(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 4 月 28 日 (28.04.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/062095 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2015/080871

(22) 国际申请日: 2015 年 6 月 5 日 (05.06.2015)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201410580006.0 2014 年 10 月 24 日 (24.10.2014) CN

(71) 申请人: **华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **复旦大学 (FUDAN UNIVERSITY)** [CN/CN]; 中国上海市杨浦区邯郸路 220 号, Shanghai 200433 (CN)。

(72) 发明人: **姜育刚 (JIANG, Yugang)**; 中国上海市浦东区张衡路 825 号计算机科学大楼 413-5 室, Shanghai 201203 (CN)。 **吴祖煊 (WU, Zuxuan)**; 中国上海市浦东区张衡路 825 号计算机科学大楼 413-5 室, Shanghai 201203 (CN)。 **薛向阳 (XUE, Xiangyang)**; 中国上海市浦东区张衡路 825 号计算机科学大楼 413-5 室, Shanghai 201203 (CN)。 **顾子晨 (GU, Zichen)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **柴振华 (CHAI, Zhenhua)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: **北京同立钧成知识产权代理有限公司 (LEADER PATENT & TRADEMARK FIRM)**; 中国北京市海淀区西直门北大街 32 号枫蓝国际 A 座 8F-6, Beijing 100082 (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,

[见续页]

(54) **Title**: VIDEO CLASSIFICATION METHOD AND APPARATUS

(54) **发明名称** : 视频分类方法和装置

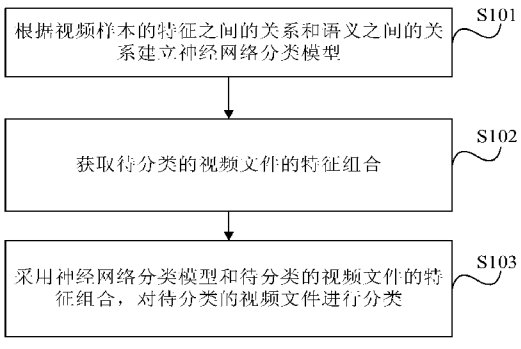


图 1 / Fig.1

S101 Establishing a neural network classification model according to a relationship among features and a relationship among semantics of a video sample S102 Acquiring a feature combination of video files to be classified S103 Using the neural network classification model and the feature combination of the video files to be classified to classify the video files to be classified

(57) **Abstract**: A video classification method and apparatus. The method comprises: establishing a neural network classification model according to a relationship among features and a relationship among semantics of a video sample (S101); acquiring a feature combination of video files to be classified (S102); and using the neural network classification model and the feature combination of the video files to be classified to classify the video files to be classified (S103). Since a neural network classification model is established according to a relationship among features and a relationship among semantics of a video sample, and the relationship among the features and the relationship among the semantics are fully considered, the accuracy of video classification can be improved.

(57) **摘要**: 一种视频分类方法和装置，通过根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型（S101）；获取待分类的视频文件的特征组合（S102）；采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类（S103）。由于神经网络分类模型是根据视频样本的特征之间的关系和语义之间的关系建立的，充分考虑了特征之间的关系和语义之间的关系，因此，可以提高视频分类的准确性。

WO 2016/062095 A1

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))。

视频分类方法和装置

本申请要求于 2014 年 10 月 24 日提交中国专利局、申请号为 201410580006.0、发明名称为“视频分类方法和装置”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。

技术领域

本发明实施例涉及计算机技术，尤其涉及一种视频分类方法和装置。

背景技术

视频分类是指利用视频的视觉信息、听觉信息以及动作信息对视频进行处理和分析，并判断和识别出视频中出现的动作和事件。视频分类应用非常广泛，例如：进行智能监控、视频数据管理等。

现有技术中，通过早期融合的技术进行视频分类，具体地，将从视频文件中提取出的不同特征或者不同特征的核矩阵线性组合起来，输入到分类器中进行分析，从而，对视频进行分类。然而，采用现有技术的方法，忽略了特征之间和语义之间的关系，因此，视频分类的准确性不高。

发明内容

本发明实施例提供一种视频分类方法和装置，以提高视频分类的准确性。

本发明实施例第一方面提供一种视频分类方法，包括：

根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型；

获取待分类的视频文件的特征组合；

采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类。

结合第一方面，在第一种可能的实现方式中，所述根据视频样本的特征

之间的关系和语义之间的关系建立神经网络分类模型，包括：

根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

根据所述神经网络分类模型融合层的权重矩阵和所述神经网络分类层的权重矩阵建立神经网络的分类模型。

结合第一方面的第一种可能的实现方式，在第二种可能的实现方式中，所述根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，包括：

