

(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
2020 年 3 月 5 日 (05.03.2020)



(10) 国际公布号
WO 2020/042597 A1

- (51) 国际专利分类号:
G06F 16/53 (2019.01) **G06F 16/33** (2019.01)
G06F 16/58 (2019.01)
- (21) 国际申请号: PCT/CN2019/078657
- (22) 国际申请日: 2019 年 3 月 19 日 (19.03.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201811008853.4 2018 年 8 月 31 日 (31.08.2018) CN
- (71) 申请人: 深圳大学 (SHENZHEN UNIVERSITY) [CN/CN]; 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。
- (72) 发明人: 曹文明 (CAO, Wenming); 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。 林秋斌 (LIN, Qiubin); 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。

- (74) 代理人: 深圳市恒申知识产权事务所 (普通合伙) (HENSEN INTELLECTUAL PROPERTY FIRM); 中国广东省深圳市福田区南园路 68 号上步大厦 10H, Guangdong 518000 (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,

(54) Title: CROSS-MODAL RETRIEVAL METHOD AND SYSTEM

(54) 发明名称: 一种跨模态检索方法及系统

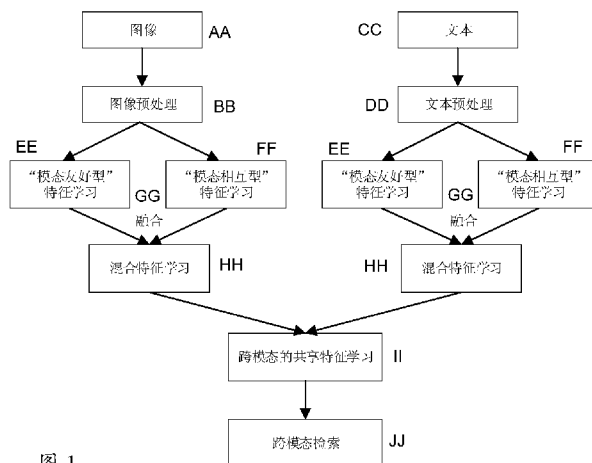


图 1

AA IMAGE
BB IMAGE PRE-PROCESSING
CC TEXT
DD TEXT PRE-PROCESSING
EE "MODAL FRIENDLY" FEATURE LEARNING
FF "MODAL INTERACTIVE" FEATURE LEARNING
GG COMBINATION
HH MIXED FEATURE LEARNING
II CROSS-MODAL SHARED FEATURE LEARNING
JJ CROSS-MODAL RETRIEVAL

(57) Abstract: A cross-modal retrieval method comprises: using a stacked restricted Boltzmann machine and a multimodal deep belief network to extract a modal friendly feature and a modal interactive feature of each of an image and text, wherein the modal friendly feature causes statistical characteristics between the acquired features to be more similar to input characteristics, and the modal interactive feature acquires mutual information lost in an original input instance; combining the two features to acquire a mixed feature; and acquiring a final shared feature by means of multiple bi-modal automatic encodings. In the method, the stacked restricted Boltzmann machine is used to extract internal features of each modality, and the deep belief network is used to mine lost mutual information between modalities; the two features are combined to construct a mixed feature suitable for cross-modal retrieval, and a multi-layer bi-modal automatic encoding network is used to mine cross-modal complex information, thereby effectively improving the accuracy and retrieval speed of a cross-modal retrieval task.

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)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：一种跨模态检索方法，包括采用堆叠式受限玻尔兹曼机和多模态深度置信网络提取图像和文本各自的模态友好型特征和模态相互型特征，模态友好型特征能够使所获得的特征之间的统计特性方面与输入的更相似，模态相互型特征能够得到在原始输入实例中丢失的相互信息，将两种特征进行融合得到混合特征，通过多个双模态自动编码获得最后的共享特征。该方法利用堆叠式的受限玻尔兹曼机提取各个模态的内部特征，采用深度置信网络挖掘模态之间丢失的相互信息，将两种特征进行融合，构造出适合进行跨模态检索的混合特征，利用多层双模态的自动编码网络挖掘跨模态的复杂信息，有效提高了跨模态检索任务的准确性和检索的速度。

一种跨模态检索方法及系统

技术领域

本发明属于检索技术领域，尤其涉及一种跨模态检索方法及系统。

背景技术

跨模态检索是一种新型的检索方法，其能够检索多模态数据。例如，输入图像，在文本数据库中检索对应的文本；给定文本，在图像数据库中找到相应的图像。

当前，基于深度神经网络的跨模态检索方法主要包含两个步骤：（1）提取各个模态的内部特征以及模态之间的特征；（2）在两个模态之间建立各自的共享特征。但是，在第一步骤中，往往丢失了模态之间的相互信息；在第二步骤中，目前的方法使用的都是比较浅层的网络，难以挖掘跨模态的复杂信息。

