

(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
2016 年 12 月 29 日 (29.12.2016)



(10) 国际公布号
WO 2016/206188 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/088391
- (22) 国际申请日: 2015 年 8 月 28 日 (28.08.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510349469.0 2015 年 6 月 23 日 (23.06.2015) CN
- (71) 申请人: 百度在线网络技术(北京)有限公司
(BAIDU ONLINE NETWORK TECHNOLOGY
(BEIJING) CO., LTD.) [CN/CN]; 中国北京市海淀区
上地十街 10 号百度大厦三层, Beijing 100085 (CN)。
- (72) 发明人: 吴金虎 (WU, Jinhu); 中国北京市海淀区上
地十街 10 号百度大厦三层, Beijing 100085 (CN)。
吴名宇 (WU, Mingyu); 中国北京市海淀区上地十
街 10 号百度大厦三层, Beijing 100085 (CN)。

- (74) 代理人: 北京鸿德海业知识产权代理事务所(普
通合伙) (BEIJING WISPRO INTELLECTUAL
PROPERTY LLP.); 中国北京市海淀区知春路 6 号
锦秋国际大厦 A508, Beijing 100088 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,
GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,
JP, KE, KG, KN, KP, KR, 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, 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)。

(54) Title: DATA PROCESSING METHOD, DEVICE AND APPARATUS FOR OUTPUT OBJECT AND NON-VOLATILE COMPUTER STORAGE MEDIUM

(54) 发明名称: 输出对象的数据处理方法、装置、设备及非易失性计算机存储介质

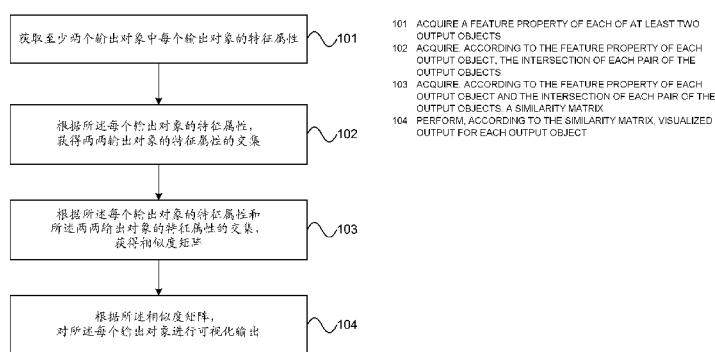


图 1

(57) Abstract: A data processing method, device and apparatus for an output object and a non-volatile computer storage medium. The method comprises: acquiring a feature property of each of at least two output objects (101); then acquiring, according to the feature property of each output object, the intersection of each pair of the output objects (102); acquiring, according to the feature property of each output object and the intersection of each pair of the output objects, a similarity matrix (103); and thereby, performing, according to the similarity matrix, visualized output for each output object (104). A bidirectional relationship between each pair of output objects can be represented by a similarity matrix obtained according to a feature property of each output object and the intersection of each pair of the output objects, and therefore obtained similarities between all pairs of the output objects are the real similarities of the pairs of the output objects, thus improving the reliability of visualization of the output objects.

(57) 摘要:

[见续页]



WO 2016/206188 A1

**本国际公布:**

— 包括国际检索报告(条约第 21 条(3))。

一种输出对象的数据处理方法、装置、设备及非易失性计算机存储介质。该方法通过获取至少两个输出对象中每个输出对象的特征属性(101)，进而根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集(102)，以及根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵(103)，使得能够根据所述相似度矩阵，对所述每个输出对象进行可视化输出(104)，由于根据每个输出对象的特征属性和两两输出对象的特征属性的交集所获得的相似度矩阵，能够表示两两输出对象之间的双向关系，使得所获得的两两输出对象的相似度是两两输出对象真正的相似度，从而提高了输出对象可视化的可靠性。

输出对象的数据处理方法、装置、设备及非易失性计算机存储介质

本申请要求了申请日为 2015 年 06 月 23 日，申请号为 201510349469.0 发明名称为“输出对象的数据处理方法及装置”的中国专利申请的优先权。

技术领域

本发明涉及数据处理技术，特别涉及一种输出对象的数据处理方法、装置、设备及非易失性计算机存储介质。

背景技术

随着互联网的不断发展，不同行业的用户所面临的数据量越来越大，从这些海量数据中找到他们所关心的内容即输出对象，并且可视化是一个很大的挑战。例如，广告商可以通过选择不同的推广产品，来向公众提供推广服务，其需要知道推广产品与其他推广产品之间的关系，或者推广产品与媒体属性、人口属性和地域属性之间的关系，等等，并且还需要将这些多维度关系展现在一个如二维平面空间等较低维度空间上。在展现的过程中，可以将两两输出对象的特征属性的交集与该两两输出对象的特征属性的并集的比值，作为该两两输出对象的相似度。

