



(12) 发明专利

(10) 授权公告号 CN 110298303 B

(45) 授权公告日 2022.03.25

(21) 申请号 201910566847.9

CN 103892829 A, 2014.07.02

(22) 申请日 2019.06.27

CN 107480635 A, 2017.12.15

(65) 同一申请的已公布的文献号

CN 105640500 A, 2016.06.08

申请公布号 CN 110298303 A

CN 107348958 A, 2017.11.17

(43) 申请公布日 2019.10.01

US 2015305965 A1, 2015.10.29

(73) 专利权人 西北工业大学

US 2017193294 A1, 2017.07.06

地址 710072 陕西省西安市友谊西路127号

Olivier Le Meur 等. Visual Attention Saccadic Models Learn to Emulate Gaze Patterns From Childhood to Adulthood. 《IEEE TRANSACTIONS ON IMAGE PROCESSING》. 2017, 第26卷 (第10期),

(72) 发明人 夏辰 韩军伟 郭雷 李宽

李红霞

Chen Xia 等. Bottom-Up Visual Saliency Estimation With Deep Autoencoder-Based Sparse Reconstruction. 《IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS》. 2016, 第27卷 (第6期),

(74) 专利代理机构 西北工业大学专利中心

61204

代理人 刘新琼

李瑜 等. 精神病超高危人群自由视图眼动特征分析. 《中华精神科杂志》. 2017, 第50卷 (第1期), (续)

(51) Int. Cl.

G06V 40/19 (2022.01)

G06V 20/52 (2022.01)

G06V 10/26 (2022.01)

G06N 3/04 (2006.01)

(56) 对比文件

CN 109447096 A, 2019.03.08

CN 109101876 A, 2018.12.28

CN 105873515 A, 2016.08.17

CN 106934337 A, 2017.07.07

审查员 张露

权利要求书2页 说明书11页 附图2页

(54) 发明名称

一种基于长短时记忆网络扫视路径学习的人群识别方法

(57) 摘要

本发明涉及一种基于长短时记忆网络扫视路径学习的人群识别方法,采集不同人群组别的眼动数据构建训练样本集,利用LSTM网络来有监督地学习出不同类别扫视模式与人群组别标记之间的映射关系,根据训练后的网络来对待评估观察者的扫视序列进行得分判定,实现基于扫视路径学习的人群识别,与其他手段结合帮助提高人群识别的准确性,促进相关应用的发展。

CN 110298303 B



[转续页]

[接上页]

(56) 对比文件

Antoine Coutrot 等. Scanpath modeling and classification with hidden Markov models.《Behav Res》.2017,

高帅. 基于机器学习的扫视路径估计方法研究.《中国优秀硕士学位论文全文数据库信息科技辑》.2019,第2019年卷(第2期),

1. 一种基于长短时记忆网络扫视路径学习的人群识别方法,其特征在于步骤如下:

步骤1: 构建测试图像库

首先根据具体人群识别问题来选择图像构建测试图像库;

步骤2: 采集不同人群组别扫视序列

组织眼动实验分别采集不同组别观察者在测试图像库上的眼动数据,获取其中扫视路径序列;

步骤3: 提取扫视序列特征

提取观察者在测试图像上扫视序列对应的特征,为下一步训练LSTM网络做准备;

3a) 图像过分割

采用简单线性迭代聚类算法SLIC将图像分成多个过分割区域,其中每个过分割区域由相邻且具有相似特性的像素点组成;

3b) 提取扫视序列特征

利用语义哈希算法来对采集的扫视路径进行特征提取:对于当前扫视路径上的每一个注视点,找到其所在图像过分割区域,提取该过分割区域中心大小为 $d \times d \times 3$ 的原始图像块,将其对应的向量输入到语义哈希编码特征提取网络中,提取中心层 N_c 维二值特征;即对于长度为 m 的扫视路径序列,得到其所对应的 $N_c \times m$ 扫视路径特征;

步骤4: 训练LSTM网络

首先构建训练LSTM网络需要用到的训练样本集:训练样本集由扫视路径特征与对应的人群组别标记组成,LSTM网络的学习目标是建立由扫视路径特征到人群组别标记之间的映射关系;其次是利用训练样本集训练LSTM网络:依次将一个扫视路径上连续注视点对应的语义哈希编码特征作为网络输入,迭代得到最顶层回归层的人群组别预测得分,再使用基于时间的反向传播算法根据预测得分与对应标记的预测误差迭代调整网络参数,得到基于扫视路径的人群识别模型;

步骤5: 测试待评估观察者扫视序列

利用眼动仪记录当前待评估观察者在测试图像上的扫视路径;采用同训练过程同样的方式提取各个扫视路径特征,即对于每一个扫视路径,找到其上每一个注视点所在的过分割区域,并提取该过分割区域中心对应的语义哈希编码,得到编码后的扫视序列,将该序列输入到训练后的人群识别模型中,得到该序列对应的回归层得分;

步骤6: 判断待评估观察者人群组别

根据训练集上观察者的平均得分来依次计算两个相邻组别之间的平均得分划分阈值;对于每两个相邻组别,首先利用训练后LSTM网络预测该两个相邻组别下各个观察者在所有测试图像上的平均得分,再利用最大类间方差法来对这两个相邻组别观察者的平均得分进行阈值划分;以此类推,对于 n 个组别的人群识别问题,得到 $n-1$ 个相邻组别划分阈值;最后计算待评估观察者扫视路径序列的平均得分,并判断该平均得分所在组别区间,完成人群识别判断。

