的相关变量。如图 3A 中所示，传统的方法目标相关系数分布的比较乱，

需要人工来识别三个以上相关变量。

在本实施例中，通过获取相关系数表，查找相关系数表中相关系数的绝对值大于预设阈值的目标相关系数，根据目标相关系数将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号，根据分组的组号调整相关系数表中变量的排列顺序，将具有同一组号的多个变量调整为相邻的变量，然后将调整后的相关系数表中具有标记的目标相关系数进行突出显示，此时，多个相关变量聚集在了一起，通过将相应的目标相关系数进行突出显示，从而实现了快速识别多个相关变量，从而提高了数据建模速度。

如图 4 所示，在一个实施例中，根据目标相关系数值将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号的步骤包括：

步骤 206A，获取待聚类的目标变量，判断该目标变量是否已经被分组，若已经被分组，则进入步骤 206B，若未被分组，则进入步骤 206C。

在本实施例中，将相关的多个变量采用遍历的方法进行聚类，首先，确定一个待聚类的目标变量，然后找出与这个目标变量相关的其他变量。具体地，将要被聚类的变量称为目标变量，然后判断该目标变量是否已经被分组，若已经被分组，则需要根据目标相关系数获取与目标变量相关的第一变量，因为目标相关系数对应的两个变量是相关变量，所以首先根据目标相关系数获取与目标变量相关的第一变量，然后在再查找与第一变量组号相同的第二变量，将第一变量和第二变量的组号修改为与目标变量的组号相同。若未被分组，则获取与目标变量相关的第一变量以及和第一变量组号相同的第二变量，为该目标变量、第一变量和第二变量分配一个新的组号。

步骤 206B，根据目标相关系数获取与目标变量相关的第一变量以及和第一变量组号相同的第二变量，将第一变量和第二变量的组号修改为与目标变量的组号相同。

在本实施例中，若在终端中目标变量已经被分组，说明目标变量已经有了组号，那么查找与该目标变量相关的第一变量，然后还要查找与该第一变

量组号相同的第二变量，第一变量和第二变量组号相同说明两者已经是相关的变量，这样，就可以将与目标变量直接相关的第一变量、以及与目标变量间接相关的第二变量全部查找到，从而就找到了与目标变量相关的所有变量，然后将与该目标变量相关的所有变量的组号修改为与该目标变量的组号相同，即将第一变量和第二变量的组号修改为与目标变量的组号相同。其中，第一变量用来表示与目标变量直接相关的变量，第二变量用来表示通过第一变量与目标变量相关的变量。第一变量和第二变量分别表示的是一类变量，并不用于限制数量。具体的，假设目标变量为 A，首先，查找与目标变量 A 相关的第一变量，假设查找到的第一变量为 B、C，然后再查找与第一变量组号相同的第二变量，比如，查找到与 B 组号相同的为 D，与 C 组号相同的为 E，那么将 B、C、D 和 E 的组号都设置为与目标变量 A 的组号相同。

步骤 206C，根据目标相关系数获取与目标变量相关的第一变量以及和第一变量组号相同的第二变量，为目标变量、第一变量和第二变量分配一个新的组号。

在本实施例中，若待聚类的目标变量还没有被分组，那么说明该目标变量还没有组号，不过，同样要获取与该目标变量相关的第一变量，以及和第一变量组号相同的第二变量，然后为该目标变量、第一变量和第二变量分配一个新的组号。具体的，可以按照递增的顺序为该目标变量、第一变量和第二变量分配一个新的组号，比如，若 G 表示当前的总组数，则分配新的组号为 G+1。

如图 5 所示，若目标变量未被分组，则获取与目标变量相关的第一变量以及和第一变量组号相同的第二变量，为目标变量、第一变量和第二变量分配一个新的组号的步骤包括：

步骤 502，若目标变量未被分组，则获取与目标变量相关的第一变量以及和第一变量组号相同的第二变量。

在本实施例中，若目标变量未被分组，说明当前目标变量还没有组号，获取与目标变量相关的第一变量，即根据相关系数获取与目标变量直接相关

的第一变量；当第一变量已经被分过组，还需要获取与该第一变量组号相同的第二变量。然后为目标变量、第一变量以及第二变量统一分配一个新的组号。

步骤 504，将目标变量、第一变量以及第二变量的组号统一赋值为 $G+1$ ，其中， G 表示当前的总组数。

在本实施例中，若变量还未被分组，则说明变量还没有组号，为了便于后续可以更直观的看到多个变量之间的关系，采用递增的顺序进行分组，即当目标变量还未被分组，则查找与目标变量相关的第一变量以及和第一变量组号相同的第二变量，然后将目标变量、第一变量以及第二变量的组号统一赋值为 $G+1$ ，其中， G 表示当前的总组数。

在一个具体的实施例中，假设有五个变量， $A1$ 、 $A2$ 、 $A3$ 、 $A4$ 和 $A5$ ，若根据相关系数表中的相关系数可知： $A1$ 与 $A2$ 相关、 $A2$ 和 $A3$ 相关、 $A4$ 和 $A5$ 相关。在开始阶段 $A1$ 、 $A2$ 、 $A3$ 、 $A4$ 和 $A5$ 都还未被分组，采用遍历聚类的方法，首先，将 $A1$ 作为目标变量，查找与目标变量 $A1$ 相关的第一变量，查找到的第一变量只有 $A2$ ，由于 $A2$ 此时还未分组，即 $A2$ 还没有组号，所以此时不存在与 $A2$ 组号相同的第二变量，即查找到的与 $A1$ 相关只有 $A2$ ，此时，为 $A1$ 和 $A2$ 分配一个组号 $G+1$ ，由于之前没有组，即 G 初始为 0，所以分配给 $A1$ 和 $A2$ 的组号为 1。然后将 $A2$ 作为目标变量，此时 $A2$ 已经被分组，同样的，需要查找与目标变量 $A2$ 相关的第一变量，查找到的第一变量有 $A1$ 和 $A3$ ，然后再分别获取与第一变量组号相同的第二变量，由于 $A3$ 还未被分组，所以此时不存在与 $A3$ 相关的第二变量，而 $A1$ 已经被分组，而与 $A1$ 组号相同的只有 $A2$ 本身，所以，查找到的与 $A2$ 相关的只有 $A1$ 和 $A3$ ，那么将 $A2$ 、 $A1$ 和 $A3$ 的组号都修改为与 $A2$ 相同，即组号为 1，依次类推，通过该方法将所有相关的变量聚类到同一组，当然不相关的分别在不同的组。