然而，由于两两输出对象的特征属性的交集与该两两输出对象的特征属性的并集的比值只能够表示两两输出对象之间的单向关系，使得所获得的两两输出对象的相似度并不是两两输出对象真正的相似度，从而导致了输出对象可视化的可靠性的降低。

发明内容

本发明的多个方面提供一种输出对象的数据处理方法、装置、设备及非易失性计算机存储介质，用以提高输出对象可视化的可靠性。

本发明的一方面，提供一种输出对象的数据处理方法，包括：

获取至少两个输出对象中每个输出对象的特征属性；

根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集；

根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵；

根据所述相似度矩阵，对所述每个输出对象进行可视化输出。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述输出对象的特征属性包括：

输出对象所属搜索关键词所对应的人数。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵，包括：

根据两两输出对象的特征属性的交集的平方与该两两输出对象的特征属性的乘积的比值，获得该两两输出对象的相似度；

根据该两两输出对象的相似度，获得所述相似度矩阵。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述根据所述相似度矩阵,对所述每个输出对象进行可视化输出,包括:

对所述相似度矩阵进行 SVD 处理,以获得分解矩阵;

根据所述分解矩阵,获得所述每个输出对象的输出位置;

根据所述每个输出对象的输出位置,输出所述每个输出对象。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述根据所述相似度矩阵,对所述每个输出对象进行可视化输出,包括:

根据所述相似度矩阵,计算两两输出对象的理论距离;

根据预先设置的至少一个特定种子,利用随机函数,获得所述每个输出对象的至少一个随机位置;

分别以所述每个输出对象的至少一个随机位置中每个随机位置为该输出对象的初始位置,根据所述两两输出对象的理论距离,进行迭代处理,以获得所述每个输出对象的迭代位置;

根据所述每个输出对象的迭代位置,获得所述每个输出对象的输出位置;

根据所述每个输出对象的输出位置,输出所述每个输出对象。

如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述根据所述相似度矩阵,计算两两输出对象的距离,包括:

将 1 与两两输出对象的相似度的比值,作为该两两输出对象的距离;
或者

若两两输出对象的相似度不为 0，将 1 与该两两输出对象的相似度的比值的 N 次方，作为该两两输出对象的距离；若两两输出对象的相似度为 0，将 1 与最小的两两输出对象的相似度的比值的 N 次方，作为该两两输出对象的距离； N 为大于 0 且小于 1 的数。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式， N 为 0.5。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述根据所述相似度矩阵，对所述每个输出对象进行可视化输出，包括：

根据所述相似度矩阵，在二维平面空间或三维立体空间上，对所述每个输出对象进行可视化输出。

本发明的另一方面，提供一种输出对象的数据处理装置，包括：

获取单元，用于获取至少两个输出对象中每个输出对象的特征属性；

分析单元，用于根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集；

所述分析单元，还用于根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵；

输出单元，用于根据所述相似度矩阵，对所述每个输出对象进行可视化输出。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，

所述输出对象的特征属性包括：

输出对象所属搜索关键词所对应的人数。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述分析单元，具体用于

根据两两输出对象的特征属性的交集的平方与该两两输出对象的特征属性的乘积的比值，获得该两两输出对象的相似度；以及

根据该两两输出对象的相似度，获得所述相似度矩阵。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述输出单元，具体用于

对所述相似度矩阵进行 SVD 处理，以获得分解矩阵；

根据所述分解矩阵，获得所述每个输出对象的输出位置；以及

根据所述每个输出对象的输出位置，输出所述每个输出对象。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述输出单元，具体用于

根据所述相似度矩阵，计算两两输出对象的理论距离；

根据预先设置的至少一个特定种子，利用随机函数，获得所述每个输出对象的至少一个随机位置；

分别以所述每个输出对象的至少一个随机位置中每个随机位置为该输出对象的初始位置，根据所述两两输出对象的理论距离，进行迭代处理，以获得所述每个输出对象的迭代位置；

根据所述每个输出对象的迭代位置，获得所述每个输出对象的输出位置；以及

根据所述每个输出对象的输出位置，输出所述每个输出对象。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述输出单元，具体用于

将 1 与两两输出对象的相似度的比值，作为该两两输出对象的距离；或者

若两两输出对象的相似度不为 0，将 1 与该两两输出对象的相似度的比值的 N 次方，作为该两两输出对象的距离；若两两输出对象的相似度为 0，将 1 与最小的两两输出对象的相似度的比值的 N 次方，作为该两两输出对象的距离； N 为大于 0 且小于 1 的数。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式， N 为 0.5。