发明内容

本发明所要解决的技术问题在于提供一种跨模态检索方法及系统，旨在解决现有技术在进行跨模态检索时丢失模态之间的相互关系，难以挖掘跨模态的复杂信息的问题。

本发明是这样实现的，一种跨模态检索方法，包括：

分别对图像和文本进行预处理，得到图像特征和文本特征；

根据所述图像特征和所述文本特征，利用堆叠式的受限玻尔兹曼机提取所述图像的模态友好型特征和所述文本的模态友好型特征，及利用多模态的深度置信网络提取所述图像的模态相互型特征和所述文本的模态相互型特征；

利用联合的自动编码器将所述图像的模态友好型特征和模态相互型特征结合，得到所述图像的混合特征，及将所述文本的模态友好型特征和模态相互型

特征结合，得到所述文本的混合特征；

采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系，得到所述图像和所述文本的共享特征；

根据所述共享特征进行跨模态检索。

进一步地，对图像进行预处理，得到图像特征的步骤包括：

采用 VGGNet 模型提取所述图像 FC7 层的卷积特征。

进一步地，对文本进行预处理，得到文本特征的步骤包括：

采用 bag of words 模型提取所述文本 1000 维的文本特征。

进一步地，所述堆叠式的受限玻尔兹曼机包括三层受限玻尔兹曼机，根据所述图像特征和所述文本特征，利用堆叠式的受限玻尔兹曼机提取所述图像的模态友好型特征和所述文本的模态友好型特征包括：

利用高斯的受限玻尔兹曼机和重复的 softmax 的受限玻尔兹曼机分别提取所述图像特征和所述文本特征；

利用简单的受限玻尔兹曼机移除所述图像特征和所述文本特征各自模态的特性，以获得高层次的语义特征；

完成堆叠式的受限玻尔兹曼机的训练，得到所述图像的模态友好型特征 $X_{friendly}^i$ 和所述文本的模态友好型特征 $X_{friendly}^t$ 。

进一步地，根据所述图像特征和所述文本特征，利用多模态的深度置信网络提取所述图像的模态相互型特征和所述文本的模态相互型特征包括：

分别采用所述图像和所述文本的深度置信网络提取各自模态的特征；

在顶层采用联合的受限玻尔兹曼机进行模态相互型特征的学习，得到所述图像的模态相互型特征 X_{mutual}^i 和所述文本的模态相互型特征 X_{mutual}^t 。

进一步地，所述利用联合的自动编码器将所述图像的模态友好型特征和模态相互型特征结合，得到所述图像的混合特征，及将所述文本的模态友好型特征和模态相互型特征结合，得到所述文本的混合特征包括：

利用联合的自动编码器将所述图像或所述文本各自的模态友好型特征和模

态相互型特征结合，得到所述图像的混合特征和所述文本的混合特征；

利用带有 softmax 层的三层前向神经网络通过有监督学习，挖掘出所述图像和所述文本更有用的混合特征。

进一步地，所述采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系，得到所述图像和所述文本的共享特征包括：

采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系；

根据所述相互关系进行所述图像和所述文本的跨模态的共享特征的学习，得到所述图像和所述文本的共享特征。

进一步地，在提取共享特征时，双模态自动编码网络的输入维度和输出维度一致，共享特征层的维度是输入维度的一半，从共享特征层中，学习得到最后的跨模态的共享特征。

进一步地，将 softmax 层与最后的输出层连接，优化所述共享特征。

本发明实施例还提供了一种跨模态检索系统，包括：

预处理单元，用于分别对图像和文本进行预处理，得到图像特征和文本特征；

特征提取单元，用于根据所述图像特征和所述文本特征，利用堆叠式的受限玻尔兹曼机提取所述图像的模态友好型特征和所述文本的模态友好型特征，及利用多模态的深度置信网络提取所述图像的模态相互型特征和所述文本的模态相互型特征；

特征结合单元，用于利用联合的自动编码器将所述图像的模态友好型特征和模态相互型特征结合，得到所述图像的混合特征，及将所述文本的模态友好型特征和模态相互型特征结合，得到所述文本的混合特征；

特征挖掘单元，用于采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系，得到所述图像和所述文本的

共享特征;