通过优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

所述目标函数为：

$$\min_{W,\Omega} \zeta + \frac{\lambda_1}{2}\|W_E\|_{2,1} + \frac{\lambda_2}{2} tr(\mathrm{W}_{L-1}\,\Omega\,\mathrm{W}_{L-1}^T)$$

$$\mathrm{s.t} \quad \Omega \geq 0 \quad tr(\Omega) = 1$$

其中，ζ 表示视频样本的预测值和真实值之间的偏差，λ_1 表示预设的第一权重系数，λ_2 表示预设的第二权重系数，W_E 表示所述神经网络分类模型融合层的权重矩阵，W_E 的每一列对应一种特征，W_{L-1} 表示所述神经网络分类模型分类器层的权重矩阵，W_{L-1}^T 表示所述 W_{L-1} 的转置，$\|W_E\|_{2,1}$ 表示 W_E 的 2,1 范数，Ω 表示一个半正定的对称矩阵，用于表征语义之间的关系，Ω 初始值为单位矩阵。

结合第一方面的第二种可能的实现方式，在第三种可能的实现方式中，所述通过优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，包括：

采用近端梯度算法优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵。

结合第一方面的第三种可能的实现方式，在第四种可能的实现方式中，所述采用近端梯度算法优化目标函数，包括：

初始化所述目标函数中的所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

通过输入视频样本的特征，获取输出的预测值和实际值的偏差；

根据所述偏差调整所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，直到所述偏差小于预设阈值。

本发明实施例第二方面提供一种视频分类装置，包括：

模型建立模块，用于根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型；

特征提取模块，用于获取待分类的视频文件的特征组合；

分类模块，用于采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类。

结合第二方面，在第一种可能的实现方式中，所述模型建立模块具体用于根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；根据所述神经网络分类模型融合层的权重矩阵和所述神经网络分类层的权重矩阵建立神经网络的分类模型。

结合第二方面的第一种可能的实现方式，在第二种可能的实现方式中，所述模型建立模块具体用于通过优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

所述目标函数为：

$$\min_{W,\Omega} \zeta + \frac{\lambda_1}{2}\|W_E\|_{2,1} + \frac{\lambda_2}{2} tr(\mathrm{W}_{L-1}\,\Omega\,\mathrm{W}_{L-1}^T)$$

$$\text{s.t} \quad \Omega \geq 0 \quad tr(\Omega) = 1$$

其中，ζ 表示视频样本的预测值和真实值之间的偏差，λ_1 表示预设的第一权重系数，λ_2 表示预设的第二权重系数，W_E 表示所述神经网络分类模型融合层的权重矩阵，W_E 的每一列对应一种特征，W_{L-1} 表示所述神经网络分类模型分类器层的权重矩阵，W_{L-1}^T 表示所述 W_{L-1} 的转置，$\|W_E\|_{2,1}$ 表示 W_E 的 2,1 范数，Ω 表示一个半正定的对称矩阵，用于表征语义之间的关系，Ω 初始值为单位矩阵。

结合第二方面的第二种可能的实现方式，在第三种可能的实现方式中，所述模型建立模块具体用于采用近端梯度算法优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵。

结合第二方面的第三种可能的实现方式，在第四种可能的实现方式中，所述模型建立模块具体用于初始化所述目标函数中的所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；通过输入视频样本的特征，获取输出的预测值和实际值的偏差；根据所述偏差调整所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩

阵，直到所述偏差小于预设阈值。

本发明实施例提供的视频分类方法和装置，通过根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型；获取待分类的视频文件的特征组合；采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类。由于神经网络分类模型是根据视频样本的特征之间的关系和语义之间的关系建立的，充分考虑了特征之间的关系和语义之间的关系，因此，可以提高视频分类的准确性。

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 为本发明视频分类方法实施例一的流程示意图；

图 2 为本发明视频分类方法实施例二的流程示意图；

图 3 为本发明视频分类装置实施例一的结构示意图；

图 4 为本发明视频分类装置实施例二的结构示意图。

具体实施方式

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，所描述的实施例仅仅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做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