2. 根据权利要求1所述的一种基于长短时记忆网络扫视路径学习的人群识别方法,其特征在于所述的步骤2中记载的眼动实验:采用善睐眼动平台2.0下的便携式眼动仪,眼动实验屏幕分辨率为 1920×1080 ,考虑到儿童视觉角问题,测试图像大小设置为 800×600 并放置于屏幕中心,其余部分用灰色模板填充;观察者坐在距离屏幕60cm-75cm左右的位置;

每个测试图像呈现3s,不同场景间隔为0.5s,用一幅灰色模板来实现。

一种基于长短时记忆网络扫视路径学习的人群识别方法

技术领域

[0001] 本发明属于图像处理领域,涉及一种人群识别方法,即通过对不同人群的扫视路径进行有监督学习,来建立扫视模式与人群组别之间的映射关系,实现基于扫视路径学习的人群识别。

背景技术

[0002] 每秒钟大约有 10^8 阶比特数据流进入人眼,远超过大脑能够全部处理和吸收的信息量。面对如此大量与复杂的视觉输入,人类视觉系统仍然能够有效地获取场景重要内容,来增加对外部世界的了解。这主要是因为选择性视觉注意机制的存在。视觉注意是人类视觉感知的一项重要特征,是大部分视觉任务的关键因素,其本质在于快速提取场景重要区域,并对其分配主要计算资源。对人类的视觉注意过程进行研究不仅有助于探索人类视觉加工的内在机理,也能够为大数据时代下资源的合理分配与信息的高效处理提供解决方案,来帮助解决医疗诊断、社会安全等多个领域中的应用问题。

[0003] 在目前关于视觉注意的研究中,研究者们主要关注于人们在观察视觉输入时表现出的共性眼动特征。Laurent Itti等人在文章“A Model of Saliency-Based Visual Attention for Rapid Scene Analysis”,IEEE Transactions on Pattern Analysis and Machine Intelligence,vol.20,no.11,pp.1254-1259,1998中通过提出模拟视网膜神经元细胞感受野的中心-周围对比度机制来估计场景不同区域被人们关注的概率。基于深度学习强大的表示与学习能力,近期视觉注意的研究则倾向于探索由原始图像到被关注概率图的直接映射关系。Wenguan Wang和Jianbing Shen在文章“Deep visual attention prediction,”IEEE Transactions on Image Processing,vol.27,no.5,pp.2368-2378,2018中通过设计改进卷积神经网络来预测像素级的被关注概率图,相对于之前工作进一步提升对于图像关注区域估计的有效性与稳定性。尽管对图像关注区域进行估计能够揭示出人类关注区域位置等重要信息,却无法对视觉注意中注视点跳转及整个扫视路径的生成过程进行描述。为了更加全面地理解与阐明视觉注意,研究者们开始关注到扫视路径的预测问题,即估计连续多个时刻的注视点位置,以生成扫视路径的方式来建模人类视觉系统对于当前场景的动态解析过程。Ming Jiang等人在文章“Learning to predict sequences of human visual fixations,”IEEE Transactions on Neural Networks and Learning Systems,vol.27,no.6,pp.1241-1252,2016中提出利用马尔可夫决策过程来建模扫视过程的时变性,在不同的扫视阶段学习不同的参数,并通过迭代地寻找新动作与基于该动作提取对应的特征组合来生成扫视路径。

[0004] 相对于场景关注区域估计与扫视路径预测等共性机制的研究,目前仅有少数工作关注到不同群体在观察同组图像时呈现出的视觉注意差异性。Johannes Hewig等人在文章“Gender Differences for Specific Body Regions When Looking at Men and Women”,Journal of Nonverbal Behavior,vol.32,pp.67-78,2008中首次聚焦于不同性别人群的眼动差异性分析。他们分别为不同性别组提供30幅人像图像,并对人像进行手动区域划分。

最后通过统计不同组别下各个划分区域的注视点总数及关注时间等特征来发掘不同人群组别的视觉注意差异性。Olivier Le Meur在文章“Visual Attention Saccadic Models Learn to Emulate Gaze Patterns From Childhood to Adulthood”, IEEE Transactions on Image Processing, vol.26, no.10, pp.4777-4789, 2017中通过统计来自5个年龄段组别101个观察者在30幅儿童读物图像上的眼动数据来发现不同组别下扫视幅度和方向联合概率分布的差异性,并提出针对不同年龄段组别建立不同的扫视路径预测模型。

[0005] 虽然在传统聚焦于视觉注意共性机制研究的基础上出现了针对不同人群视觉注意差异性分析的工作,目前的研究主要是利用统计的手段来获取不同人群在某些特定眼动指标(注视点数、关注时间、扫视幅度和方向联合概率分布)下的差异性。一方面,大多数指标主要关注于单个注视点特征,即使是扫视幅度和方向联合概率分布也只针对相邻的两个注视点进行计算,没有考虑前期扫视路径序列对于当前注视点选择的影响,缺少对于整个动态扫视过程的学习与分析。完整的扫视路径不仅能够揭示多个关注区域的位置信息,还包含关注区域连接关系以及序列动态信息,具有重要的学习与挖掘价值。另一方面,之前研究的主要目标在于进行不同人群视觉注意差异性的分析,缺少直接基于扫视路径学习的人群识别模型。

发明内容

[0006] 要解决的技术问题

[0007] 为了避免现有技术的不足之处,本发明提出一种基于长短时记忆网络扫视路径学习的人群识别方法。利用长短时记忆(Long Short-Term Memory, LSTM)网络的序列数据学习能力来探索扫视路径数据内在本质与相互关系,实现基于扫视路径学习的人群识别。在学习模型的基础上,通过待评估观察者的扫视路径表现给出人群组别判定的结果。