特征检索单元, 用于根据所述共享特征进行跨模态检索。

本发明与现有技术相比, 有益效果在于: 本发明实施例采用堆叠式的受限玻尔兹曼机和多模态的深度置信网络提取图像和文本各自的模态友好型特征和模态相互型特征, 其中模态友好型特征能够使所获得的特征之间的统计特性方面比输入实例之间的更相似, 而模态相互型特征能够得到在原始输入实例中丢失的相互信息, 将所得到的这两种特征进行融合, 得到混合特征, 进一步通过多个双模态自动编码获得最后的共享特征。本发明实施例利用堆叠式的受限玻尔兹曼机提取各个模态的内部特征, 采用多模态的深度置信网络挖掘模态之间丢失的相互信息, 将两种特征进行融合, 并构造出适合进行跨模态检索的混合特征, 利用多层双模态的自动编码网络挖掘跨模态之间的相互关系, 从而获得适合跨模态检索的共享特征, 有效提高了跨模态检索任务的准确性, 提高了检索的速度。

附图说明

图 1 是本发明实施例提供的一种跨模态检索方法的流程图;

图 2 是本发明实施例提供的堆叠式的受限玻尔兹曼机的示意图;

图 3 是本发明实施例提供的多模态的深度置信网络的示意图;

图 4 是本发明实施例提供的双模态自动编码网络的示意图;

图 5 是本发明实施例提供的一种跨模态检索系统的结构示意图。

具体实施方式

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。

图 1 示出了本发明实施例提供的一种跨模态检索方法, 包括:

S101, 分别对图像和文本进行预处理, 得到图像特征和文本特征;

S102, 根据所述图像特征和所述文本特征, 利用堆叠式的受限玻尔兹曼机提取所述图像的模态友好型特征和所述文本的模态友好型特征, 及利用多模态的深度置信网络提取所述图像的模态相互型特征和所述文本的模态相互型特征;

S103, 利用联合的自动编码器将所述图像的模态友好型特征和模态相互型特征结合, 得到所述图像的混合特征, 及将所述文本的模态友好型特征和模态相互型特征结合, 得到所述文本的混合特征;

S104, 采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系, 得到所述图像和所述文本的共享特征;

S105, 根据所述共享特征进行跨模态检索。

下面结合图 2 到图 4 对本发明实施例进行进一步地阐述:

一、预处理阶段:

(1) 对于图像, 采用 VGGNet 模型对原始图像提取“fc7”层的卷积特征;

(2) 对于文本, 采用 bag of words 模型提取 1000 维的文本特征。

当输入的图片或者文本不同时, 其大小、维度是不一致的, 因此预处理阶段, 对图像和文本进行预处理后分别得到的卷积特征和文本特征是为了获得统一大小的维度特征, 如图像是 4096, 文本是 1000。同时, 预处理也能够获得图像和文本比较浅层次的特征。

二、模态友好型特征和模态相互型特征的学习:

(1) 模态友好型特征学习:

如图 2 所示, 利用堆叠式的受限玻尔兹曼机提取该卷积特征和文本特征的模态友好型特征, 以通过堆叠式的受限玻尔兹曼机的提取操作, 获得较高层次的特征。在前两层, 堆叠式的受限玻尔兹曼机利用高斯的受限玻尔兹曼机和重复的 softmax 的受限玻尔兹曼机分别提取该卷积特征和该文本特征。然后, 在第三层, 利用简单的受限玻尔兹曼机移除模态各自的特性, 从而获得高层次的

语义特征，该语义特征即为模态友好型特征，其相对于原始的图像和文本输入具有更好的统计特性。在这包括三层受限玻尔兹曼机进行模态友好型学习的过程中，每一个模态的堆叠式受限玻尔兹曼机的输出概率如下：

$$P(h_i^{(3)}) = \sum_{v_i, h_i^1, h_i^2} P(h_i^{(3)} | h_i^{(2)}) P(h_i^{(2)} | h_i^{(1)}) P(h_i^{(1)} | v_i);$$

$$P(h_t^{(3)}) = \sum_{v_t, h_t^1, h_t^2} P(h_t^{(3)} | h_t^{(2)}) P(h_t^{(2)} | h_t^{(1)}) P(h_t^{(1)} | v_t);$$

其中， v_i 、 v_t 分别表示图像特征和文本特征的输入， $h_i^j (j=1,2,3)$ 表示图像通道的第 j 层隐含层的输出特征， $h_t^j (j=1,2,3)$ 表示文本通道的第 j 层隐含层的输出特征。

在训练完堆叠式受限玻尔兹曼机之后，能够得到图像的模态友好型特征 $X_{friendly}^i$ 、文本的模态友好型特征 $X_{friendly}^t$ ，其模态友好型特征的统计特性比输入的图像和文本的原始模态的统计特性更相似。具体地，图像和文本的模态友好型特征在统计特性方面这两者之间更相似，统计特性的相似性指的是两种模态之间的，并不是特指某一模态，如输入图像和图像的模态友好型特征之间的统计特性。