本发明通过结合视频样本的特征之间的关系和语义之间的关系训练神经网络分类模型，获取神经网络分类模型中各连接的最优的权重，从而，提高视频分类的准确性。

下面以具体地实施例对本发明的技术方案进行详细说明。下面这几个具体的实施例可以相互结合，对于相同或相似的概念或过程可能在某

些实施例不再赘述。

图 1 为本发明视频分类方法实施例一的流程示意图，如图 1 所示，本实施例的方法如下：

S101：根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型。

本发明实施例中所描述的神经网络是指人工神经网络，人工神经网络是一种模拟生物神经系统的计算模型，包括多层，每一层都是上一层的非线性变化，人工神经网络包括深度神经网络和传统的神经网络，深度神经网络和传统的神经网络相比可以得到从低到高不同层次的复杂特征表达，深度神经网络的结构和人体大脑皮层的多层感知结构非常类似，从而具有一定的生物理论基础，是目前研究的热点。

神经网络是一组连接的输入/输出单元，每一个输入/输出单元叫做神经元，其中，每个连接都与一个权重相关联。在神经网络的训练阶段，通过调整每个连接的相关的权重，能够较为准确的输出预测结果。

本发明实施例中所描述的视频样本是指用于训练神经网络分类模型时，所采用的视频文件。

本发明实施例借助深度神经网络的结构，根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；根据所述神经网络分类模型融合层的权重矩阵和所述神经网络分类层的权重矩阵建立神经网络的分类模型。

其中，根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵具体地，通过优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，其中，目标函数带有精心设计的正则化约束条件，从而，能够在同一个神经网络分类模型中充分考虑特征之间的关系和语义之间的关系，从而，提高视频分类的准确性。

本发明实施例带有正则化约束条件的目标函数如下所示：

$$\min_{W,\Omega} \zeta + \frac{\lambda_1}{2}\|W_E\|_{2,1} + \frac{\lambda_2}{2} tr(\mathrm{W}_{L-1}\,\Omega\,\mathrm{W}_{L-1}^{T})$$

$$\text{s.t} \quad \Omega \geq 0 \quad tr(\Omega) = 1$$

其中，ζ表示视频样本的预测值和真实值之间的偏差，λ_1表示预设的第一权重系数，λ_2表示预设的第二权重系数，W_E表示所述神经网络分类模型融合层的权重矩阵，W_E的每一列对应一种特征，W_{L-1}表示所述神经网络分类模型分类器层的权重矩阵，W_{L-1}^T表示所述W_{L-1}的转置，$\|W_E\|_{2,1}$表示W_E的2,1范数，Ω表示一个半正定的对称矩阵，用于表征语义之间的关系，Ω初始值为单位矩阵。

通常情况下，神经网络分类模型的权重矩阵一般随机进行初始化，在训练阶段，通过前向传播算法不断的对视频样本的特征（原始输入）进行非线性的映射，从而得到视频样本的预测值，视频样本的预测值和真实值之间往往有一定的偏差，通过不断的调整融合层的权重矩阵和分类器层的权重矩阵，使得针对不同的视频样本，预测值和真实值之间的偏差最小，ζ即是用来衡量整个数据集上所有视频样本的真实值和通过网络前向传播得到的预测值偏差的经验损失。

本发明为了充分利用特征之间的关系和语义之间的关系，提高视频分类的准确性，目标函数中增加了$\|W_E\|_{2,1}$项和$tr(\mathrm{W}_{L-1}\Omega\mathrm{W}_{L-1}^T)$项，其中，$W_E$表示所述神经网络分类模型融合层的权重矩阵，$W_E$的每一列对应一种特征，$\mathrm{W}_{L-1}$表示所述神经网络分类模型分类器层的权重矩阵。

最小化不同范数的含义如下所示：

特征之间的关系（融合层权重）：

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & & \vdots \\ w_{i1} & w_{i2} & \cdots & w_{im} \end{bmatrix} \xrightarrow{\text{最小化}\mathrm{I}_{21}\text{范数}} \begin{bmatrix} 0 & 0 & \cdots & 0 \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

语义之间的关系（分类器层权重）

$$\begin{bmatrix} 10\cdots 0 \\ 01\cdots \vdots \\ \vdots\vdots\ddots\vdots \\ 00\cdots 1 \end{bmatrix} \xrightarrow{\text{最小化trace范数}} \begin{bmatrix} w_{12}\cdots w_{14} \\ 1\ \cdots\ \vdots \\ \vdots\ \ddots\ \vdots \\ w_{41}\cdots\ 1 \end{bmatrix}$$

$$\Omega = \frac{(\mathrm{w}_{L-1}^T\mathrm{w}_{L-1})^{\frac{1}{2}}}{tr((\mathrm{w}_{L-1}^T\mathrm{w}_{L-1})^{\frac{1}{2}})}$$