[0008] 技术方案

[0009] 一种基于长短时记忆网络扫视路径学习的人群识别方法,其特征在于步骤如下:

[0010] 步骤1:构建测试图像库

[0011] 首先根据具体人群识别问题来选择图像构建测试图像库;

[0012] 步骤2:采集不同人群组别扫视序列

[0013] 组织眼动实验分别采集不同组别观察者在测试图像库上的眼动数据,获取其中扫视路径序列;

[0014] 步骤3:提取扫视序列特征

[0015] 提取观察者在测试图像上扫视序列对应的特征,为下一步训练LSTM网络做准备;

[0016] 3a) 图像过分割

[0017] 采用简单线性迭代聚类算法SLIC将图像分成多个过分割区域,其中每个过分割区域由相邻且具有相似特性的像素点组成;

[0018] 3b) 提取扫视序列特征

[0019] 利用语义哈希算法来对采集的扫视路径进行特征提取:对于当前扫视路径上的每一个注视点,找到其所在图像过分割区域,提取该过分割区域中心大小为 $d \times d \times 3$ 的原始图像块,将其对应的向量输入到语义哈希编码特征提取网络中,提取中心层 N_c 维二值特征;即对于长度为 m 的扫视路径序列,得到其所对应的 $N_c \times m$ 扫视路径特征;

[0020] 步骤4:训练LSTM网络

[0021] 首先构建训练LSTM网络需要用到的训练样本集:训练样本集由扫视路径特征与对应的人群组别标记组成,LSTM网络的学习目标是建立由扫视路径特征到人群组别标记之间的映射关系;其次是利用训练样本集训练LSTM网络:依次将一个扫视路径上连续注视点对应的语义哈希编码特征作为网络输入,迭代得到最顶层回归层的人群组别预测得分,再使用基于时间的反向传播算法根据预测得分与对应标记的预测误差迭代调整网络参数,得到基于扫视路径的人群识别模型;

[0022] 步骤5:测试待评估观察者扫视序列

[0023] 利用眼动仪记录当前待评估观察者在测试图像上的扫视路径;采用同训练过程同样的方式提取各个扫视路径特征,即对于每一个扫视路径,找到其上每一个注视点所在的过分割区域,并提取该过分割中心对应的语义哈希编码,得到编码后的扫视序列,将该序列输入到训练后的人群识别模型中,得到该序列对应的回归层得分;

[0024] 步骤6:判断待评估观察者人群组别

[0025] 根据训练集上观察者的平均得分来依次计算两个相邻组别之间的平均得分划分阈值;对于每两个相邻组别,首先利用训练后LSTM网络预测该两个组别下各个观察者在所有测试图像上的平均得分,再利用最大类间方差法来对这两个相邻组别观察者的平均得分进行阈值划分;以此类推,对于n个组别的人群识别问题,得到n-1个相邻组别划分阈值;最后计算待评估观察者扫视路径序列的平均得分,并判断该平均得分所在组别区间,完成人群识别判断。

[0026] 所述的步骤2中记载的眼动实验:采用善睐眼动平台2.0下的便携式眼动仪,眼动实验屏幕分辨率为 1920×1080 ,考虑到儿童视觉角问题,测试图像大小设置为 800×600 并放置于屏幕中心,其余部分用灰色模板填充;观察者坐在距离屏幕65cm左右的位置;每个测试图像呈现3s,不同场景间隔为0.5s,用一幅灰色模板来实现。

[0027] 有益效果

[0028] 本发明提出的一种基于长短时记忆网络扫视路径学习的人群识别方法,具有如下优点:

[0029] 1) 本发明从连续扫视路径序列的角度入手,相对于之前注视点统计的研究更加完整地利用了眼动数据中蕴含的动态关联信息。同时结合在序列表示与处理上具有突出能力的LSTM网络来对扫视路径序列间的内在表示进行挖掘,更好地探索动态扫视路径序列与人群标记之间的本质关系。

[0030] 2) 本发明根据扫视路径上各个注视点位置对应的语义哈希编码来对扫视数据进行表示,实现由底层图像块向量到高层语义编码向量的映射,获取扫视路径序列的抽象性描述。相对于底层像素级模型具有更强的抽象与表示能力,能够更好地发掘不同人群组别视序列中呈现出的语义关联差异性,得到与实际人群组别标记更一致的识别结果。

[0031] 3) 本发明从学习的角度来直接建立从特征提取扫视路径序列到人群组别标记的预测模型,相对于之前进行眼动属性差异性分析的模型更进一步地研究了视觉注意属性与人群组别之间的关系,实现了人群组别的分类与识别。

[0032] 本发明可通过测试图像的改变扩展到年龄段识别、性别识别、自闭症识别等多种类型人群识别与分类应用中,也可以作为特征并结合其他机器学习方法用于目标检测、识

别等应用中。

附图说明

- [0033] 图1为本发明的实现总流程图；
- [0034] 图2为本发明中语义哈希编码特征提取网络结构图；
- [0035] 图3为本发明中扫视路径特征提取示意图；
- [0036] 图4为本发明中LSTM网络示意图；
- [0037] 图5为本发明人群识别平均得分阈值划分示意图。

具体实施方式

[0038] 现结合实施例、附图对本发明作进一步描述：

[0039] 本发明采集不同人群组别的眼动数据构建训练样本集，利用LSTM网络来有监督地学习出不同组别扫视模式与人群组别标记之间的映射关系，根据训练后的LSTM网络来对待评估观察者的扫视路径进行得分判定，实现基于扫视路径学习的人群识别，与其他手段结合帮助提高人群识别的准确性，促进相关应用的发展。其实现步骤包括如下：