在一个实施例中，根据分组的组号调整相关系数表中变量的排列顺序，将具有同一组号的多个变量调整为相邻的变量的步骤包括：将具有同一组号的多个变量调整为相邻的变量，并根据组号的大小按照从大到小的顺序调整

相关系数表中变量的排列顺序。

在本实施例中，为了快速识别相关变量，将具有同一组号的多个变量调整为相邻的变量，即将同一组号的多个变量聚集在一起，然后根据组号的大小按照从大到小的顺序调整相关系数表中变量的排列顺序。这样便于更有规律
5 的识别多个相关变量。

如图 6 所示，在一个实施例中，上述相关变量识别方法还包括：

步骤 212，采用主成分分析从相同组号对应的目标相关系数中筛选出一个代表该组的相关系数。

在本实施例中，终端将调整后的相关系数表中的具有标记的目标相关系数进行突出显示识别出多个相关变量后，为了消除共线性，采用主成分分析
10 方法从相同组号对应的多个目标相关系数中筛选出一个代表该组的相关系数，然后根据筛选出的相关系数进行后续的处理，比如，根据筛选出的相关系数建立线性回归模型等。

15 如图 7 所示，在一个实施例中，提出了一种相关变量识别装置 700，该装置包括：

获取模块 702，用于获取待处理的相关系数表，相关系数表中记载了多个变量彼此之间的相关系数。

查找模块 704，用于查找相关系数表中相关系数的绝对值大于预设阈值
20 的相关系数，将大于预设阈值的相关系数作为目标相关系数，并将目标相关系数进行标记。

聚类模块 706，用于根据目标相关系数将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号。

调整模块 708，用于根据分组的组号调整相关系数表中变量的排列顺序，
25 将具有同一组号的多个变量调整为相邻的变量。

显示模块 710，用于将调整后的相关系数表中的具有标记的目标相关系数进行突出显示。

如图 8 所示, 在一个实施例中, 聚类模块 706 包括:

判断模块 706A, 用于获取待聚类的目标变量, 判断该目标变量是否已经被分组。

组号修改模块 706B, 用于若目标变量已经被分组, 则根据目标相关系数
5 获取与目标变量相关的第一变量以及和第一变量组号相同的第二变量, 将第一变量和第二变量的组号修改为与目标变量的组号相同。

组号分配模块 706C, 用于若目标变量未被分组, 则获取与目标变量相关的第一变量以及和第一变量组号相同的第二变量, 为目标变量、第一变量和第二变量分配一个新的组号。

10 在一个实施例中, 组号分配模块还用于若目标变量未被分组, 则获取与目标变量相关的第一变量以及和第一变量组号相同的第二变量, 按照组号递增的规则将目标变量、第一变量以及第二变量的组号统一赋值为 $G+1$, 其中, G 表示当前的总组数。

15 在一个实施例中, 调整模块还用于将具有同一组号的多个变量调整为相邻的变量, 并按照组号的大小从大到小的顺序调整相关系数表中变量的排列顺序。

如图 9 所示, 在一个实施例中, 提出了一种相关变量识别装置 900, 除了包括模块 712 至模块 710, 还包括:

20 筛选模块 712, 用于采用主成分分析从相同组号对应的目标相关系数中筛选出一个代表该组的相关系数。

上述基于相关变量识别装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。其中, 网络接口可以是以太网卡或无线网卡等。上述各模块可以硬件形式内嵌于或独立于终端中的处理器中, 也可以以软件形式存储于终端中的存储器中, 以便于处理器调用执行以上各个模块对应的操作。该
25 处理器可以为中央处理单元 (CPU)、微处理器、单片机等。

在一个实施例中, 图 1 终端中所述计算机可读指令被处理器执行时, 使

得处理器执行以下步骤：获取待处理的相关系数表，所述相关系数表中记载了多个变量彼此之间的相关系数；查找所述相关系数表中相关系数的绝对值大于预设阈值的相关系数，将所述绝对值大于预设阈值的相关系数作为目标相关系数，并将所述目标相关系数进行标记；根据所述目标相关系数将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号；根据分组的组号调整所述相关系数表中变量的排列顺序，将具有同一组号的多个变量调整为相邻的变量；及将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

在一个实施例中，所述处理器所执行根据所述目标相关系数将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号的步骤包括：获取待聚类的目标变量，判断该目标变量是否已经被分组；若所述目标变量已经被分组，则根据所述目标相关系数获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量，将所述第一变量和第二变量的组号修改为与所述目标变量的组号相同；若所述目标变量未被分组，则根据所述目标相关系数获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量，为所述目标变量、第一变量和第二变量分配一个新的组号。

在一个实施例中，所述处理器所执行的若所述目标变量未被分组，则获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量，为所述目标变量、第一变量和第二变量分配一个新的组号包括：若所述目标变量未被分组，则获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量；按照组号递增的规则将所述目标变量、第一变量以及第二变量的组号统一赋值为 $G+1$ ，其中， G 表示当前的总组数。