$\|\mathbf{W}_E\|_{2,1}$即先对矩阵的每一行求2范数得到一个向量，再对这个向量求1范

数。当最小化这一范数的时候，在极少行为非0的情况下对应的目标函数会最小，从而使得矩阵行稀疏，于是保留下来的非零行即为所有不同特征间所共享的一个具有相同的模式，可反映出特征之间的一致性。

Ω 是一个半正定的对称矩阵用来刻画语义之间的关系，它最初初始化为一个单位矩阵，在神经网络分类模型的训练过程中利用分类器层的权重对其进行更新，从而得到语义间的关系，它的非对角线上的每一元素衡量的是不同语义之间的关系。

上述目标函数，可在后向传播的框架中采用基于近端梯度算法（Proximal Gradient Method，以下简称：PGM）优化目标函数。近端梯度算法是求解大规模数据时最为常用的优化算法，通常可以较快收敛，高效地求解优化问题。从而，获取神经网络分类模型中各连接的权重。通常是初始化所述目标函数中的所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；通过输入视频样本的特征，获取输出的预测值和实际值的偏差；根据所述偏差调整所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，直到所述偏差小于预设阈值。

更具体求解算法的详细步骤如下：

1：随机初始化网络权重；

2：训练过程，重复下述步骤K次；

21）不同的特征首先通过多层非线性变换抽象到同一维度;

22）不同特征在神经网络分类模型中融合在一起；

23）融合后的特征进行分类，得到前向传播的误差，即实际值和预测值之间的偏差；

24）将误差从第L层往后传递，固定Ω，利用Ω的约束使用梯度下降更新分类器层的权重矩阵 W_{L-1}，从而在更新 W_{L-1} 时考虑语义之间的关系；对融合层的权重矩阵 W_E，在2-1范数的约束下更新 W_E，从而利用特征之间的关系，

在W_E更新后，利用更新后的W_E，学习得到Ω。

结束。

通过S101的步骤，可以训练出能够准确进行视频分类的神经网络分类模型。

S102：获取待分类的视频文件的特征组合。

获取视频文件的特征组合的方式有多种，本发明对此不作限制。

通常，会获取待分类的视频文件的多种特征从而提高分类效果。一般提取改进的密集轨迹特征作为视觉特征，密集轨迹特征包括 30 维的轨迹特征，96 维的梯度直方图（histogram of gradients）的特征，108 维的光流直方图(histogram of optical flow)特征以及 192 维的运动的二元直方图(motion binary histogram)特征。这四种特征进一步被转换为了 4000 维的词袋(bag-of-words)的特征表达。还会提取梅尔倒谱系数（Mel-Frequency Cepstral Coefficients，以下简称：MFCC）以及基于频谱图（Spectrogram）的尺度不变特征（Scale Invariant Feature Transform，以下简称：SIFT）等音频特征。

S103：采用神经网络分类模型和待分类的视频文件的特征组合，对待分类的视频文件进行分类。

即，将待分类的视频文件的特征组合作为神经网络分类模型的输入，通过神经网络分类模型输出待分类的视频文件所属的分类。

采用神经网络分类模型进行视频分类处理，几乎可以实时完成，效率较高。

本实施例中，通过根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型；获取待分类的视频文件的特征组合；采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类。由于神经网络分类模型是根据视频样本的特征之间的关系和语义之间的关系建立的，充分考虑了特征之间的关系和语义之间的关系，因此，可以提高视频分类的准确性。

利用本发明的技术方案产生视频分类的结果可应用与其他视频相关技术之中，如视频摘要和视频检索等。在视频摘要中，可以将视频分成多个片段，之后利用本发明中的视频分类技术对视频进行语义分析，提取出具有意义的

视频片段作为视频摘要的结果。在视频检索中，可以利用本发明中的视频分类技术提取出视频内容的语义信息，从而对视频进行检索。

本发明还提供一种一个实施例，如图 2 所示，图 2 为本发明视频分类方法实施例二的流程示意图，如图 2 所示：

S201：从给定的视频文件中提取视觉特征和听觉特征；

S202：对提取的特征进行量化，得到特征对应的词袋模型；

S203：将每个词袋模型表征为对应的向量，对向量进行前向特征变换；

S204：对进行前向特征变换后的特征进行融合特征处理。

S205：输出视频分类结果。

采用本发明的方法，视频分类处理，几乎可以实时完成，效率较高，并且视频分类的准确性较高。

图 3 为本发明视频分类装置实施例一的结构示意图，本实施例的装置包括模型建立模块 301、特征提取模块 302 和分类模块 303，其中，模型建立模块 301 用于根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型；