[0040] (1) 构建测试图像库

[0041] 首先根据具体人群识别问题来选择测试图像构建测试图像库。

[0042] (2) 采集不同人群组别扫视路径序列

[0043] 组织眼动实验，分别采集不同组别观察者在所有测试图像上的眼动数据，获取测试图像上对应的扫视路径序列。眼动记录设备可采用眼动仪。眼动实验屏幕分辨率为 1920×1080 ，考虑到儿童视觉角问题，测试图像大小可设置为 800×600 并放置于屏幕中心，其余部分用灰色模板填充。观察者坐在距离屏幕60cm-75cm左右的位置。每个测试图像呈现3s，不同测试图像间隔为0.5s，用一幅灰色模板来实现。

[0044] (3) 提取扫视路径特征

[0045] 提取观察者在测试图像上扫视路径对应的特征，为下一步训练LSTM网络做准备。

[0046] (3a) 图像过分割

[0047] 采用简单线性迭代聚类(simple linear iterative clustering, SLIC)算法来将图像分割为多个过分割区域，其中每个过分割区域由相邻且具有相似特性的像素点组成。接下来再以这些过分割区域代替像素点作为图像基本处理单元，来减小图像噪声等对最终识别结果的影响。

[0048] (3b) 提取扫视路径特征

[0049] 利用语义哈希(Semantic Hashing)算法来对采集的扫视路径进行特征提取。将扫视路径上各个注视点所在过分割区域对应的原始图像数据变换到具有更强辨别能力的语义空间，利用进而提升基于LSTM网络的识别能力。语义哈希算法采用深层自动编码器作为学习网络。该网络由共用中心层且结构对称的编码器与解码器两部分组成。中心层为二值神经元，其余所有神经元均为逻辑神经元。一个重要特点是编码器神经元个数逐层减少，在编码器的顶层(中心层)神经元数量通常远小于输入层神经元数量。另一方面，又需要通过解码器来在网络输出端重构出输入，这就迫使网络中心层能够发掘出高维输入数据中的内在结构与关系，获得从底层图像信息到语义特征的映射关系。具体训练中，利用从OSIE数据

库700幅图像中所有过分割块中心提取的大小为 $d \times d \times 3$ 的图像块来进行语义哈希编码特征提取网络训练,并将网络中心层 N_c 维输出作为注视点的特征。

[0050] 在训练语义哈希编码特征提取网络的基础上,下面需要对扫视路径进行特征提取。对于当前扫视路径上的每一个注视点,找到其所在图像过分割区域,提取该过分割区域中心大小为 $d \times d \times 3$ 的原始图像块,将其对应的向量输入到语义哈希编码特征提取网络中,提取中心层 N_c 维二值特征。即对于长度为 m 的扫视路径序列,得到其所对应的 $N_c \times m$ 扫视路径特征。

[0051] (4) 训练LSTM网络

[0052] 首先构建训练LSTM网络需要用到的训练样本集。训练样本集由扫视路径特征与对应的人群组别标记组成,LSTM网络的学习目标是建立由扫视路径特征到人群组别标记之间的映射关系。其次是利用训练样本集训练LSTM网络。依次将一个扫视路径上连续注视点对应的语义哈希编码特征作为网络输入,迭代得到最顶层回归层的人群组别预测得分,再使用基于时间的反向传播算法(Back Propagation Through Time,BPTT)算法根据预测得分与对应标记的预测误差迭代调整网络参数,得到基于扫视路径的人群识别模型。

[0053] (5) 测试待评估观察者扫视路径序列

[0054] 利用眼动仪记录当前待评估观察者在所有测试图像上的扫视路径。采用同训练过程同样的方式提取各个扫视路径特征,即对于每一个扫视路径,找到其上每一个注视点所在的过分割区域,并提取该过分割区域中心对应的语义哈希编码,得到扫视路径对应的特征序列,将该序列输入到训练后的LSTM模型中,得到该序列对应的回归层得分。

[0055] (6) 判断待评估观察者人群组别

[0056] 首先根据训练集上观察者的平均得分来依次计算两个相邻组别之间的平均得分划分阈值。对于每两个相邻组别,首先利用训练后LSTM网络预测该两个组别下各个观察者在所有测试图像上的平均得分,再利用最大类间方差法来对这两个相邻组别观察者的平均得分进行阈值划分。以此类推,对于 n 个组别的人群识别问题,得到 $n-1$ 个相邻组别划分阈值。最后计算待评估观察者扫视路径序列的平均得分,并判断该平均得分所在组别区间,完成人群识别判断。

[0057] 实施例1:基于扫视路径学习的性别识别

[0058] 参照图1,本发明的具体实现步骤如下:

[0059] 步骤1,构建测试图像库

[0060] 本发明共选择两类图像来记录眼动。第一类图像来自30幅人像图像;第二类图像来自30幅具有不同性别购物偏好物体的合成图像,例如连衣裙与男生运动鞋的合成图像、化妆品与数码产品的合成图像等等。

[0061] 步骤2,采集不同人群组别扫视路径序列

[0062] 组织眼动实验,分别采集不同性别组别观察者在所有测试图像上的眼动数据,获取测试图像上对应的扫视路径序列。对于性别识别问题,共采集年龄在20-35周岁20名男性观察者及20名女性观察者的眼动数据。眼动记录设备采用善睐眼动平台2.0下的便携式眼动仪。眼动实验屏幕分辨率为 1920×1080 ,观察者坐在距离屏幕60cm-75cm左右的位置。每个测试图像呈现3s,不同测试图像间隔为0.5s,用一幅灰色模板来实现。