在一个实施例中，所述处理器所执行的所述根据分组的组号调整所述相关系数表中变量的排列顺序，将具有同一组号的多个变量调整为相邻的变量的步骤包括：将具有同一组号的多个变量调整为相邻的变量，并按照组号的大小从大到小的顺序调整相关系数表中变量的排列顺序。

在一个实施例中，所述处理器还用于执行以下步骤：采用主成分分析从

相同组号对应的目标相关系数中筛选出一个代表该组的相关系数。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，该计算机程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施
5 5 施例的流程。其中，前述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory, ROM）等非易失性存储介质，或随机存储记忆体（Random Access Memory, RAM）等。

以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未
10 10 对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

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
15 15 变形和改进，这些都属于本发明的保护范围。因此，本发明的保护范围应以所附权利要求为准。

权利要求书

1、一种相关变量识别方法，包括：

获取待处理的相关系数表，所述相关系数表中记载了多个变量彼此之间的相关系数；

查找所述相关系数表中相关系数的绝对值大于预设阈值的相关系数，将
5 所述绝对值大于预设阈值的相关系数作为目标相关系数，并将所述目标相关系数进行标记；

根据所述目标相关系数将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号；

根据分组的组号调整所述相关系数表中变量的排列顺序，将具有同一组
10 号的多个变量调整为相邻的变量；及

将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

2、根据权利要求1所述的方法，其特征在于，所述根据所述目标相关系数将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号包
15 括：

获取待聚类的目标变量，判断该目标变量是否已经被分组；

若所述目标变量已经被分组，则根据所述目标相关系数获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量，将所述第一变量和第二变量的组号修改为与所述目标变量的组号相同；

20 若所述目标变量未被分组，则根据所述目标相关系数获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量，为所述目标变量、第一变量和第二变量分配一个新的组号。

3、根据权利要求2所述的方法，其特征在于，所述若所述目标变量未被分组，则获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的
25 的第二变量，为所述目标变量、第一变量和第二变量分配一个新的组号包括：

若所述目标变量未被分组，则获取与所述目标变量相关的第一变量以及

和所述第一变量组号相同的第二变量;

按照组号递增的规则将所述目标变量、第一变量以及第二变量的组号统一赋值为 $G+1$, 其中, G 表示当前的总组数。

4、根据权利要求 1 所述的方法, 其特征在于, 所述根据分组的组号调整所述相关系数表中变量的排列顺序, 将具有同一组号的多个变量调整为相邻的变量包括:

将具有同一组号的多个变量调整为相邻的变量, 并按照组号的大小从大到小的顺序调整相关系数表中变量的排列顺序。

5、根据权利要求 1 所述的方法, 其特征在于, 所述方法还包括:

采用主成分分析从相同组号对应的目标相关系数中筛选出一个代表该组的相关系数。

6、一种相关变量识别装置, 包括:

获取模块, 用于获取待处理的相关系数表, 所述相关系数表中记载了多个变量彼此之间的相关系数;

查找模块, 用于查找所述相关系数表中相关系数的绝对值大于预设阈值的相关系数, 将所述绝对值大于预设阈值的相关系数作为目标相关系数, 并将所述目标相关系数进行标记;

聚类模块, 用于根据所述目标相关系数将相关的多个变量进行聚类归为同一组, 并为该组分配一个唯一的组号;

调整模块, 用于根据分组的组号调整所述相关系数表中变量的排列顺序, 将具有同一组号的多个变量调整为相邻的变量; 及

显示模块, 用于将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

7、根据权利要求 6 所述的装置, 其特征在于, 所述聚类模块包括:

判断模块, 用于获取待聚类的目标变量, 判断该目标变量是否已经被分组;

组号修改模块, 用于若所述目标变量已经被分组, 则根据所述目标相关

系数获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量，将所述第一变量和第二变量的组号修改为与所述目标变量的组号相同；

组号分配模块，用于若所述目标变量未被分组，则获取与所述目标变量
5 相关的第一变量以及和所述第一变量组号相同的第二变量，为所述目标变量、第一变量和第二变量分配一个新的组号。

8、根据权利要求7所述的装置，其特征在于，所述组号分配模块还用于若所述目标变量未被分组，则获取与所述目标变量相关的第一变量以及和所述
10 第一变量组号相同的第二变量，按照组号递增的规则将所述目标变量、第一变量以及第二变量的组号统一赋值为 $G+1$ ，其中， G 表示当前的总组数。

9、根据权利要求6所述的装置，其特征在于，所述调整模块还用于将具有同一组号的多个变量调整为相邻的变量，并按照组号的大小从大到小的顺序调整相关系数表中变量的排列顺序。

10、根据权利要求6所述的装置，其特征在于，所述装置还包括：

15 筛选模块，用于采用主成分分析从相同组号对应的目标相关系数中筛选出一个代表该组的相关系数。

11、一种终端，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行以下步骤：

20 获取待处理的相关系数表，所述相关系数表中记载了多个变量彼此之间的相关系数；

查找所述相关系数表中相关系数的绝对值大于预设阈值的相关系数，将所述绝对值大于预设阈值的相关系数作为目标相关系数，并将所述目标相关系数进行标记；

25 根据所述目标相关系数将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号；

根据分组的组号调整所述相关系数表中变量的排列顺序，将具有同一组

号的多个变量调整为相邻的变量；及

将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

12、根据权利要求 11 所述的终端，其特征在于，所述处理器所执行根据
5 所述目标相关系数将相关的多个变量进行聚类归为同一组，并为该组分配一个唯一的组号的步骤包括：

获取待聚类的目标变量，判断该目标变量是否已经被分组；

若所述目标变量已经被分组，则根据所述目标相关系数获取与所述目标
10 变量相关的第一变量以及和所述第一变量组号相同的第二变量，将所述第一变量和第二变量的组号修改为与所述目标变量的组号相同；