在本步骤中，第三层受限玻尔兹曼机的输入是第二层的输出，前两层的输入是预处理步骤中得到的卷积特征或者文本特征，经过本步骤的操作，可以认为第三层移除的是原始图像和原始文本的特性，使得最后的输出特征在统计特性方面的距离更近，更具有一般性。

(2) 模态相互型特征学习：

如图 3 所示，利用多模态的深度置信网络提取该卷积特征和文本特征的模态相互型特征。该多模态的深度置信网络分别采用图像和文本利用自身的深度置信网络提取各自模态的特征，然后在顶层采用联合的受限玻尔兹曼机学习模态相互型特征。例如，在给定文本特征 v_t 的情况下，生成的图像，其条件概率分布如下：

$$P(h_i^{(3)} | h_i^{(2)}, h_t^{(2)}) = \sigma(W_i^{(3)} h_i^{(2)} + W_i^{(3)} h_t^{(2)} + a_i);$$

$$P(h_i^{(2)} | h_i^{(3)}) = \sigma(W_i^{(3)} h_i^{(3)} + a_i);$$

其中, W 表示权重, a 表示偏置, $\sigma(x) = 1 / (1 + e^{-x})$, 将 $h_i^{(3)}$ 作为图像的模态相互型特征 X_{mutual}^i 。同理, 可以得到文本的模态相互型特征 X_{mutual}^t 。

三、混合特征学习:

首先, 利用联合的自动编码器将模态友好型特征和模态相互型特征结合起来, 得到混合特征。例如, 对于文本通道, 其联合分布如下:

$$P(v_1, v_2) = \sum_{h_1^{(1)}, h_2^{(1)}, h^{(2)}} P(h_1^{(1)}, h_2^{(1)}, h^{(2)}) \sum_{h_1^{(1)}} P(v_1 | h_1^{(1)}) \times \sum_{h_2^{(1)}} P(v_2 | h_2^{(1)});$$

其中, v_1 表示文本的模态友好型特征 $X_{friendly}^t$, v_2 表示文本的模态相互型特征 X_{mutual}^t 。图像通道同理可得其联合分布。

其次, 利用一个带有 softmax 层的三层前向神经网络通过有监督学习, 挖掘出更有用的混合特征。具体地, 上述其他步骤所有的操作都没有用到标签信息, 即都是无标签的学习, 又叫无监督学习, 而在本步骤中, 第一次使用标签信息, 也叫有监督学习, 将两种特征进行融合的时候, 可以得到比较初级的混合特征, 其效果相对差一些, 然后再通过前向神经网络的处理能够得到更高级、有用的混合特征, 其效果有助于提升整个系统的检索结果。

四、跨模态的共享特征学习:

图 4 为单个双模态自动编码网络, 通过采用多个双模态自动编码网络挖掘模态之间的相互关系, 从而学习到跨模态的共享特征, 其概率分布如下式所示:

$$P(s) = \sum_{v_i, v_t} P(s | v_i, v_t) P(v_i) P(v_t);$$

其中, s 表示共享特征, v_i 表示图像的输入特征, v_t 表示文本的输入特征。

接着重建图像、文本特征, 获得共享特征层, 最后生成各自模态的适合跨模态检索的共享特征。其中, 在重建图像、文本特征时, 双模态自动编码网络的输入维度和输出维度一致, 共享特征层的维度是输入维度的一半, 从共享特征层中, 可以学习到各自模态的共享特征。另外, 使用 softmax 层与最后的输出层连接, 从而能够进一步优化学习到的共享特征。通过多个双模态自动编码

网络的叠加，可以不断对学习到的共享特征进行降维，这样操作可以使得最终学习到的共享特征维度比较小，从而提高检索的速度。

在图 2 至图 4 中，图像通道和文本通道分别指的是预处理过程中得到的图像的卷积特征和文本的文本特征。

本发明实施例还提供了如图 5 所示的一种跨模态检索系统，包括：

预处理单元 501，用于分别对图像和文本进行预处理，得到图像特征和文本特征；

特征提取单元 502，用于根据所述图像特征和所述文本特征，利用堆叠式的受限玻尔兹曼机提取所述图像的模态友好型特征和所述文本的模态友好型特征，及利用多模态的深度置信网络提取所述图像的模态相互型特征和所述文本的模态相互型特征；

特征结合单元 503，用于利用联合的自动编码器将所述图像的模态友好型特征和模态相互型特征结合，得到所述图像的混合特征，及将所述文本的模态友好型特征和模态相互型特征结合，得到所述文本的混合特征；

特征挖掘单元 504，用于采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系，得到所述图像和所述文本的共享特征；