[0063] 步骤3,提取扫视路径序列特征

[0064] 这一步骤的目标在于构建扫视路径特征提取网络,提取观察者在测试图像上扫视路径序列对应的特征,为下一步训练LSTM网络做准备。

[0065] (3a) 图像过分割

[0066] 首先采用简单线性迭代聚类(simple linear iterative clustering,SLIC)算法来将图像分割为多个过分割区域,其中每个过分割区域由相邻且具有相似特性的像素点组成。接下来再以这些过分割区域代替像素点作为图像基本处理单元,来减小图像噪声等对最终识别结果的影响。

[0067] (3b) 提取扫视路径特征

[0068] 接下来任务是利用语义哈希(Semantic Hashing)算法来对采集的扫视路径序列进行特征提取,将扫视路径序列上各个注视点所在过分割区域对应的原始图像数据变换到具有更强辨别能力的语义空间,进而提升基于LSTM网络的识别能力。参照图2,语义哈希算法采用深层自动编码器作为学习网络。该网络由共用中心层且结构对称的编码器与解码器两部分组成。中心层为二值神经元,其余所有神经元均为逻辑神经元。编码器由5层网络结构组成,相邻两层网络是全连接的关系, N_0 为输入层的神经元数量,其大小由输入图像块对应的向量维数 $d \times d \times 3$ 决定。该网络一个重要特点是编码器神经元个数逐层减少, $N_i \geq 2N_{i+1}, i=0,1,2, N_3 \geq 2N_c$,中心层神经元数量 N_c 通常远小于输入层神经元数量 N_0 。另一方面,又需要通过解码器来在网络输出端重构出输入,这就迫使网络中心层能够发掘出高维输入数据中的内在结构与关系,获得从底层图像信息到语义特征的映射关系。

[0069] 为了获取网络参数,利用从OSIE数据库700幅图像中所有过分割块中心提取的340,000个大小为 $d \times d \times 3$ 的图像块来进行网络训练。对于每个样本点 x ,将其对应的长度为 d 的原始图像块 $p(x)$ 转化为 $d \times d \times 3$ 维列向量 $f(x)$ 作为网络输入,再将输入向量 $f(x)$ 作为对应的网络标记输出,来构建语义哈希编码特征提取网络训练样本。

[0070] 语义哈希编码特征提取网络的训练过程分为预训练与反向传播微调参数两个阶段,首先是预训练的过程,来给出网络一组初始参数 $H(\cdot, \theta)$ 。首先将编码器看作一组受限玻尔兹曼机(Restricted Boltzmann Machine,RBM)的组合,训练当前RBM得到隐层单元的激活概率将被作为下一个RBM的可见层单元继续迭代训练,来逐层初始化编码器,解码器的初值是编码器的转置。这样能将整个学习任务分解成多个简单的任务,并消除了在定向生成式模型中出现的推理问题。

[0071] 接下来是反向传播阶段,在初始网络 $H(\cdot, \theta)$ 下,计算网络输出 $H(f(x), \theta)$,以及输出向量 $H(f(x), \theta)$ 和标记输出 $f(x)$ 之间的交叉熵误差 $CE(x)$:

$$[0072] \quad CE(x) = -\sum_i f(x)_i \log(H(f(x), \theta)_i) - \sum_i (1 - f(x)_i) \log(1 - H(f(x), \theta)_i),$$

[0073] 其中, $f(x)_i$ 为标记输出 $f(x)$ 的第 i 维元素, $H(f(x), \theta)_i$ 为实际输出 $H(f(x), \theta)$ 的第 i 维元素。根据所有训练样本交叉熵误差的平均值来利用反向传播算法自顶向下微调整整个网络参数 $H(\cdot, \theta)$,提高模型的鉴别能力。

[0074] 在训练语义哈希编码特征提取网络的基础上,下面需要对扫视路径序列进行特征提取。参照图3,对于长度为 m 扫视路径序列 $(q_1, q_2, \dots, q_t, \dots, q_m)$ 上 t 时刻注视点 q_t ,找到其在图像过分割区域,提取该区域中心大小为 $d \times d \times 3$ 的原始图像块,将其对应的向量 $f(q_t)$ 输入到语义哈希编码特征提取网络中,提取中心层 N_c 维二值特征 $H(q_t)$ 。将长度为 m 的扫视路

径序列,转化为对应的 $N_c \times m$ 特征。

[0075] 步骤4,训练LSTM网络

[0076] 首先在上一步骤的基础上构建训练LSTM网络需要用到的训练样本集。对每个来自男性组别的扫视路径序列提取特征,并给出对应的标记1;对每个来自女性组别的扫视路径序列提取特征,并给出对应的标记2。

[0077] 参照图4,构建用于性别识别的LSTM网络。该网络由两部分组成,迭代循环部分以及顶层回归层。循环部分包含两个状态:单元状态 c 和隐含层 h 。在 t 时刻,LSTM网络将受到三个输入的共同作用:当前时刻LSTM网络的输入值 x_t 、上一时刻LSTM网络的输出值 h_{t-1} 、以及上一时刻的单元状态 c_{t-1} 。同样在 t 时刻,LSTM网络有两部分输出:当前时刻LSTM网络的输出值 h_t ,以及当前时刻的单元状态 c_t 。所有输入输出都是向量的形式。

[0078] LSTM网络用两个门来控制单元状态 c 的内容,一个是遗忘门(forget gate),它决定了上一时刻的单元状态有多少保留到当前时刻;另一个是输入门(input gate),它决定了当前时刻网络的输入有多少保存到单元状态。LSTM网络用输出门(output gate)来控制单元状态有多少输出到LSTM的当前输出值。