若所述目标变量未被分组，则根据所述目标相关系数获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量，为所述目标变量、第一变量和第二变量分配一个新的组号。

13、根据权利要求 12 所述的终端，其特征在于，所述处理器所执行的若
15 所述目标变量未被分组，则获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量，为所述目标变量、第一变量和第二变量分配一个新的组号包括：

若所述目标变量未被分组，则获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量；

20 按照组号递增的规则将所述目标变量、第一变量以及第二变量的组号统一赋值为 $G+1$ ，其中， G 表示当前的总组数。

14、根据权利要求 11 所述的终端，其特征在于，所述处理器所执行的所述根据分组的组号调整所述相关系数表中变量的排列顺序，将具有同一组号的多个变量调整为相邻的变量的步骤包括：

25 将具有同一组号的多个变量调整为相邻的变量，并按照组号的大小从大到小的顺序调整相关系数表中变量的排列顺序。

15、根据权利要求 11 所述的终端，其特征在于，所述处理器还用于执行

以下步骤:

采用主成分分析从相同组号对应的目标相关系数中筛选出一个代表该组的相关系数。

16、一个或多个存储有计算机可读指令的非易失性可读存储介质, 所述计算机可读指令被一个或多个处理器执行时, 使得所述一个或多个处理器执行以下步骤:

获取待处理的相关系数表, 所述相关系数表中记载了多个变量彼此之间的相关系数;

查找所述相关系数表中相关系数的绝对值大于预设阈值的相关系数, 将所述绝对值大于预设阈值的相关系数作为目标相关系数, 并将所述目标相关系数进行标记;

根据所述目标相关系数将相关的多个变量进行聚类归为同一组, 并为该组分配一个唯一的组号;

根据分组的组号调整所述相关系数表中变量的排列顺序, 将具有同一组号的多个变量调整为相邻的变量; 及

将调整后的相关系数表中的具有标记的所述目标相关系数进行突出显示。

17、根据权利要求 16 所述的非易失性可读存储介质, 其特征在于, 所述处理器所执行的根据所述目标相关系数将相关的多个变量进行聚类归为同一组, 并为该组分配一个唯一的组号的步骤包括:

获取待聚类的目标变量, 判断该目标变量是否已经被分组;

若所述目标变量已经被分组, 则根据所述目标相关系数获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量, 将所述第一变量和第二变量的组号修改为与所述目标变量的组号相同;

若所述目标变量未被分组, 则根据所述目标相关系数获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量, 为所述目标变量、第一变量和第二变量分配一个新的组号。

18、根据权利要求 17 所述的非易失性可读存储介质，其特征在于，所述处理器所执行的若所述目标变量未被分组，则获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量，为所述目标变量、第一变量和第二变量分配一个新的组号包括：

5 若所述目标变量未被分组，则获取与所述目标变量相关的第一变量以及和所述第一变量组号相同的第二变量；

按照组号递增的规则将所述目标变量、第一变量以及第二变量的组号统一赋值为 $G+1$ ，其中， G 表示当前的总组数。

19、根据权利要求 16 所述的非易失性可读存储介质，其特征在于，所述
10 处理器所执行的所述根据分组的组号调整所述相关系数表中变量的排列顺序，将具有同一组号的多个变量调整为相邻的变量的步骤包括：