如上所述的方面和任一可能的实现方式，进一步提供一种实现方式，所述输出单元，具体用于

根据所述相似度矩阵，在二维平面空间或三维立体空间上，对所述每个输出对象进行可视化输出。

本发明的另一方面，提供一种设备，包括：

一个或者多个处理器；

存储器；

一个或者多个程序，所述一个或者多个程序存储在所述存储器中，当被所述一个或者多个处理器执行时：

获取至少两个输出对象中每个输出对象的特征属性；

根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集；

根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵；

根据所述相似度矩阵，对所述每个输出对象进行可视化输出。

本发明的另一方面，提供一种非易失性计算机存储介质，所述非易失性计算机存储介质存储有一个或者多个程序，当所述一个或者多个程序被一个设备执行时，使得所述设备：

获取至少两个输出对象中每个输出对象的特征属性；

根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集；

根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵；

根据所述相似度矩阵，对所述每个输出对象进行可视化输出。

由上述技术方案可知，本发明实施例通过获取至少两个输出对象中每个输出对象的特征属性，进而根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集，以及根据所述每个输出对象的特征

属性和所述两两输出对象的特征属性的交集，获得相似度矩阵，使得能够根据所述相似度矩阵，对所述每个输出对象进行可视化输出，由于根据每个输出对象的特征属性和两两输出对象的特征属性的交集所获得的相似度矩阵，能够表示两两输出对象之间的双向关系，使得所获得的两两输出对象的相似度是两两输出对象真正的相似度，从而提高了输出对象可视化的可靠性。

另外，采用本发明所提供的技术方案，通过分别以每个输出对象的至少一个随机位置中每个随机位置为该输出对象的初始位置，根据所述两两输出对象的理论距离，进行多次迭代处理，使得可以获得最优结果，能够进一步提高输出对象可视化的可靠性。

另外，采用本发明所提供的技术方案，预先设置特定种子，使得所进行的多次迭代处理中每个迭代处理的结果较为稳定，能够有效提高输出对象可视化的稳定性。

另外，采用本发明所提供的技术方案，通过改进两两输出对象的理论距离的计算方法，使得所输出的每个输出对象的输出位置间隔有所减小，能够进一步提高输出对象可视化的可靠性。

另外，采用本发明所提供的技术方案，能够有效提高用户体验。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图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性的前提下，还可以根据这些附图获得其他的附图。

图 1 为本发明一实施例提供的输出对象的数据处理方法的流程示意图；

图 2 为本发明另一实施例提供的输出对象的数据处理装置的结构示意图。

具体实施方式

为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的全部其他实施例，都属于本发明保护的范围。

需要说明的是，本发明实施例中所涉及的终端可以包括但不限于手机、个人数字助理（Personal Digital Assistant, PDA）、无线手持设备、平板电脑（Tablet Computer）、个人电脑（Personal Computer, PC）、MP3 播放器、MP4 播放器、可穿戴设备（例如，智能眼镜、智能手表、智能手环等）等。

另外，本文中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。另外，本文中字符“/”，一般表

示前后关联对象是一种“或”的关系。

图 1 为本发明一实施例提供的输出对象的数据处理方法的流程示意图，如图 1 所示。

101、获取至少两个输出对象中每个输出对象的特征属性。

所谓的输出对象，是指能够可视化输出的对象，例如，推广产品与其他推广产品之间的关系，或者推广产品与媒体属性、人口属性和地域属性之间的关系，等等。

所谓的推广产品，是指用以推广用户的信息的网络推广方式，可以包括但不限于搜索推广、网盟推广、品牌广告、百度联盟、广告管家、百度统计、百度商桥等，本实施例对此不进行特别限定。

102、根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集。

其中，两两输出对象的特征属性的交集，是指两个输出对象公共的特征属性。

103、根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵。

在一个具体的实现过程中，所述相似度矩阵中所包含的每个矩阵元素，可以代表一对输出对象的相似度，其主对角线上的矩阵元素，可以代表一个输出对象与自身的相似度。每个输出对象与自身的相似度可以预先指定，例如，0 或 1。

例如，相似度矩阵 T 可以为如下形式：

$$\begin{bmatrix} T_{11} & T_{12} & \cdots & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & \cdots & T_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ T_{n1} & T_{n2} & \cdots & \cdots & T_{nn} \end{bmatrix};$$

其中, T_{ij} ($i \neq j$) 代表一对输出对象的相似度;
 T_{ij} ($i = j$) 代表一个输出对象与自身的相似度。

104、根据所述相似度矩阵,对所述每个输出对象进行可视化输出。

需要说明的是, 101~104 的执行主体的部分或全部可以为位于本地终端的应用, 或者还可以为设置在位于本地终端的应用中的插件或软件开发工具包 (Software Development Kit, SDK) 等功能单元, 或者还可以为位于网络侧服务器中的处理引擎, 或者还可以为位于网络侧的分布式系统, 本实施例对此不进行特别限定。

可以理解的是, 所述应用可以是安装在终端上的本地程序 (nativeApp), 或者还可以是终端上的浏览器的一个网页程序 (webApp), 本实施例对此不进行限定。