[0079] 首先是遗忘门输出的计算:

$$[0080] \quad f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f),$$

[0081] 其中, W_f 与 b_f 分别为遗忘门的权重与偏置项, $\sigma(\cdot)$ 为sigmoid函数。

[0082] 其次是输入门输出计算:

$$[0083] \quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i),$$

[0084] 其中, W_i 与 b_i 分别为输入门的权重与偏置项。

[0085] 再次,计算用于描述输入的状态 \bar{c}_t :

$$[0086] \quad \bar{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c),$$

[0087] 其中, W_c 与 b_c 分别为权重与偏置项。 $\tanh(\cdot)$ 为tanh函数。

[0088] 接下来,则是计算当前 t 时刻的单元状态 c_t :

$$[0089] \quad c_t = f_t \odot c_{t-1} + i_t \odot \bar{c}_t,$$

[0090] 其中 \odot 表示按元素乘操作,通过计算 c_t ,可以将当前短期记忆 \bar{c}_t 与长期记忆 c_{t-1} 的影响组合起来。由于遗忘门的设定,单元状态可以保存很久之前的信息,由于输入门的设定,可以避免将当前无关紧要的内容存入记忆。

[0091] 下面,是输出门的计算:

$$[0092] \quad o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o),$$

[0093] 其中, W_o 与 b_o 分别为输出门的权重与偏置项。在计算输出门的基础上,最后是当前时刻下最终输出的计算:

$$[0094] \quad h_t = o_t \odot \tanh(c_t).$$

[0095] 假设扫视路径序列长度为 m ,在前向传播到最后一个注视点时,将最终输出 h_m 与神经元个数为50的全连接层 $FF(\cdot)$ 连接,得到全连接层输出 y_m :

$$[0096] \quad y_m = FF(h_m)$$

[0097] 再将 y_m 与包含一个神经元的回归层全连接起来,得到该扫视路径序列对应的得分score。

[0098] $score = regression(y_m)$

[0099] 在将训练样本集扫视路径序列特征输入到网络中进行前向传播的基础上,使用BPTT算法来更新网络参数。具体做法是将连续注视点对应的 N_c 维二值语义哈希编码作为网络输入,得到最顶层回归层的预测值。再计算预测值与对应人群组别标记之间的预测误差,并根据该误差计算每个权重的梯度,沿时间反向传递误差项,调整网络的参数,最终得到基于扫视路径的性别识别模型。

[0100] 输入向量维数 x_t 为 N_c ,隐含层神经元数量为200, W_f 、 W_i 、 W_c 与 W_o 大小为 $200 \times (200 + N_c)$,因此,单元状态向量的维数与隐含层相同,均为200。

[0101] 步骤5,测试待评估观察者扫视路径序列

[0102] 利用眼动仪记录当前待评估观察者在所有测试图像上的扫视路径。采用同训练过程同样的方式提取各个扫视路径特征,即对于每一个扫视路径,找到其上每一个注视点所在的过分割区域,并提取该过分割区域中心对应的语义哈希编码,得到扫视路径对应的特征序列,将该序列输入到训练后的LSTM模型中,得到该序列对应的回归层得分。

[0103] 步骤6,判断待评估观察者人群组别

[0104] 首先根据训练集上观察者的平均得分来计算两个相邻组别之间的平均得分划分阈值。参照图5,计算两个相邻组别中20名观察者在60幅图像上的平均得分,得到40个平均得分的数据集合,再利用最大类间方差法来对该集合数据进行阈值划分,得到男性组别与女性组别扫视路径平均得分的划分阈值 $thr1$,将平均得分划分为2个区间。

[0105] 最后计算待评估观察者扫视路径序列的平均得分,并判断该平均得分所在组别区间,完成观察者性别识别判断。

[0106] 实施例2:基于扫视路径学习的年龄段识别

[0107] 参照图1,本发明的具体实现步骤如下:

[0108] 步骤1,构建测试图像库

[0109] 本发明共选择两类图像来记录眼动。第一类图像来自OSIE数据库的30幅蕴含语义情感特性及社会性的自然图像,包括面部表情、游戏场景、对话场景等等;第二类图像来自30幅儿童读物与绘本图像。

[0110] 步骤2,采集不同人群组别扫视路径序列

[0111] 组织眼动实验,分别采集不同性别组别观察者在60幅测试图像上的眼动数据,获取测试图像上对应的扫视路径序列。对于年龄段识别问题,人群组别年龄段分别设定为2-6岁,6-10岁,10岁-18岁,18岁以上。对于4个年龄段组别,各采集20名观察者的眼动数据。眼动记录设备可采用善睐眼动平台2.0下的便携式眼动仪。眼动实验屏幕分辨率为 1920×1080 ,考虑到儿童视觉角问题,测试图像大小设置为 800×600 并放置于屏幕中心,其余部分用灰色模板填充。观察者坐在距离屏幕60cm-75cm左右的位置。每个测试图像呈现3s,不同测试图像间隔为0.5s,用一幅灰色模板来实现。

[0112] 步骤3,提取扫视路径序列特征

[0113] 这一步骤的目标在于构建扫视路径特征提取网络,提取观察者在测试图像上扫视路径序列对应的特征,为下一步训练LSTM网络做准备。

[0114] (3a) 图像过分割

[0115] 首先采用简单线性迭代聚类(simple linear iterative clustering,SLIC)算法

来将图像分割为多个过分割区域,其中每个过分割区域由相邻且具有相似特性的像素点组成。接下来再以这些过分割区域代替像素点作为图像基本处理单元,来减小图像噪声等对最终识别结果的影响。