将具有同一组号的多个变量调整为相邻的变量，并按照组号的大小从大到小的顺序调整相关系数表中变量的排列顺序。

20、根据权利要求 16 所述的非易失性可读存储介质，其特征在于，所述
15 处理器还用于执行以下步骤：

采用主成分分析从相同组号对应的目标相关系数中筛选出一个代表该组的相关系数。

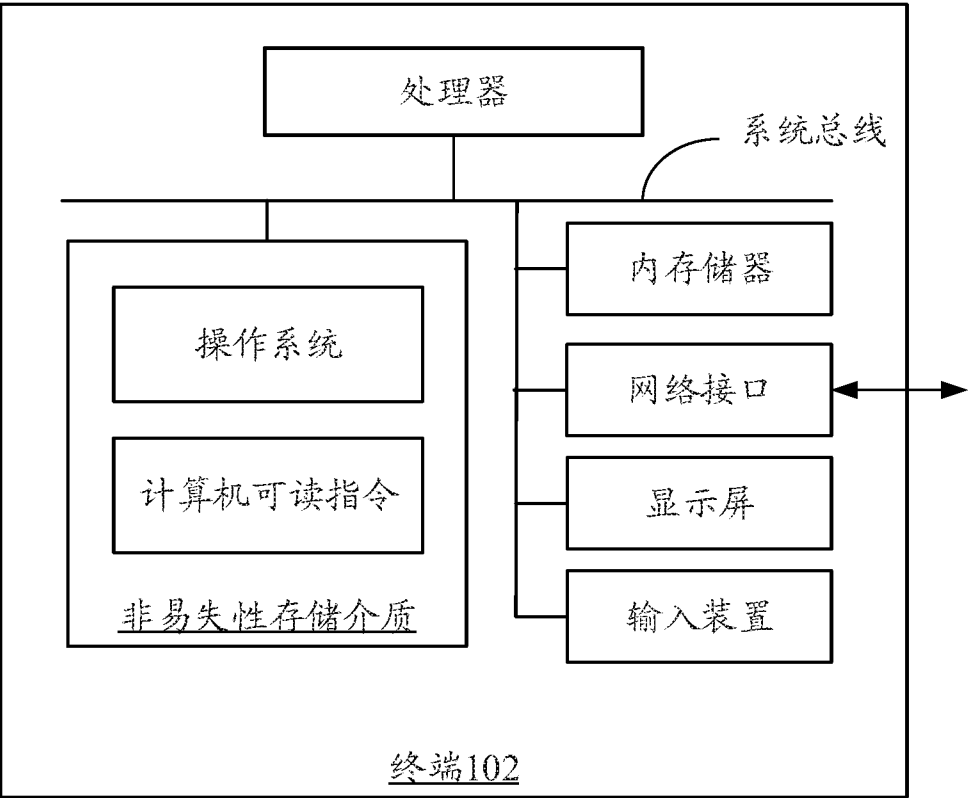


图 1

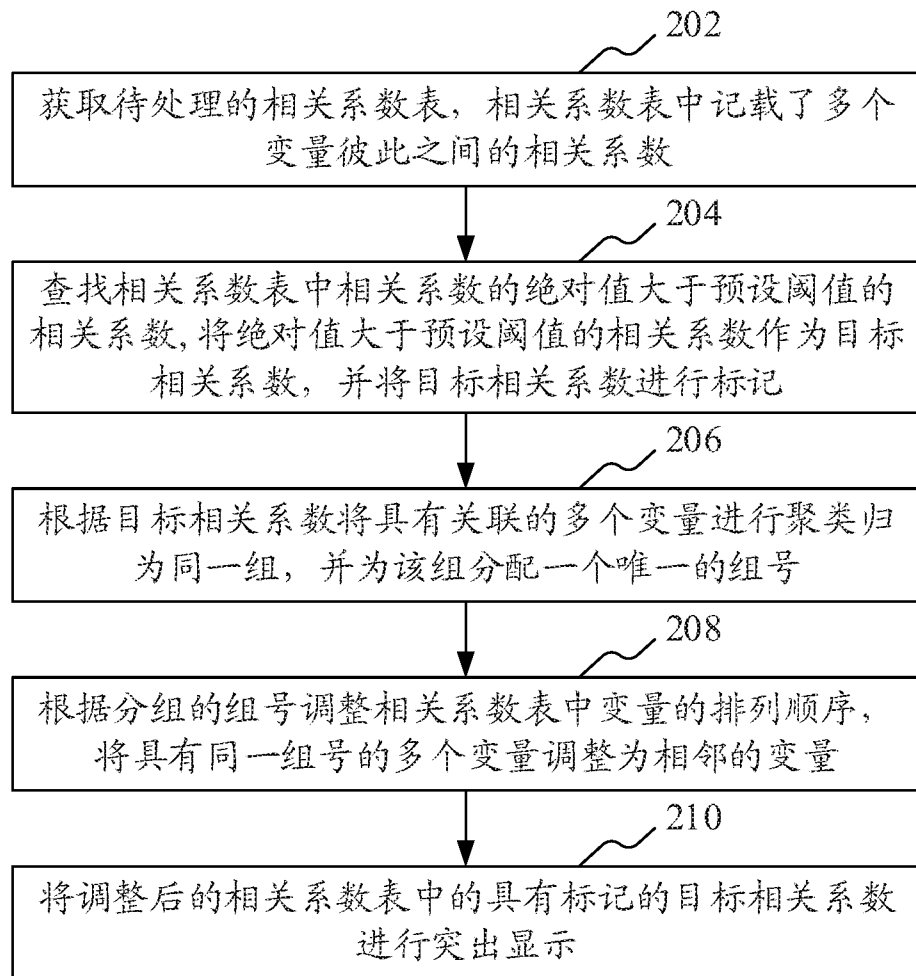


图 2

NAME	A3	A4	A5	A6	A7	A8	A10	A11	A12	A14	A18	A19	A21	A35
A3		0.558506	0.618657	0.650419	0.682635	0.650026	0.457825	0.651911	0.607511	0.470215	0.12091	0.144574	0.120957	0.640448
A4	0.558506		0.818173	0.537445	0.804938	0.510088	0.521656	0.727416	0.461769	0.417218	0.062686	0.094817	0.064005	0.533655
A5	0.618657	0.818173		0.565915	0.941708	0.544917	0.837368	0.856383	0.491573	0.435058	0.173675	0.211968	0.174317	0.571604
A6	0.650419	0.537445	0.565915		0.714946	0.908023	0.378535	0.734741	0.26623	0.4547	0.186112	0.220825	0.1868	0.884591
A7	0.682635	0.804938	0.941708	0.714946		0.678739	0.786733	0.906384	0.623991	0.478199	0.185121	0.224465	0.185883	0.692771
A8	0.650026	0.510088	0.544917	0.908023	0.678739		0.365607	0.698172	0.937608	0.46464	0.182896	0.215782	0.183564	0.974601
A10	0.457825	0.521656	0.837368	0.378535	0.786733	0.365607		0.720159	0.32339	0.296784	0.23923	0.27179	0.239315	0.385876
A11	0.651911	0.727416	0.856383	0.734741	0.906384	0.698172	0.720159		0.706515	0.458316	0.236427	0.2773	0.236836	0.703905
A12	0.607511	0.461769	0.491573	0.26623	0.623991	0.937608	0.32339	0.706515		0.431171	0.193554	0.224302	0.193952	0.912738
A14	0.470215	0.417218	0.435058	0.4547	0.478199	0.46464	0.296784	0.458316	0.431171		0.030637	0.042647	0.030384	0.478741
A18	0.12091	0.062686	0.173675	0.186112	0.185121	0.182896	0.23923	0.236427	0.193554	0.030637		0.299122	0.979027	0.171961
A19	0.144574	0.094817	0.211968	0.220825	0.224465	0.215782	0.27179	0.2773	0.224302	0.042647	0.299122		0.803009	0.20822
A21	0.120957	0.064005	0.174317	0.1868	0.185883	0.183564	0.239315	0.236836	0.193952	0.030384	0.979027	0.803009		0.17294
A35	0.640448	0.533655	0.571604	0.884591	0.692771	0.974601	0.385876	0.703905	0.912738	0.478741	0.171961	0.20822	0.17294	