特征检索单元 505，用于根据所述共享特征进行跨模态检索。

具体地，预处理单元 501 具体用于：

采用 VGGNet 模型提取所述图像 FC7 层的卷积特征；

采用 bag of words 模型提取所述文本 1000 维的文本特征。

进一步地，所述堆叠式的受限玻尔兹曼机包括三层受限玻尔兹曼机，特征提取单元 502 具体用于：

利用高斯的受限玻尔兹曼机和重复的 softmax 的受限玻尔兹曼机分别提取所述图像特征和所述文本特征；

利用简单的受限玻尔兹曼机移除所述图像特征和所述文本特征各自模态的

特性，以获得高层次的语义特征；

完成堆叠式的受限玻尔兹曼机的训练，得到所述图像的模态友好型特征 $X_{friendly}^i$ 和所述文本的模态友好型特征 $X_{friendly}^t$ 。

特征提取单元 502 还用于：

分别采用所述图像和所述文本的深度置信网络提取各自模态的特征；

在顶层采用联合的受限玻尔兹曼机进行模态相互型特征的学习，得到所述图像的模态相互型特征 X_{mutual}^i 和所述文本的模态相互型特征 X_{mutual}^t 。

特征结合单元 503 具体用于：

利用联合的自动编码器将所述图像或所述文本各自的模态友好型特征和模态相互型特征结合，得到所述图像的混合特征和所述文本的混合特征；

利用带有 softmax 层的三层前向神经网络通过有监督学习，挖掘出所述图像和所述文本更有用的混合特征。

特征挖掘单元 504 具体用于：

采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系；

根据所述相互关系进行所述图像和所述文本的跨模态的共享特征的学习，得到所述图像和所述文本的共享特征。

进一步地，特征挖掘单元 504 在提取共享特征时，双模态自动编码网络的输入维度和输出维度一致，共享特征层的维度是输入维度的一半，从共享特征层中，学习得到最后的跨模态的共享特征。

进一步地，特征挖掘单元 504 还用于将 softmax 层与最后的输出层连接，优化所述共享特征。

本发明提供的上述实施例可以应用在以图搜文、以文搜图、视频检索、图片标注的等领域，本发明实施例采用堆叠式受限玻尔兹曼机和多模态深度置信网络各自提取模态友好型和模态相互型特征，其中模态友好型特征能够使所获得的特征在统计特性方面比输入实例之间的更相似，而模态相互型特征能够得

到在原始输入实例中丢失的相互信息。接着，将所得到的这两种特征进行融合，得到混合特征，进一步通过多个双模态自动编码获得最后的共享特征。本发明实施例能有效提高跨模态检索任务的准确性，如以图搜文、以文搜图等，以及提高检索的速度。

在本申请所提供的几个实施例中，应该理解到，所揭露的方法和装置，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，所述模块的划分，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，例如多个模块或组件可以结合或者可以集成到另一个装置，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，装置或模块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模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理模块，即可以位于一个地方，或者也可以分布到多个网络模块上。可以根据实际的需要选择其中的部分或者全部模块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模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。

所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品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）执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器（ROM，Read-Only Memory）、

随机存取存储器（RAM, Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

需要说明的是，对于前述的各方法实施例，为了简便描述，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它顺序或者同时进行。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都是本发明所必须的。

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，可以参见其它实施例的相关描述。

以上为对本发明所提供的一种跨模态检索方法及系统的描述，对于本领域的技术人员，依据本发明实施例的思想，在具体实施方式及应用范围上均会有改变之处，综上，本说明书内容不应理解为对本发明的限制。

权 利 要 求 书

1、一种跨模态检索方法，其特征在于，包括：

分别对图像和文本进行预处理，得到图像特征和文本特征；

根据所述图像特征和所述文本特征，利用堆叠式的受限玻尔兹曼机提取所述图像的模态友好型特征和所述文本的模态友好型特征，及利用多模态的深度置信网络提取所述图像的模态相互型特征和所述文本的模态相互型特征；

利用联合的自动编码器将所述图像的模态友好型特征和模态相互型特征结合，得到所述图像的混合特征，及将所述文本的模态友好型特征和模态相互型特征结合，得到所述文本的混合特征；

采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系，得到所述图像和所述文本的共享特征；

根据所述共享特征进行跨模态检索。

2、如权利要求 1 所述的跨模态检索方法，其特征在于，对图像进行预处理，得到图像特征的步骤包括：

采用 VGGNet 模型提取所述图像 FC7 层的卷积特征。

3、如权利要求 1 所述的跨模态检索方法，其特征在于，对文本进行预处理，得到文本特征的步骤包括：

采用 bag of words 模型提取所述文本 1000 维的文本特征。

4、如权利要求 1 所述的跨模态检索方法，其特征在于，所述堆叠式的受限玻尔兹曼机包括三层受限玻尔兹曼机，根据所述图像特征和所述文本特征，利用堆叠式的受限玻尔兹曼机提取所述图像的模态友好型特征和所述文本的模态友好型特征包括：