[0116] (3b) 提取扫视路径特征

[0117] 接下来任务是利用语义哈希 (Semantic Hashing) 算法来对采集的扫视路径序列进行特征提取,将扫视路径序列上各个注视点所在过分割区域对应的原始图像数据变换到具有更强辨别能力的语义空间,进而提升基于LSTM网络的识别能力。参照图2,语义哈希算法采用深层自动编码器作为学习网络。该网络由共用中心层且结构对称的编码器与解码器两部分组成。中心层为二值神经元,其余所有神经元均为逻辑神经元。编码器由5层网络结构组成,相邻两层网络是全连接的关系, N_0 为输入层的神经元数量,其大小由输入图像块对应的向量维数 $d \times d \times 3$ 决定。该网络一个重要特点是编码器神经元个数逐层减少, $N_i \geq 2N_{i+1}, i=0,1,2, N_3 \geq 2N_c$,中心层神经元数量 N_c 通常远小于输入层神经元数量 N_0 。另一方面,又需要通过解码器来在网络输出端重构出输入,这就迫使网络中心层能够发掘出高维输入数据中的内在结构与关系,获得从底层图像信息到语义特征的映射关系。

[0118] 为了获取网络参数,利用从OSIE数据库700幅图像中所有过分割块中心提取的340,000个大小为 $d \times d \times 3$ 的图像块来进行网络训练。对于每个样本点 x ,将其对应的长度为 d 的原始图像块 $p(x)$ 转化为 $d \times d \times 3$ 维列向量 $f(x)$ 作为网络输入,再将输入向量 $f(x)$ 作为对应的网络标记输出,来构建语义哈希编码特征提取网络训练样本。

[0119] 语义哈希编码特征提取网络的训练过程分为预训练与反向传播微调参数两个阶段,首先是预训练的过程,来给出网络一组初始参数 $H(\cdot, \theta)$ 。首先将编码器看作一组受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM) 的组合,训练当前RBM得到隐层单元的激活概率将被作为下一个RBM的可见层单元继续迭代训练,来逐层初始化编码器,解码器的初值是编码器的转置。这样能将整个学习任务分解成多个简单的任务,并消除了在定向生成式模型中出现的推理问题。

[0120] 接下来是反向传播阶段,在初始网络 $H(\cdot, \theta)$ 下,计算网络输出 $H(f(x), \theta)$,以及输出向量 $H(f(x), \theta)$ 和标记输出 $f(x)$ 之间的交叉熵误差 $CE(x)$:

$$[0121] \quad CE(x) = -\sum_i f(x)_i \log(H(f(x), \theta)_i) - \sum_i (1 - f(x)_i) \log(1 - H(f(x), \theta)_i),$$

[0122] 其中, $f(x)_i$ 为标记输出 $f(x)$ 的第 i 维元素, $H(f(x), \theta)_i$ 为实际输出 $H(f(x), \theta)$ 的第 i 维元素。根据所有训练样本交叉熵误差的平均值来利用反向传播算法自顶向下微调整个网络参数 $H(\cdot, \theta)$,提高模型的鉴别能力。

[0123] 在训练语义哈希编码特征提取网络的基础上,下面需要对扫视路径序列进行特征提取。参照图3,对于长度为 m 扫视路径序列 $(q_1, q_2, \dots, q_t, \dots, q_m)$ 上 t 时刻注视点 q_t ,找到其在图像过分割区域,提取该区域中心大小为 $d \times d \times 3$ 的原始图像块,将其对应的向量 $f(q_t)$ 输入到语义哈希编码特征提取网络中,提取中心层 N_c 维二值特征 $H(q_t)$ 。将长度为 m 的扫视路径序列,转化为对应的 $N_c \times m$ 特征。

[0124] 步骤4,训练LSTM网络

[0125] 首先在上一步骤的基础上构建训练LSTM网络需要用到的训练样本集。对于年龄段识别问题,提取各个年龄段扫视路径序列特征,并给出对应的人群组别标记1(2-6岁)、2(6-

10岁)、3(10岁-18岁)、4(18岁以上)。

[0126] 参照图4,构建用于年龄段识别的LSTM网络。该网络由两部分组成,迭代循环部分以及顶层回归层。循环部分包含两个状态:单元状态 c 和隐含层 h 。在 t 时刻,LSTM网络将受到三个输入的共同作用:当前时刻LSTM网络的输入值 x_t 、上一时刻LSTM网络的输出值 h_{t-1} 、以及上一时刻的单元状态 c_{t-1} 。同样在 t 时刻,LSTM网络有两部分输出:当前时刻LSTM网络的输出值 h_t ,以及当前时刻的单元状态 c_t 。所有输入输出都是向量的形式。

[0127] LSTM网络用两个门来控制单元状态 c 的内容,一个是遗忘门(forget gate),它决定了上一时刻的单元状态有多少保留到当前时刻;另一个是输入门(input gate),它决定了当前时刻网络的输入有多少保存到单元状态。LSTM网络用输出门(output gate)来控制单元状态有多少输出到LSTM的当前输出值。

[0128] 首先是遗忘门输出的计算:

$$[0129] \quad f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f),$$

[0130] 其中, W_f 与 b_f 分别为遗忘门的权重与偏置项, $\sigma(\cdot)$ 为sigmoid函数。

[0131] 其次是输入门输出计算:

$$[0132] \quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i),$$

[0133] 其中, W_i 与 b_i 分别为输入门的权重与偏置项。

[0134] 再次,计算用于描述输入的状态 \bar{c}_t :

$$[0135] \quad \bar{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c),$$

[0136] 其中, W_c 与 b_c 分别为权重与偏置项。 $\tanh(\cdot)$ 为tanh函数。

[0137] 接下来,则是计算当前 t 时刻的单元状态 c_t :

$$[0138] \quad c_t = f_t \odot c_{t-1} + i_t \odot \bar{c}_t,$$

[0139] 其中 \odot 表示按元素乘操作,通过计算 c_t ,可以将当前短期记忆 \bar{c}_t 与长期记忆 c_{t-1} 的影响组合起来。由于遗忘门的设定,单元状态可以保存很久之前的信息,由于输入门的设定,可以避免将当前无关紧要的内容存入记忆。