图 3A

NAME	A72	A74	A71	A73	A58	A61	A66	A88	A55	A57	A59	A60	A63	A65
A72		0.83523	0.68405	0.59739	0.60212	0.63175	0.56178	0.61297	0.62549	0.65855	0.48912	0.6907	0.48228	0.53188
A74	0.83523		0.56637	0.69696	0.46092	0.48358	0.40911	0.47463	0.59331	0.58798	0.44495	0.59995	0.44104	0.45542
A71	0.68405	0.56637		0.80742	0.43847	0.4535	0.41201	0.4729	0.4852	0.46771	0.29738	0.5029	0.29448	0.41437
A73	0.59739	0.69696	0.80742		0.35332	0.37123	0.32201	0.38529	0.46805	0.42062	0.26957	0.4438	0.2688	0.36118
A58	0.60212	0.46092	0.43847	0.35332		0.95999	0.84467	0.88093	0.44182	0.43781	0.25853	0.61696	0.25408	0.60713
A61	0.63175	0.48358	0.4535	0.37123	0.95999		0.86753	0.91866	0.43086	0.42386	0.24925	0.59263	0.24488	0.58389
A66	0.56178	0.40911	0.41201	0.32201	0.84467	0.86753		0.80894	0.37443	0.35433	0.18518	0.50882	0.18107	0.57504
A88	0.61297	0.47463	0.4729	0.38529	0.88093	0.91866	0.80894		0.41808	0.41467	0.25436	0.56219	0.24958	0.54367
A55	0.62549	0.59331	0.4852	0.46805	0.44182	0.43086	0.37443	0.41808		0.80396	0.50653	0.77572	0.50024	0.64219
A57	0.65855	0.58798	0.46771	0.42062	0.43781	0.42386	0.35433	0.41467	0.80396		0.82612	0.9173	0.82028	0.7716
A59	0.48912	0.44495	0.29738	0.26957	0.25853	0.24925	0.18518	0.25436	0.50653	0.82612		0.76254	0.99954	0.62007
A60	0.6907	0.59995	0.5029	0.4438	0.61696	0.59263	0.50882	0.56219	0.77572	0.9173	0.76254		0.75806	0.85123
A63	0.48228	0.44104	0.29448	0.2688	0.25408	0.24488	0.18107	0.24958	0.50024	0.82028	0.99954	0.75806		0.61839
A65	0.53188	0.45542	0.41437	0.36118	0.60713	0.58389	0.57504	0.54367	0.64219	0.7716	0.62007	0.85123	0.61839	