这样, 通过获取至少两个输出对象中每个输出对象的特征属性, 进而根据所述每个输出对象的特征属性, 获得两两输出对象的特征属性的交集, 以及根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集, 获得相似度矩阵, 使得能够根据所述相似度矩阵, 对所述每个输出对象进行可视化输出, 由于根据每个输出对象的特征属性和两两输出对象的特征属性的交集所获得的相似度矩阵, 能够表示两两输出对象之间的双向关系, 使得所获得的两两输出对象的相似度是两两输出对象真正的相似度, 从而提高了输出对象可视化的可靠性。

可选地, 在本实施例的一个可能的实现方式中, 所述输出对象的特征属性可以包括但不限于如下属性:

输出对象所属搜索关键词所对应的人数。

可选地，在本实施例的一个可能的实现方式中，在 103 中，具体可以根据两两输出对象的特征属性的交集的平方与该两两输出对象的特征属性的乘积的比值，获得该两两输出对象的相似度，进而，则可以根据该两两输出对象的相似度，获得所述相似度矩阵。

这种实现方式中，实际上可以理解为，将，两两输出对象的特征属性的交集分别与该两两输出对象中的每个输出对象的比值，相乘，得到该两两输出对象的相似度。再由两两输出对象的相似度，构成所述相似度矩阵。

这种根据每个输出对象的特征属性和两两输出对象的特征属性的交集所获得的相似度，能够表示两两输出对象之间的双向关系，这样，所获得的两两输出对象的相似度是两两输出对象真正的相似度，能够更加准确地表征出两两输出对象之间的关系。

可选地，在本实施例的一个可能的实现方式中，在 104 中，具体可以对所述相似度矩阵进行奇异值分解（Singular Value Decomposition, SVD）处理，以获得分解矩阵，进而，则可以根据所述分解矩阵，获得所述每个输出对象的输出位置。然后，则根据所述每个输出对象的输出位置，输出所述每个输出对象。

在一个具体的实现过程中，SVD 处理的详细描述可以参见现有技术中的相关内容，此处不再赘述。

在另一个具体的实现过程中，除了根据所述分解矩阵，获得所述每个输出对象的输出位置即以指定坐标轴定义的输出空间中的坐标值，之外，还可以进一步根据所述分解矩阵，获得该输出空间中的所述指定坐

标轴的贡献度，以供输出给用户。

可选地，在本实施例的一个可能的实现方式中，在 104 中，具体可以根据所述相似度矩阵，计算两两输出对象的理论距离，进而，则可以根据预先设置的至少一个特定种子，利用随机函数，获得所述每个输出对象的至少一个随机位置。接着，分别以所述每个输出对象的至少一个随机位置中每个随机位置为该输出对象的初始位置，根据所述两两输出对象的理论距离，进行迭代处理，以获得所述每个输出对象的迭代位置，进而根据所述每个输出对象的迭代位置，获得所述每个输出对象的输出位置。然后，则根据所述每个输出对象的输出位置，输出所述每个输出对象。

这样，通过分别以每个输出对象的至少一个随机位置中每个随机位置为该输出对象的初始位置，根据所述两两输出对象的理论距离，进行多次迭代处理，使得可以获得最优结果，能够进一步提高输出对象可视化的可靠性。

在这个实现方式中，所谓的输出对象的随机位置，并不是一个真正的随机位置，而实际上是一个伪随机位置。获得这个伪随机位置，所依据的伪随机函数，是以一个种子随机函数作为初始条件，然后用一定的算法不停迭代所产生的。这样，预先设置特定种子，使得所进行的多次迭代处理中每个迭代处理的结果较为稳定，能够有效提高输出对象可视化的稳定性。

在一个具体的实现过程中，具体可以将 1 与两两输出对象的相似度的比值，作为该两两输出对象的距离。这种传统的距离定义方式，所获得的距离的精度非常差，尤其是当一个相似度矩阵中所包含的矩阵元素

的取值相差比较大时,所输出的每个输出对象的输出位置间隔会非常大,这样,会导致输出对象的可视化的可靠性的降低。

在另一个具体的实现过程中,具体可以若两两输出对象的相似度不为 0,将 1 与该两两输出对象的相似度的比值的 N 次方,作为该两两输出对象的距离;若两两输出对象的相似度为 0,将 1 与最小的两两输出对象的相似度的比值的 N 次方,作为该两两输出对象的距离; N 为大于 0 且小于 1 的数。这种改进的距离定义方式,使得所输出的每个输出对象的输出位置间隔有所减小,能够进一步提高输出对象可视化的可靠性。

优选地, N 可以取值为 0.5。

可选地,在本实施例的一个可能的实现方式中,在 104 中,具体可以根据所述相似度矩阵,在二维平面空间或三维立体空间等输出空间上,对所述每个输出对象进行可视化输出。这样,通过将所述每个输出对象在输出空间上进行可视化输出,使得能够清晰地确定每个输出对象之间的关系。