特征提取模块 302 用于获取待分类的视频文件的特征组合；

分类模块 303 用于采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类。

在上述实施例中，所述模型建立模块 301 具体用于根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；根据所述神经网络分类模型融合层的权重矩阵和所述神经网络分类层的权重矩阵建立神经网络的分类模型。

在上述实施例中，所述模型建立模块 301 具体用于通过优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

所述目标函数为：

$$\min_{W,\Omega} \zeta + \frac{\lambda_1}{2}\|W_E\|_{2,1} + \frac{\lambda_2}{2} tr(\mathrm{W}_{L-1}\,\Omega\,\mathrm{W}_{L-1}^T)$$

$$\text{s.t} \quad \Omega \geq 0 \quad tr(\Omega) = 1$$

其中，ζ 表示视频样本的预测值和真实值之间的偏差，λ_1 表示预设的第一权重系数，λ_2 表示预设的第二权重系数，W_E 表示所述神经网络分类模型融合层的权重矩阵，W_E 的每一列对应一种特征，W_{L-1} 表示所述神经网络分类模

型分类器层的权重矩阵，W_{L-1}^T表示所述W_{L-1}的转置，$\|W_E\|_{2,1}$表示W_E的2,1范数，Ω表示一个半正定的对称矩阵，用于表征语义之间的关系，Ω初始值为单位矩阵。

在上述实施例中，所述模型建立模块301具体用于采用近端梯度算法优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵。

在上述实施例中，所述模型建立模块301具体用于初始化所述目标函数中的所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；通过输入视频样本的特征，获取输出的预测值和实际值的偏差；根据所述偏差调整所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，直到所述偏差小于预设阈值。

图3的装置的其它功能和操作可以参照上面图1的方法实施例的过程，为了避免重复，此处不再赘述。

图3所示实施例的装置，通过模型建立模块根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型；特征提取模块获取待分类的视频文件的特征组合；分类模块采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类。由于神经网络分类模型是根据视频样本的特征之间的关系和语义之间的关系建立的，充分考虑了特征之间的关系和语义之间的关系，因此，可以提高视频分类的准确性。

图4为本发明视频分类装置实施例二的结构示意图，如图4所示，本实施例的装置包括存储器410和处理器420，存储器410可以包括随机存储器、闪存、只读存储器、可编程只读存储器、非易失性存储器或寄存器等。处理器420可以是中央处理器（Central Processing Unit，CPU）。存储器410用于存储可执行指令。处理器420可以执行存储器410中存储的可执行指令，例如，处理器420用于根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型；获取待分类的视频文件的特征组合；采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类。

可选地，作为一个实施例，处理器420可用于根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；根据所述神经网络分类模型融合层

的权重矩阵和所述神经网络分类层的权重矩阵建立神经网络的分类模型。

可选地，作为一个实施例，处理器 420 可用于通过优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

所述目标函数为：

$$\min_{W,\Omega} \zeta + \frac{\lambda_1}{2}\|W_E\|_{2,1} + \frac{\lambda_2}{2} tr(\mathrm{W}_{L-1}\,\Omega\,\mathrm{W}_{L-1}^T)$$

$$\text{s.t} \quad \Omega \geq 0 \quad tr(\Omega) = 1$$

其中，ζ 表示视频样本的预测值和真实值之间的偏差，λ_1 表示预设的第一权重系数，λ_2 表示预设的第二权重系数，W_E 表示所述神经网络分类模型融合层的权重矩阵，W_E 的每一列对应一种特征，W_{L-1} 表示所述神经网络分类模型分类器层的权重矩阵，W_{L-1}^T 表示所述 W_{L-1} 的转置，$\|W_E\|_{2,1}$ 表示 W_E 的 2,1 范数，Ω 表示一个半正定的对称矩阵，用于表征语义之间的关系，Ω 初始值为单位矩阵。

可选地，作为一个实施例，处理器 420 可用于采用近端梯度算法优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵。

可选地，作为一个实施例，处理器420可用于初始化所述目标函数中的所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

通过输入视频样本的特征，获取输出的预测值和实际值的偏差；

根据所述偏差调整所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，直到所述偏差小于预设阈值。

图 4 的装置的其它功能和操作可以参照上面图 1 的方法实施例的过程，为了避免重复，此处不再赘述。

本领域普通技术人员可以理解：实现上述各方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成。前述的程序可以存储于一计算机可读取存储介质中。该程序在执行时，执行包括上述各方法实施例的步骤；而前述的存储介质包括：ROM、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：以上各实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述各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，或者对其中部分或者全部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范围。