利用高斯的受限玻尔兹曼机和重复的 softmax 的受限玻尔兹曼机分别提取所述图像特征和所述文本特征；

利用简单的受限玻尔兹曼机移除所述图像特征和所述文本特征各自模态的特性，以获得高层次的语义特征；

完成堆叠式的受限玻尔兹曼机的训练，得到所述图像的模态友好型特征 $X_{friendly}^i$ 和所述文本的模态友好型特征 $X_{friendly}'$ 。

5、如权利要求 1 所述的跨模态检索方法，其特征在于，根据所述图像特征和所述文本特征，利用多模态的深度置信网络提取所述图像的模态相互型特征和所述文本的模态相互型特征包括：

分别采用所述图像和所述文本的深度置信网络提取各自模态的特征；

在顶层采用联合的受限玻尔兹曼机进行模态相互型特征的学习，得到所述图像的模态相互型特征 X_{mutual}^i 和所述文本的模态相互型特征 X_{mutual}' 。

6、如权利要求 1 所述的跨模态检索方法，其特征在于，所述利用联合的自动编码器将所述图像的模态友好型特征和模态相互型特征结合，得到所述图像的混合特征，及将所述文本的模态友好型特征和模态相互型特征结合，得到所述文本的混合特征包括：

利用联合的自动编码器将所述图像或所述文本各自的模态友好型特征和模态相互型特征结合，得到所述图像的混合特征和所述文本的混合特征；

利用带有 softmax 层的三层前向神经网络通过有监督学习，挖掘出所述图像和所述文本更有用的混合特征。

7、如权利要求 1 所述的跨模态检索方法，其特征在于，所述采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系，得到所述图像和所述文本的共享特征包括：

采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系；

根据所述相互关系进行所述图像和所述文本的跨模态的共享特征的学习，得到所述图像和所述文本的共享特征。

8、如权利要求 7 所述的跨模态检索方法，其特征在于，在提取共享特征时，双模态自动编码网络的输入维度和输出维度一致，共享特征层的维度是输入维度的一半，从共享特征层中，学习得到最后的跨模态的共享特征。

9、如权利要求 8 所述的跨模态检索方法，其特征在于，将 softmax 层与最后的输出层连接，优化所述共享特征。

10、一种跨模态检索系统，其特征在于，包括：

预处理单元，用于分别对图像和文本进行预处理，得到图像特征和文本特征；

特征提取单元，用于根据所述图像特征和所述文本特征，利用堆叠式的受限玻尔兹曼机提取所述图像的模态友好型特征和所述文本的模态友好型特征，及利用多模态的深度置信网络提取所述图像的模态相互型特征和所述文本的模态相互型特征；

特征结合单元，用于利用联合的自动编码器将所述图像的模态友好型特征和模态相互型特征结合，得到所述图像的混合特征，及将所述文本的模态友好型特征和模态相互型特征结合，得到所述文本的混合特征；