图 3B

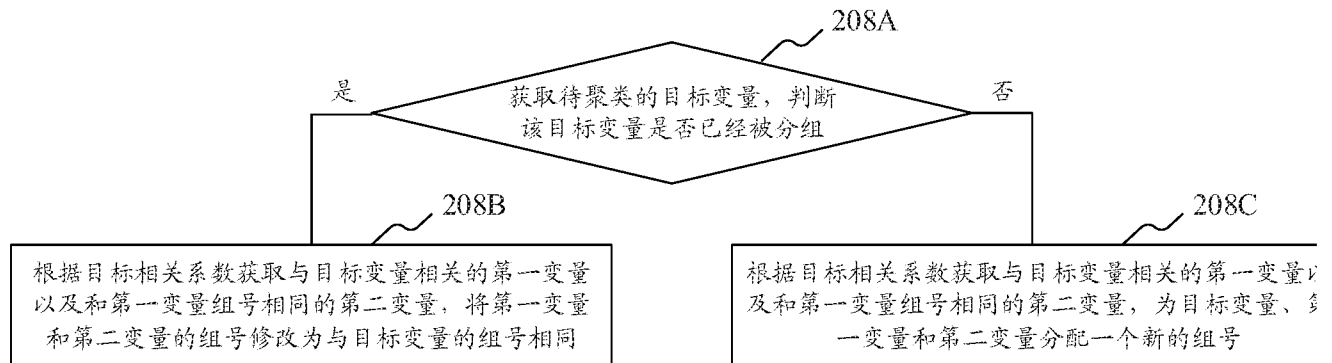


图 4

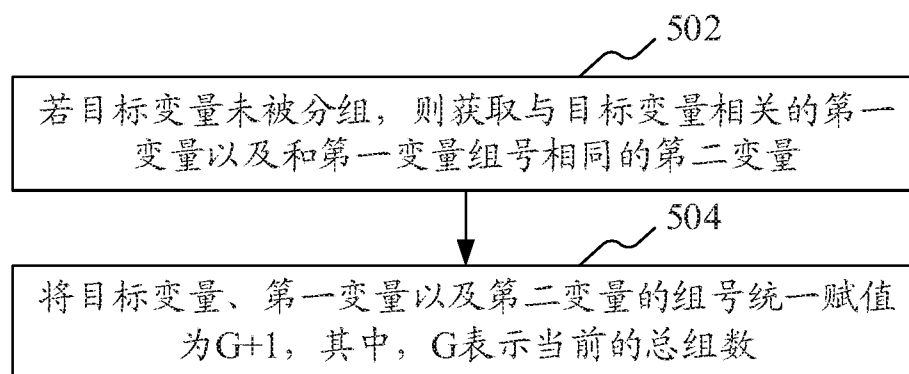


图 5

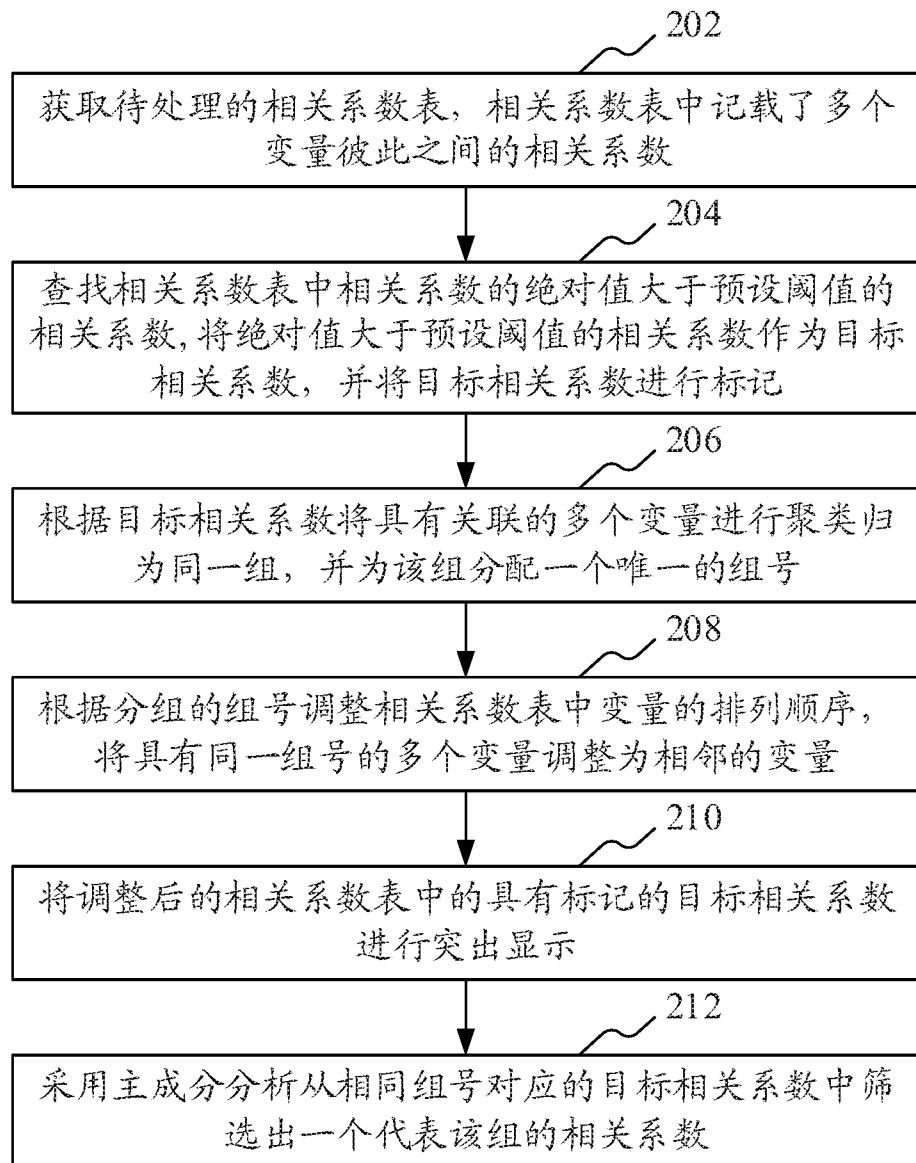


图 6

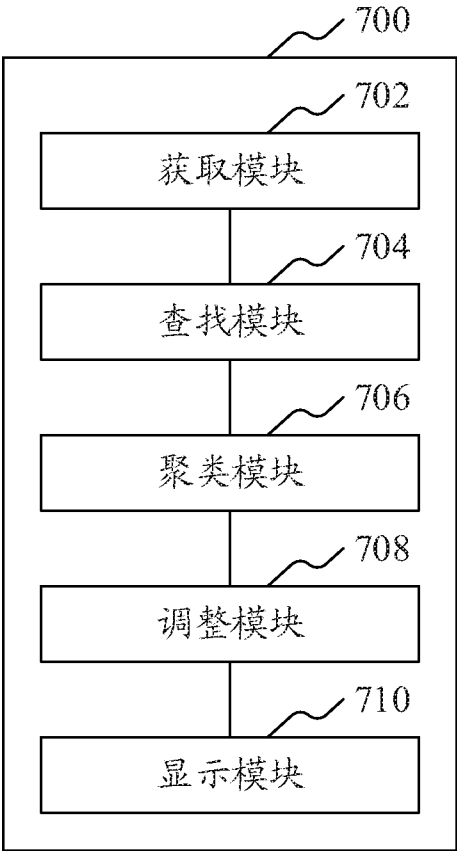


图 7

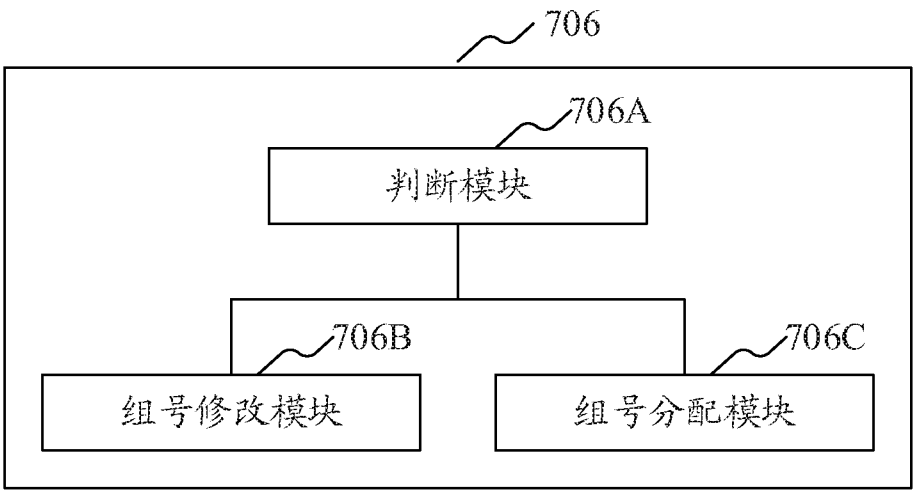


图 8

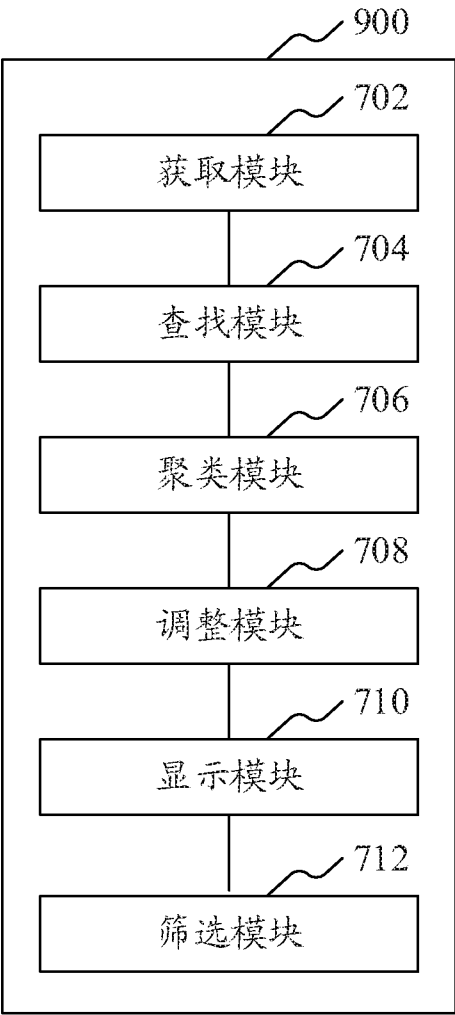


图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/090578

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, EPODOC, WPI, GOOGLE, CNKI, IEEE: 相关, 变量, 系数, 阈值, 聚类, 分组, 排序, 排列, 顺序, 显示, correlation, variable, coefficient, threshold, class, group, cluster, order, display

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104252627 A (GUANGZHOU HUADUO INTERNET TECHNOLOGY CO., LTD.), 31 December 2014 (31.12.2014), the abstract, and description, paragraphs [0037]-[0055] and [0149]	1-20
A	CN 106324405 A (NANJING INSTITUTE OF TECHNOLOGY), 11 January 2017 (11.01.2017), entire document	1-20
A	CN 106156791 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD. et al.), 23 November 2016 (23.11.2016), entire document	1-20
A	US 2015120639 A1 (SAMSUNG SDS CO., LTD.), 30 April 2015 (30.04.2015), entire document	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 07 November 2017	Date of mailing of the international search report 20 November 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer WANG, Ying Telephone No. (86-10) 62414071

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/090578

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104252627 A	31 December 2014	None	
CN 106324405 A	11 January 2017	None	
CN 106156791 A	23 November 2016	None	
US 2015120639 A1	30 April 2015	KR 20150049511 A	08 May 2015
		WO 2015064829 A1	07 May 2015
		KR 101768438 B1	16 August 2017
		CN 104598500 A	06 May 2015
		US 9582736 B2	28 February 2017

国际检索报告

国际申请号

PCT/CN2017/090578

A. 主题的分类 G06K 9/62 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																	