5

权 利 要 求 书

1、一种视频分类方法，其特征在于，包括：

根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型；

获取待分类的视频文件的特征组合；

采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类。

2、根据权利要求 1 所述的方法，其特征在于，所述根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型，包括：

根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

根据所述神经网络分类模型融合层的权重矩阵和所述神经网络分类层的权重矩阵建立神经网络的分类模型。

3、根据权利要求 2 所述的方法，其特征在于，所述根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，包括：

通过优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

所述目标函数为：

$$\min_{W,\Omega} \zeta + \frac{\lambda_1}{2}\|W_E\|_{2,1} + \frac{\lambda_2}{2} tr(\mathrm{W}_{L-1}\,\Omega\,\mathrm{W}_{L-1}^T)$$

$$\text{s.t} \quad \Omega \geq 0 \quad tr(\Omega) = 1$$

其中，ζ 表示视频样本的预测值和真实值之间的偏差，λ_1 表示预设的第一权重系数，λ_2 表示预设的第二权重系数，W_E 表示所述神经网络分类模型融合层的权重矩阵，W_E 的每一列对应一种特征，W_{L-1} 表示所述神经网络分类模型分类器层的权重矩阵，W_{L-1}^T 表示所述 W_{L-1} 的转置，$\|W_E\|_{2,1}$ 表示 W_E 的 2,1 范数，Ω 表示一个半正定的对称矩阵，用于表征语义之间的关系，Ω 初始值为单位矩阵。

4、根据权利要求 3 所述的方法，其特征在于，所述通过优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，包括：

采用近端梯度算法优化目标函数，获取神经网络分类模型融合层的权重

矩阵和所述神经网络分类模型分类层的权重矩阵。

5、根据权利要求 4 所述的方法，其特征在于，所述采用近端梯度算法优化目标函数，包括：

初始化所述目标函数中的所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

通过输入视频样本的特征，获取输出的预测值和实际值的偏差；

根据所述偏差调整所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，直到所述偏差小于预设阈值。

6、一种视频分类装置，其特征在于，包括：

模型建立模块，用于根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型；

特征提取模块，用于获取待分类的视频文件的特征组合；

分类模块，用于采用所述神经网络分类模型和所述待分类的视频文件的特征组合，对所述待分类的视频文件进行分类。

7、根据权利要求 6 所述的装置，其特征在于，所述模型建立模块具体用于根据视频样本的特征之间的关系和语义之间的关系，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；根据所述神经网络分类模型融合层的权重矩阵和所述神经网络分类层的权重矩阵建立神经网络的分类模型。

8、根据权利要求 7 所述的装置，其特征在于，所述模型建立模块具体用于通过优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；

所述目标函数为：

$$\min_{W,\Omega} \zeta + \frac{\lambda_1}{2}\|W_E\|_{2,1} + \frac{\lambda_2}{2} tr(\mathrm{W}_{L-1}\,\Omega\,\mathrm{W}_{L-1}^T)$$

$$\text{s.t} \quad \Omega \geq 0 \quad tr(\Omega) = 1$$

其中，ζ 表示视频样本的预测值和真实值之间的偏差，λ_1 表示预设的第一权重系数，λ_2 表示预设的第二权重系数，W_E 表示所述神经网络分类模型融合层的权重矩阵，W_E 的每一列对应一种特征，W_{L-1} 表示所述神经网络分类模型分类器层的权重矩阵，W_{L-1}^T 表示所述 W_{L-1} 的转置，$\|W_E\|_{2,1}$ 表示 W_E 的 2,1 范数，Ω 表示一个半正定的对称矩阵，用于表征语义之间的关系，Ω 初始值为单位

矩阵。

9、根据权利要求 8 所述的装置，其特征在于，所述模型建立模块具体用于采用近端梯度算法优化目标函数，获取神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵。

10、根据权利要求9所述的装置，其特征在于，所述模型建立模块具体用于初始化所述目标函数中的所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵；通过输入视频样本的特征，获取输出的预测值和实际值的偏差；根据所述偏差调整所述神经网络分类模型融合层的权重矩阵和所述神经网络分类模型分类层的权重矩阵，直到所述偏差小于预设阈值。