特征挖掘单元，用于采用多个双模态自动编码网络挖掘所述图像的混合特征与所述文本的混合特征中模态之间的相互关系，得到所述图像和所述文本的共享特征；

特征检索单元，用于根据所述共享特征进行跨模态检索。

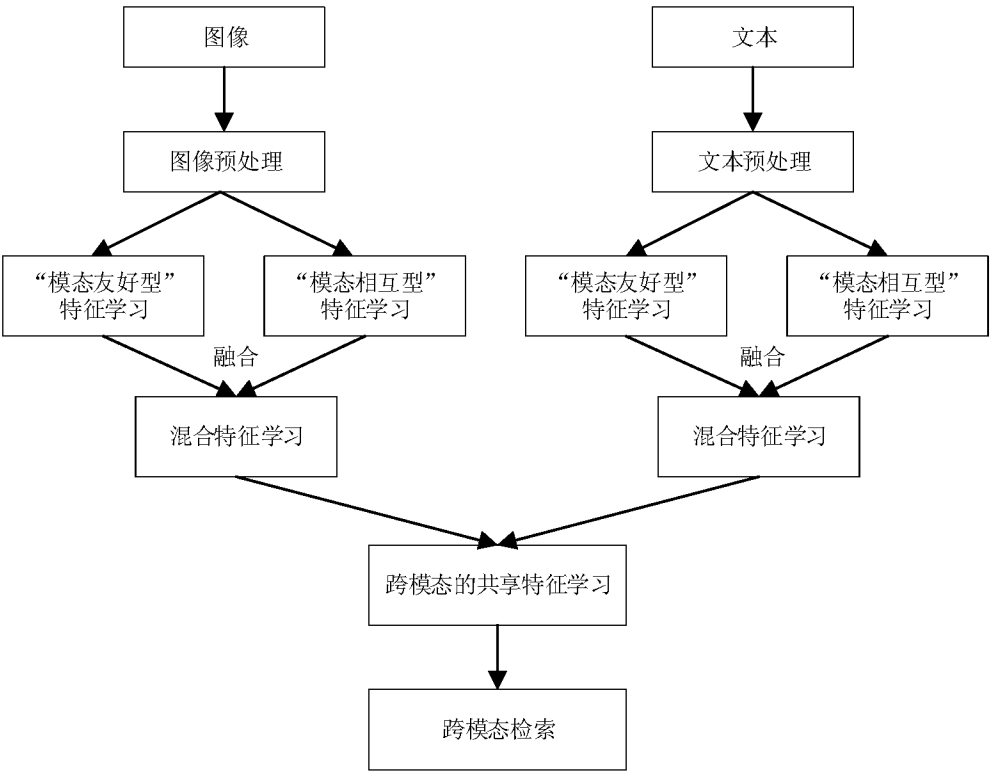


图 1

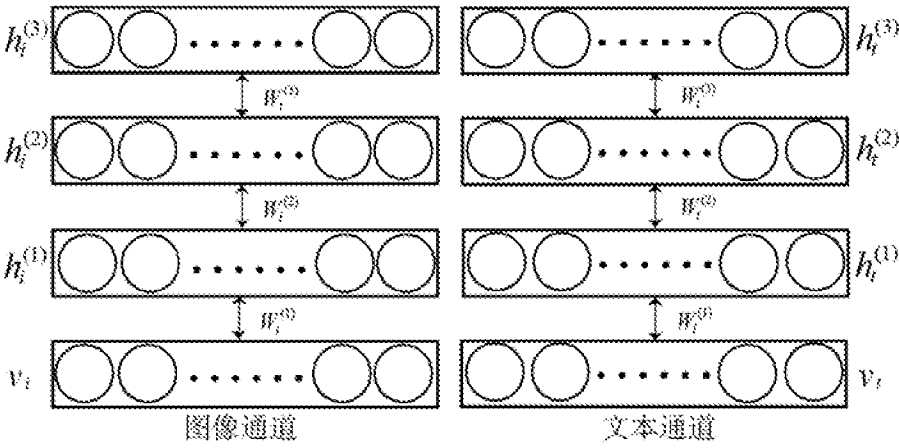


图 2

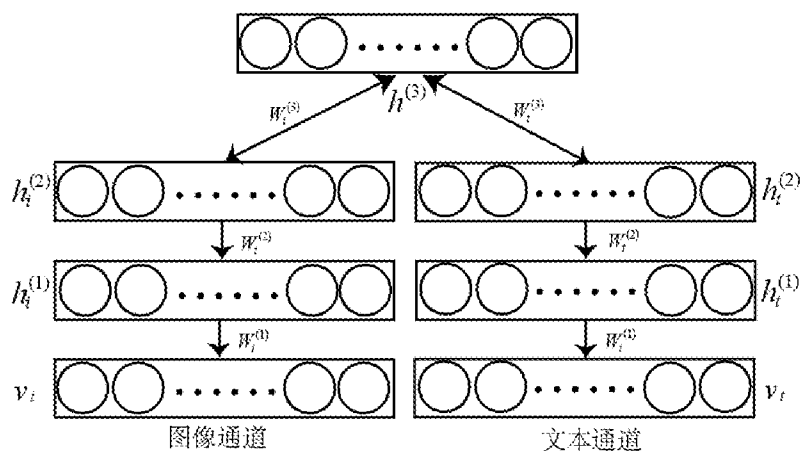


图 3

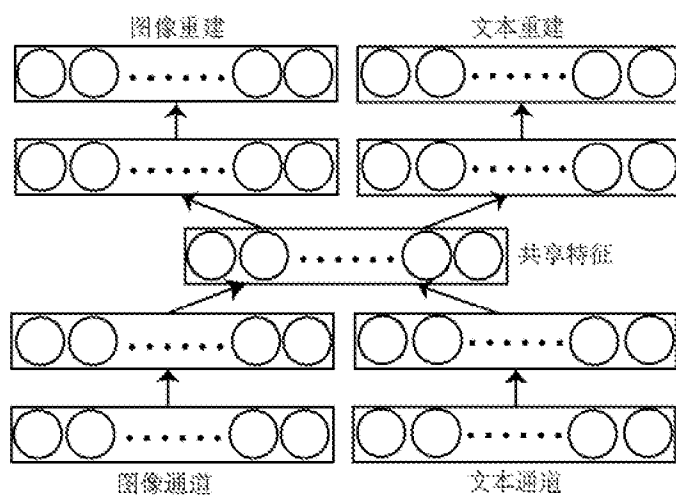


图 4

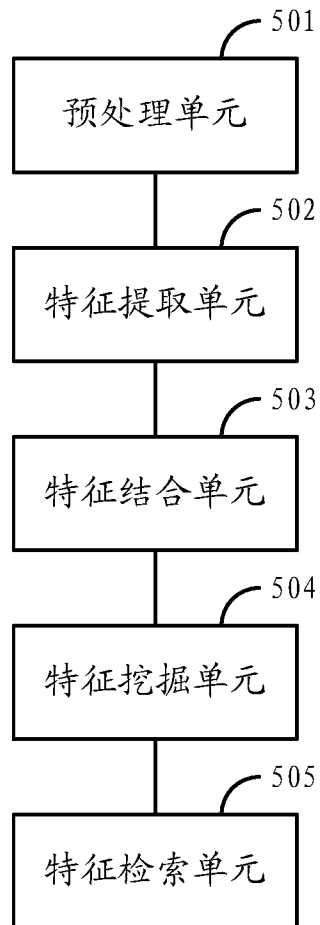


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/078657

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/53(2019.01)i; G06F 16/58(2019.01)i; G06F 16/33(2019.01)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

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: 模态, 图片, 图像, 文字, 文本, 深度置信, 深度信念, 自编码, 玻尔兹曼, 检索, 混合, 共享, modal, text, image, picture, deep belief network, DBN, Boltzmann, autoencod+, search, mix, share

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105718532 A (PEKING UNIVERSITY) 29 June 2016 (2016-06-29) description, paragraphs 23-55	1-10
X	冯方向 (FENG, Fangxiang). "基于深度学习的跨模态检索研究 (Deep Learning for Cross-Modal Reterieval)" 中国博士学位论文全文数据库 (信息科技辑) (Electronic Technology & Information Science, China Doctoral Dissertations Full-Text Database), 15 March 2016 (2016-03-15), pp. 34-80	1-10
PX	CN 109189968 A (SHENZHEN UNIVERSITY) 11 January 2019 (2019-01-11) claims 1-10	1-10
A	CN 107346328 A (PEKING UNIVERSITY) 14 November 2017 (2017-11-14) entire document	1-10
A	CN 106095893 A (PEKING UNIVERSITY SHENZHEN GRADUATE SCHOOL) 09 November 2016 (2016-11-09) entire document	1-10

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"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

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search

29 May 2019

Date of mailing of the international search report