本实施例中,通过获取至少两个输出对象中每个输出对象的特征属性,进而根据所述每个输出对象的特征属性,获得两两输出对象的特征属性的交集,以及根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集,获得相似度矩阵,使得能够根据所述相似度矩阵,对所述每个输出对象进行可视化输出,由于根据每个输出对象的特征属性和两两输出对象的特征属性的交集所获得的相似度矩阵,能够表示两两输出对象之间的双向关系,使得所获得的两两输出对象的相似度是两两输出对象真正的相似度,从而提高了输出对象可视化的可靠性。

另外,采用本发明所提供的技术方案,通过分别以每个输出对象的

至少一个随机位置中每个随机位置为该输出对象的初始位置，根据所述两两输出对象的理论距离，进行多次迭代处理，使得可以获得最优结果，能够进一步提高输出对象可视化的可靠性。

另外，采用本发明所提供的技术方案，预先设置特定种子，使得所进行的多次迭代处理中每个迭代处理的结果较为稳定，能够有效提高输出对象可视化的稳定性。

另外，采用本发明所提供的技术方案，通过改进两两输出对象的理论距离的计算方法，使得所输出的每个输出对象的输出位置间隔有所减小，能够进一步提高输出对象可视化的可靠性。

另外，采用本发明所提供的技术方案，能够有效提高用户体验。

需要说明的是，对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本发明并不受所描述的动作顺序的限制，因为依据本发明，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本发明所必须的。

在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中沒有详述的部分，可以参见其他实施例的相关描述。

图 2 为本发明另一实施例提供的输出对象的数据处理装置的结构示意图，如图 2 所示。本实施例的输出对象的数据处理装置可以包括获取单元 21、分析单元 22 和输出单元 23。其中，获取单元 21，用于获取至少两个输出对象中每个输出对象的特征属性；分析单元 22，用于根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集；

所述分析单元 22, 还用于根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集, 获得相似度矩阵; 输出单元 23, 用于根据所述相似度矩阵, 对所述每个输出对象进行可视化输出。

需要说明的是, 本实施例所提供的输出对象的数据处理装置的执行主体的部分或全部可以为位于本地终端的应用, 或者还可以为设置在位于本地终端的应用中的插件或软件开发工具包 (Software Development Kit, SDK) 等功能单元, 或者还可以为位于网络侧服务器中的处理引擎, 或者还可以为位于网络侧的分布式系统, 本实施例对此不进行特别限定。

可以理解的是, 所述应用可以是安装在终端上的本地程序 (nativeApp), 或者还可以是终端上的浏览器的一个网页程序 (webApp), 本实施例对此不进行限定。

可选地, 在本实施例的一个可能的实现方式中, 所述输出对象的特征属性可以包括但不限于如下属性:

输出对象所属搜索关键词所对应的人数。

可选地, 在本实施例的一个可能的实现方式中, 所述分析单元 22, 具体可以用于根据两两输出对象的特征属性的交集的平方与该两两输出对象的特征属性的乘积的比值, 获得该两两输出对象的相似度; 以及根据该两两输出对象的相似度, 获得所述相似度矩阵。

可选地, 在本实施例的一个可能的实现方式中, 所述输出单元 23, 具体可以用于对所述相似度矩阵进行 SVD 处理, 以获得分解矩阵; 根据所述分解矩阵, 获得所述每个输出对象的输出位置; 以及根据所述每个输出对象的输出位置, 输出所述每个输出对象。

可选地, 在本实施例的一个可能的实现方式中, 所述输出单元 23,

具体可以用于根据所述相似度矩阵，计算两两输出对象的理论距离；根据预先设置的至少一个特定种子，利用随机函数，获得所述每个输出对象的至少一个随机位置；分别以所述每个输出对象的至少一个随机位置中每个随机位置为该输出对象的初始位置，根据所述两两输出对象的理论距离，进行迭代处理，以获得所述每个输出对象的迭代位置；根据所述每个输出对象的迭代位置，获得所述每个输出对象的输出位置；以及根据所述每个输出对象的输出位置，输出所述每个输出对象。

在一个具体的实现过程中，所述输出单元 23，具体可以用于将 1 与两两输出对象的相似度的比值，作为该两两输出对象的距离。

在另一个具体的实现过程中，所述输出单元 23，具体可以用于若两两输出对象的相似度不为 0，将 1 与该两两输出对象的相似度的比值的 N 次方，作为该两两输出对象的距离；若两两输出对象的相似度为 0，将 1 与最小的两两输出对象的相似度的比值的 N 次方，作为该两两输出对象的距离； N 为大于 0 且小于 1 的数。