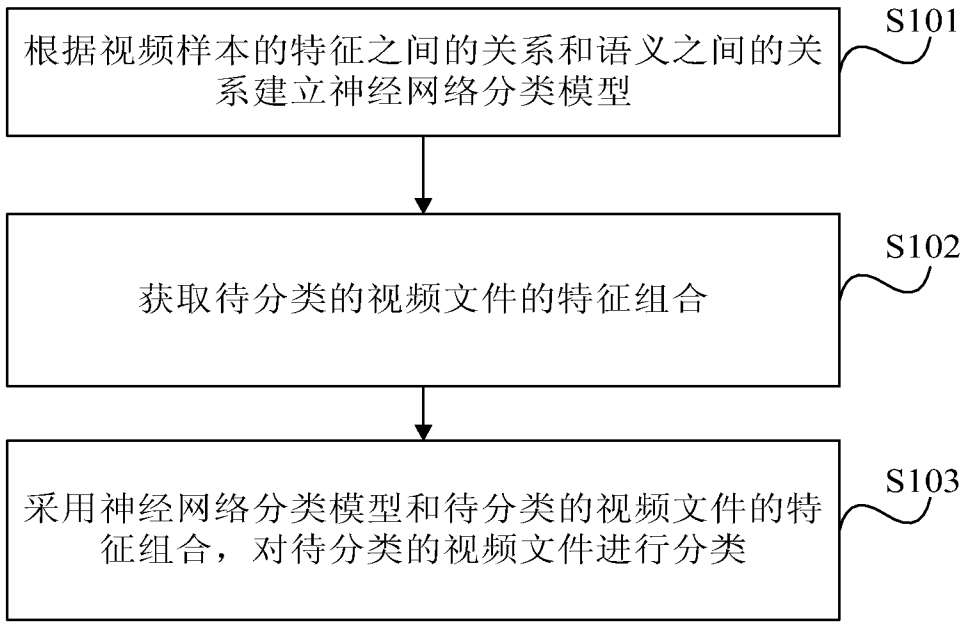

根据视频样本的特征之间的关系和语义之间的关系建立神经网络分类模型
S101
获取待分类的视频文件的特征组合
S102
采用神经网络分类模型和待分类的视频文件的特征组合，对待分类的视频文件进行分类
S103