17 June 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/078657

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	105718532	A	29 June 2016	CN	105718532	B	07 May 2019
CN	109189968	A	11 January 2019	None			
CN	107346328	A	14 November 2017	None			
CN	106095893	A	09 November 2016	CN	106095893	B	20 November 2018
				WO	2017210949	A1	14 December 2017

国际检索报告

国际申请号

PCT/CN2019/078657

A. 主题的分类

G06F 16/53(2019.01)i; G06F 16/58(2019.01)i; G06F 16/33(2019.01)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

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: 模态, 图片, 图像, 文字, 文本, 深度置信, 深度信念, 自编码, 玻尔兹曼, 检索, 混合, 共享, modal, text, image, picture, deep belief network, DBN, Boltzmann, autoencod+, search, mix, share

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 105718532 A (北京大学) 2016年 6月 29日 (2016 - 06 - 29) 说明书第23-55段	1-10
X	冯方向, “基于深度学习的跨模态检索研究” 中国博士学位论文全文数据库(信息科技辑), 2016年 3月 15日 (2016 - 03 - 15), 第34-80页	1-10
PX	CN 109189968 A (深圳大学) 2019年 1月 11日 (2019 - 01 - 11) 权利要求1-10	1-10
A	CN 107346328 A (北京大学) 2017年 11月 14日 (2017 - 11 - 14) 全文	1-10
A	CN 106095893 A (北京大学深圳研究生院) 2016年 11月 9日 (2016 - 11 - 09) 全文	1-10

☐ 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

2019年 5月 29日

国际检索报告邮寄日期

2019年 6月 17日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

武晓冬

传真号 (86-10)62019451

电话号码 86-(10)-53961385

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/078657

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	105718532	A	2016年 6月 29日	CN	105718532	B	2019年 5月 7日
CN	109189968	A	2019年 1月 11日	无			
CN	107346328	A	2017年 11月 14日	无			
CN	106095893	A	2016年 11月 9日	CN	106095893	B	2018年 11月 20日
				WO	2017210949	A1	2017年 12月 14日

表 PCT/ISA/210 (同族专利附件) (2015年1月)