优选地， N 可以取值为 0.5。

可选地，在本实施例的一个可能的实现方式中，所述输出单元 23，具体可以用于根据所述相似度矩阵，在二维平面空间或三维立体空间上，对所述每个输出对象进行可视化输出。

需要说明的是，图 1 对应的实施例中方法，可以由本实施例提供的输出对象的数据处理装置实现。详细描述可以参见图 1 对应的实施例中的相关内容，此处不再赘述。

本实施例中，通过获取单元获取至少两个输出对象中每个输出对象的特征属性，进而由分析单元根据所述每个输出对象的特征属性，获得

两两输出对象的特征属性的交集，以及根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵，使得输出单元能够根据所述相似度矩阵，对所述每个输出对象进行可视化输出，由于根据每个输出对象的特征属性和两两输出对象的特征属性的交集所获得的相似度矩阵，能够表示两两输出对象之间的双向关系，使得所获得的两两输出对象的相似度是两两输出对象真正的相似度，从而提高了输出对象可视化的可靠性。

另外，采用本发明所提供的技术方案，通过分别以每个输出对象的至少一个随机位置中每个随机位置为该输出对象的初始位置，根据所述两两输出对象的理论距离，进行多次迭代处理，使得可以获得最优结果，能够进一步提高输出对象可视化的可靠性。

另外，采用本发明所提供的技术方案，预先设置特定种子，使得所进行的多次迭代处理中每个迭代处理的结果较为稳定，能够有效提高输出对象可视化的稳定性。

另外，采用本发明所提供的技术方案，通过改进两两输出对象的理论距离的计算方法，使得所输出的每个输出对象的输出位置间隔有所减小，能够进一步提高输出对象可视化的可靠性。

另外，采用本发明所提供的技术方案，能够有效提高用户体验。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本发明所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例

仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能单元的形式实现。

上述以软件功能单元的形式实现的集成的单元，可以存储在一个计算机可读取存储介质中。上述软件功能单元存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备）或处理器（processor）执行本发明各个实施例所述方法的部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

最后应说明的是：以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，本领域的

普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

权利要求书

1、一种输出对象的数据处理方法，其特征在于，包括：

获取至少两个输出对象中每个输出对象的特征属性；

根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集；

根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵；

根据所述相似度矩阵，对所述每个输出对象进行可视化输出。

2、根据权利要求 1 所述的方法，其特征在于，所述输出对象的特征属性包括：

输出对象所属搜索关键词所对应的人数。

3、根据权利要求 1 所述的方法，其特征在于，所述根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵，包括：

根据两两输出对象的特征属性的交集的平方与该两两输出对象的特征属性的乘积的比值，获得该两两输出对象的相似度；

根据该两两输出对象的相似度，获得所述相似度矩阵。

4、根据权利要求 1 所述的方法，其特征在于，所述根据所述相似度矩阵，对所述每个输出对象进行可视化输出，包括：

对所述相似度矩阵进行 SVD 处理，以获得分解矩阵；

根据所述分解矩阵，获得所述每个输出对象的输出位置；

根据所述每个输出对象的输出位置，输出所述每个输出对象。

5、根据权利要求 1 所述的方法，其特征在于，所述根据所述相似度

矩阵, 对所述每个输出对象进行可视化输出, 包括:

根据所述相似度矩阵, 计算两两输出对象的理论距离;

根据预先设置的至少一个特定种子, 利用随机函数, 获得所述每个输出对象的至少一个随机位置;

分别以所述每个输出对象的至少一个随机位置中每个随机位置为该输出对象的初始位置, 根据所述两两输出对象的理论距离, 进行迭代处理, 以获得所述每个输出对象的迭代位置;

根据所述每个输出对象的迭代位置, 获得所述每个输出对象的输出位置;

根据所述每个输出对象的输出位置, 输出所述每个输出对象。

6、根据权利要求 5 所述的方法, 其特征在于, 所述根据所述相似度矩阵, 计算两两输出对象的距离, 包括:

将 1 与两两输出对象的相似度的比值, 作为该两两输出对象的距离;
或者

若两两输出对象的相似度不为 0, 将 1 与该两两输出对象的相似度的比值的 N 次方, 作为该两两输出对象的距离; 若两两输出对象的相似度为 0, 将 1 与最小的两两输出对象的相似度的比值的 N 次方, 作为该两两输出对象的距离; N 为大于 0 且小于 1 的数。

7、根据权利要求 6 所述的方法, 其特征在于, N 为 0.5。

8、根据权利要求 1~7 任一权利要求所述的方法, 其特征在于, 所述根据所述相似度矩阵, 对所述每个输出对象进行可视化输出, 包括:

根据所述相似度矩阵, 在二维平面空间或三维立体空间上, 对所述每个输出对象进行可视化输出。

9、一种输出对象的数据处理装置，其特征在于，包括：

获取单元，用于获取至少两个输出对象中每个输出对象的特征属性；

分析单元，用于根据所述每个输出对象的特征属性，获得两两输出对象的特征属性的交集；

所述分析单元，还用于根据所述每个输出对象的特征属性和所述两两输出对象的特征属性的交集，获得相似度矩阵；

输出单元，用于根据所述相似度矩阵，对所述每个输出对象进行可视化输出。

10、根据权利要求 9 所述的装置，其特征在于，所述输出对象的特征属性包括：

输出对象所属搜索关键词所对应的人数。

11、根据权利要求 9 所述的装置，其特征在于，所述分析单元，具体用于

根据两两输出对象的特征属性的交集的平方与该两两输出对象的特征属性的乘积的比值，获得该两两输出对象的相似度；以及

根据该两两输出对象的相似度，获得所述相似度矩阵。

12、根据权利要求 9 所述的装置，其特征在于，所述输出单元，具体用于

对所述相似度矩阵进行 SVD 处理，以获得分解矩阵；

根据所述分解矩阵，获得所述每个输出对象的输出位置；以及

根据所述每个输出对象的输出位置，输出所述每个输出对象。

13、根据权利要求 9 所述的装置，其特征在于，所述输出单元，具体用于

根据所述相似度矩阵，计算两两输出对象的理论距离；

根据预先设置的至少一个特定种子，利用随机函数，获得所述每个输出对象的至少一个随机位置；

分别以所述每个输出对象的至少一个随机位置中每个随机位置为该输出对象的初始位置，根据所述两两输出对象的理论距离，进行迭代处理，以获得所述每个输出对象的迭代位置；

根据所述每个输出对象的迭代位置，获得所述每个输出对象的输出位置；以及

根据所述每个输出对象的输出位置，输出所述每个输出对象。

14、根据权利要求 13 所述的装置，其特征在于，所述输出单元，具体用于

将 1 与两两输出对象的相似度的比值，作为该两两输出对象的距离；或者

若两两输出对象的相似度不为 0，将 1 与该两两输出对象的相似度的比值的 N 次方，作为该两两输出对象的距离；若两两输出对象的相似度为 0，将 1 与最小的两两输出对象的相似度的比值的 N 次方，作为该两两输出对象的距离； N 为大于 0 且小于 1 的数。