图 1

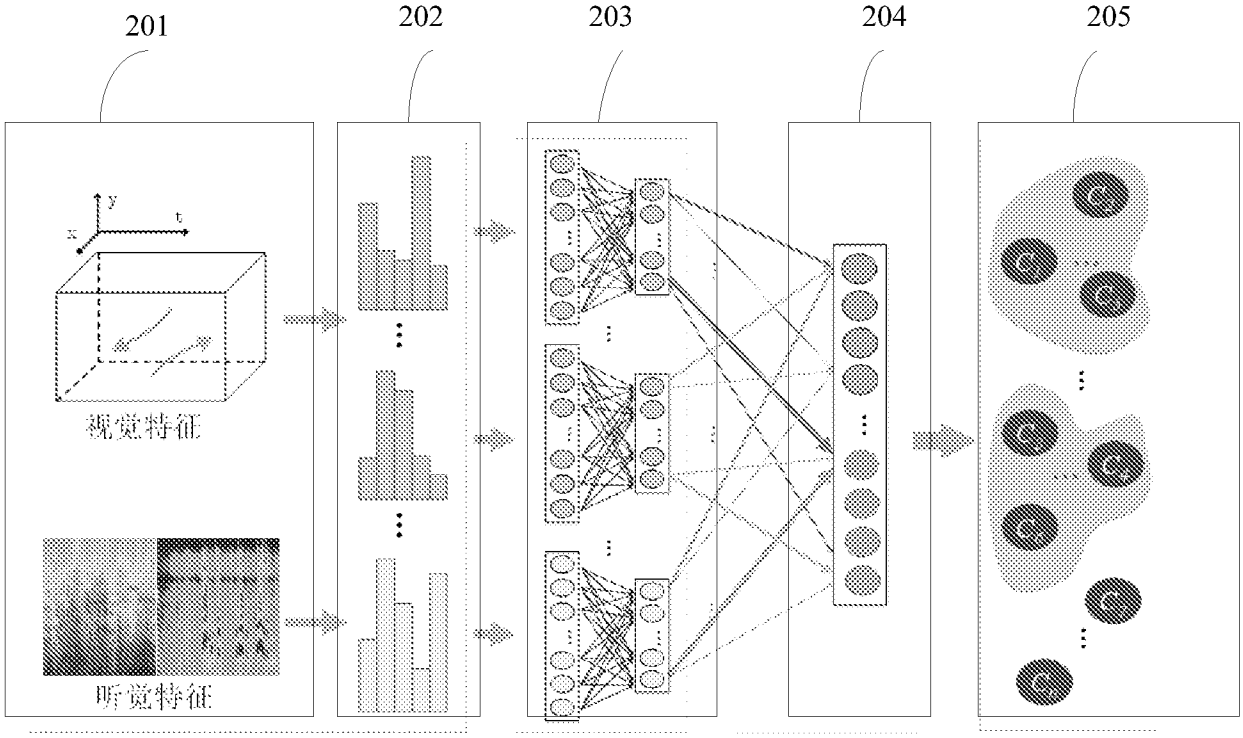

201
202
203
204
205
视觉特征
听觉特征

图 2



图 3

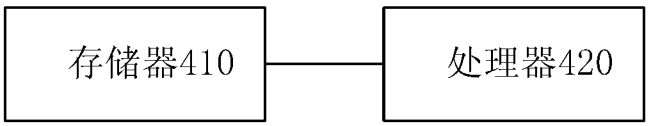


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2015/080871

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

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, CNKI, EPODOC, WPI: video, media, character+, weight, ANN, DNN, nerval, network, nerve cell, semantic, classif+, recogniz+, merge, level, power, matrix

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104331442 A (HUAWEI TECHNOLOGIES CO., LTD. et al.) 04 February 2015 (04.02.2015) claims 1-10	1-10
X	CN 102930302 A (SHANDONG UNIVERSITY) 13 February 2013 (13.02.2013) , description, paragraphs [0008]-[0037], and figures 1and 2	1, 2, 6, 7
A	CN 102930302 A (SHANDONG UNIVERSITY) 13 February 2013 (13.02.2013) the whole document	3-5, 8-10
A	CN 101593273 A (BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 02 December 2009 (02.12.2009) the whole 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	Date of mailing of the international search report
16 July 2015	28 July 2015
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 TANG, Kai Telephone No. (86-10) 62414043

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2015/080871

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 102436583 A (HARBIN ENGINEERING UNIVERSITY) 05 May 2012 (02.05.2012) the whole document	1-10
A	US 2013138436 A1 (MICROSOFT CORPORATION) 30 May 2013 (30.05.2013) the whole document	1-10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2015/080871

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104331442 A	04 February 2015	None	
CN 102930302 A	13 February 2013	None	
CN 101593273 A	02 December 2009	None	
CN 102436583 A	02 May 2012	CN 102436583 B	30 October 2013
US 2013138436 A1	30 May 2013	CN 103049792 A	17 April 2013
		HK 1183141 A0	13 December 2013

国际检索报告

国际申请号
PCT/CN2015/080871

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

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,CNKI,EPODOC,WPI:媒体，视频，特征，权重，矩阵，神经网络，神经元，语义，分类，识别，融合层，video，media，character+，ANN，DNN，nerval，network，classif+，recogniz+，merge，level，power，matrix

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
PX	CN 104331442 A (华为技术有限公司 等) 2015年 2月 4日 (2015 - 02 - 04) 权利要求1-10	1-10
X	CN 102930302 A (山东大学) 2013年 2月 13日 (2013 - 02 - 13) 说明书第[0008]-[0037]段，附图1-2	1-2，6-7
A	CN 102930302 A (山东大学) 2013年 2月 13日 (2013 - 02 - 13) 全文	3-5，8-10
A	CN 101593273 A (北京邮电大学) 2009年 12月 2日 (2009 - 12 - 02) 全文	1-10
A	CN 102436583 A (哈尔滨工程大学) 2012年 5月 2日 (2012 - 05 - 02) 全文	1-10
A	US 2013138436 A1 (MICROSOFT CORPORATION) 2013年 5月 30日 (2013 - 05 - 30) 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	国际检索报告邮寄日期
2015年 7月 16日	2015年 7月 28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) 北京市海淀区蓟门桥西土城路6号 100088 中国	康凯
传真号 (86-10)62019451	电话号码 (86-10)62414043

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2015/080871

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	104331442	A	2015年 2月 4日	无			
CN	102930302	A	2013年 2月 13日	无			
CN	101593273	A	2009年 12月 2日	无			
CN	102436583	A	2012年 5月 2日	CN	102436583	B	2013年 10月 30日
US	2013138436	A1	2013年 5月 30日	CN	103049792	A	2013年 4月 17日
				HK	1183141	A0	2013年 12月 13日