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06K 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, EPDOC, WPI, GOOGLE, CNKI, IEEE: 相关, 变量, 系数, 阈值, 聚类, 分组, 排序, 排列, 顺序, 显示, correlation, variable, coefficient, threshold, class, group, cluster, order, display																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 104252627 A (广州华多网络科技有限公司) 2014年 12月 31日 (2014 - 12 - 31) 摘要, 说明书第[0037]-[0055]段, 第[0149]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 106324405 A (南京工程学院) 2017年 1月 11日 (2017 - 01 - 11) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 106156791 A (北京京东尚科信息技术有限公司 等) 2016年 11月 23日 (2016 - 11 - 23) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2015120639 A1 (SAMSUNG SDS CO., LTD.) 2015年 4月 30日 (2015 - 04 - 30) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 104252627 A (广州华多网络科技有限公司) 2014年 12月 31日 (2014 - 12 - 31) 摘要, 说明书第[0037]-[0055]段, 第[0149]段	1-20	A	CN 106324405 A (南京工程学院) 2017年 1月 11日 (2017 - 01 - 11) 全文	1-20	A	CN 106156791 A (北京京东尚科信息技术有限公司 等) 2016年 11月 23日 (2016 - 11 - 23) 全文	1-20	A	US 2015120639 A1 (SAMSUNG SDS CO., LTD.) 2015年 4月 30日 (2015 - 04 - 30) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 104252627 A (广州华多网络科技有限公司) 2014年 12月 31日 (2014 - 12 - 31) 摘要, 说明书第[0037]-[0055]段, 第[0149]段	1-20															
A	CN 106324405 A (南京工程学院) 2017年 1月 11日 (2017 - 01 - 11) 全文	1-20															
A	CN 106156791 A (北京京东尚科信息技术有限公司 等) 2016年 11月 23日 (2016 - 11 - 23) 全文	1-20															
A	US 2015120639 A1 (SAMSUNG SDS CO., LTD.) 2015年 4月 30日 (2015 - 04 - 30) 全文	1-20															
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																	
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件													
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																
国际检索实际完成的日期	国际检索报告邮寄日期																
2017年 11月 7日	2017年 11月 20日																
ISA/CN的名称和邮寄地址	受权官员																
中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	王莹 电话号码 (86-10)62414071																

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/090578

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	104252627	A	2014年 12月 31日	无			
CN	106324405	A	2017年 1月 11日	无			
CN	106156791	A	2016年 11月 23日	无			
US	2015120639	A1	2015年 4月 30日	KR	20150049511	A	2015年 5月 8日
				WO	2015064829	A1	2015年 5月 7日
				KR	101768438	B1	2017年 8月 16日
				CN	104598500	A	2015年 5月 6日
				US	9582736	B2	2017年 2月 28日