15、根据权利要求 14 所述的装置，其特征在于， N 为 0.5。

16、根据权利要求 9~15 任一权利要求所述的装置，其特征在于，所述输出单元，具体用于

根据所述相似度矩阵，在二维平面空间或三维立体空间上，对所述每个输出对象进行可视化输出。

17、一种设备，包括：

一个或者多个处理器；

存储器；

一个或者多个程序，所述一个或者多个程序存储在所述存储器中，
当被所述一个或者多个处理器执行时：

获取至少两个输出对象中每个输出对象的特征属性；

根据所述每个输出对象的特征属性，获得两两输出对象的特征属性
的交集；

根据所述每个输出对象的特征属性和所述两两输出对象的特征属性
的交集，获得相似度矩阵；

根据所述相似度矩阵，对所述每个输出对象进行可视化输出。

18、一种非易失性计算机存储介质，所述非易失性计算机存储介质
存储有一个或者多个程序，当所述一个或者多个程序被一个设备执行时，
使得所述设备：

获取至少两个输出对象中每个输出对象的特征属性；

根据所述每个输出对象的特征属性，获得两两输出对象的特征属性
的交集；

根据所述每个输出对象的特征属性和所述两两输出对象的特征属性
的交集，获得相似度矩阵；

根据所述相似度矩阵，对所述每个输出对象进行可视化输出。

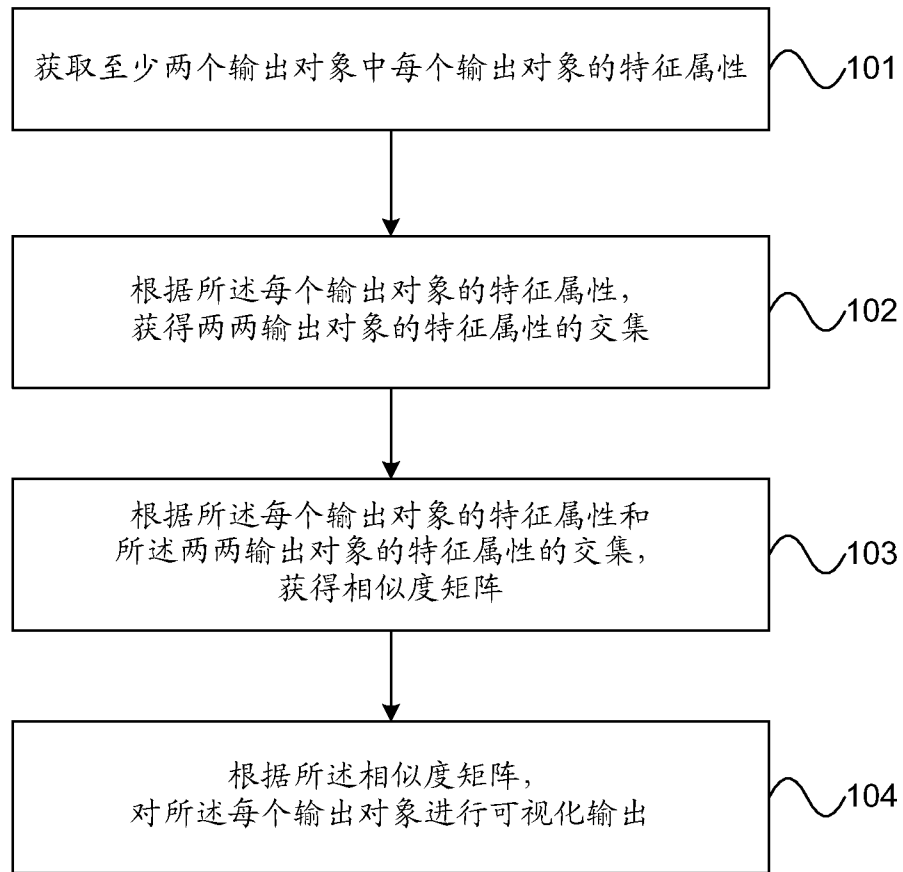


图 1

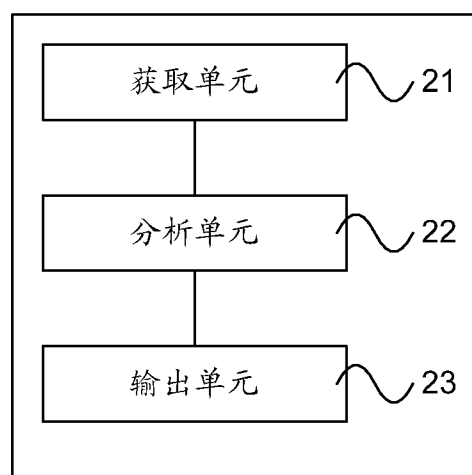


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/088391

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (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)

G06F; H04L; 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, WPI, EPODOC, CNKI: characteristic attribute, similarity matrix, intersection, similarity, output, characteristic, attribute, matrix, visualization

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103246685 A (RICOH CO., LTD.), 14 August 2013 (14.08.2013), description, paragraphs [0024]-[0055] and [0121]-[0126]	1, 2, 8-10, 16, 17, 18
A	CN 102902821 A (BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS), 30 January 2013 (30.01.2013), the whole document	1-18
A	CN 104572825 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 29 April 2015 (29.04.2015), the whole document	1-18
A	US 2008285861 A1 (CANON KABUSHIKI KAISHA), 20 November 2008 (20.11.2008), the whole document	1-18

☐ 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 16 February 2016 (16.02.2016)	Date of mailing of the international search report 01 March 2016 (01.03.2016)
Name and mailing address of the ISA/CN: 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 ZHANG, Qian Telephone No.: (86-10) 62413681

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/088391

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103246685 A	14 August 2013	None	
CN 102902821 A	30 January 2013	None	
CN 104572825 A	29 April 2015	None	
US 2008285861 A1	20 November 2008	JP 2008287437 A	27 November 2008
		JP 2013065336 A	11 April 2013

国际检索报告

国际申请号

PCT/CN2015/088391

A. 主题的分类

G06F 17/30(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)

G06F; H04L; 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, WPI, EPDOC, CNKI:交集, 相似度, 相似, 输出, 特征属性, 相似度矩阵, 可视化, intersection, similarity, output, characteristic, attribute, matrix, visualization

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 103246685 A (株式会社理光) 2013年 8月 14日 (2013 - 08 - 14) 说明书第[0024]段-第[0055]段, 第[0121]段-第[0126]段	1, 2, 8-10, 16, 17, 18
A	CN 102902821 A (北京邮电大学) 2013年 1月 30日 (2013 - 01 - 30) 全文	1-18
A	CN 104572825 A (百度在线网络技术北京有限公司) 2015年 4月 29日 (2015 - 04 - 29) 全文	1-18
A	US 2008285861 A1 (CANON KABUSHIKI KAISHA) 2008年 11月 20日 (2008 - 11 - 20) 全文	1-18

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 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

2016年 2月 16日

国际检索报告邮寄日期

2016年 3月 1日

ISA/CN的名称和邮寄地址

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

授权官员

张千

电话号码 (86-10)62413681

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2015/088391

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	103246685	A	2013年 8月 14日	无			
CN	102902821	A	2013年 1月 30日	无			
CN	104572825	A	2015年 4月 29日	无			
US	2008285861	A1	2008年 11月 20日	JP	2008287437	A	2008年 11月 27日
				JP	2013065336	A	2013年 4月 11日

表 PCT/ISA/210 (同族专利附件) (2009年7月)