[0140] 下面,是输出门的计算:

$$[0141] \quad o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o),$$

[0142] 其中, W_o 与 b_o 分别为输出门的权重与偏置项。在计算输出门的基础上,最后是当前时刻下最终输出的计算:

$$[0143] \quad h_t = o_t \odot \tanh(c_t).$$

[0144] 假设扫视路径序列长度为 m ,在前向传播到最后一个注视点时,将最终输出 h_m 与神经元个数为50的全连接层 $FF(\cdot)$ 连接,得到全连接层输出 y_m :

$$[0145] \quad y_m = FF(h_m)$$

[0146] 再将 y_m 与包含一个神经元的回归层全连接起来,得到该扫视路径序列对应的得分score。

$$[0147] \quad \text{score} = \text{regression}(y_m)$$

[0148] 在将训练样本集扫视路径序列特征输入到网络中进行前向传播的基础上,使用BPTT算法来更新网络参数。具体做法是将连续注视点对应的 N_c 维二值语义哈希编码作为网络输入,得到最顶层回归层的预测值。再计算预测值与对应人群组别标记之间的预测误差,

并根据该误差计算每个权重的梯度,沿时间反向传递误差项,调整网络的参数,最终得到基于扫视路径的年龄段识别模型。

[0149] 输入向量维数 x_t 为 N_c ,隐含层神经元数量为200, W_f 、 W_i 、 W_c 与 W_o 大小为 $200 \times (200 + N_c)$,因此,单元状态向量的维数与隐含层相同,均为200。

[0150] 步骤5,测试待评估观察者扫视路径序列

[0151] 利用眼动仪记录当前待评估观察者在所有测试图像上的扫视路径。参照图3,采用同训练过程同样的方式提取各个扫视路径特征,即对于每一个扫视路径,找到其上每一个注视点所在的过分割区域,并提取该过分割区域中心对应的语义哈希编码,得到扫视路径对应的特征序列,将该序列输入到训练后的LSTM模型中,得到该序列对应的回归层得分。

[0152] 步骤6,判断待评估观察者人群组别

[0153] 首先根据训练集上观察者的平均得分来依次计算两个相邻组别之间的平均得分划分阈值。参照图5,依次计算两个相邻组别中20名观察者在60幅图像上的平均得分,得到40个平均得分的数据集合,再利用最大类间方差法来对该集合数据计算组别划分阈值。分别得到第一组别与第二组别平均得分的划分阈值 $thr1$,第二组别与第三组别的划分阈值 $thr2$,以此类推,将4个组别人群识别问题下的平均得分划分为4个区间。

[0154] 最后计算待评估观察者扫视路径序列的平均得分,并判断该平均得分所在组别区间,完成观察者年龄段识别判断。



图1

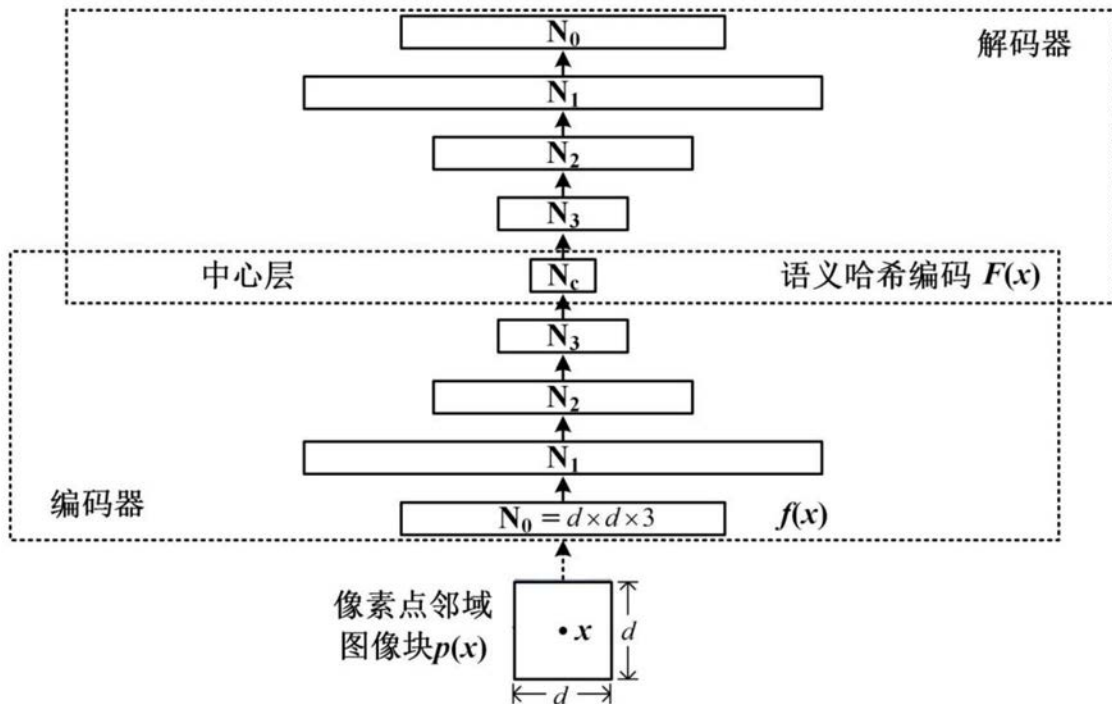


图2

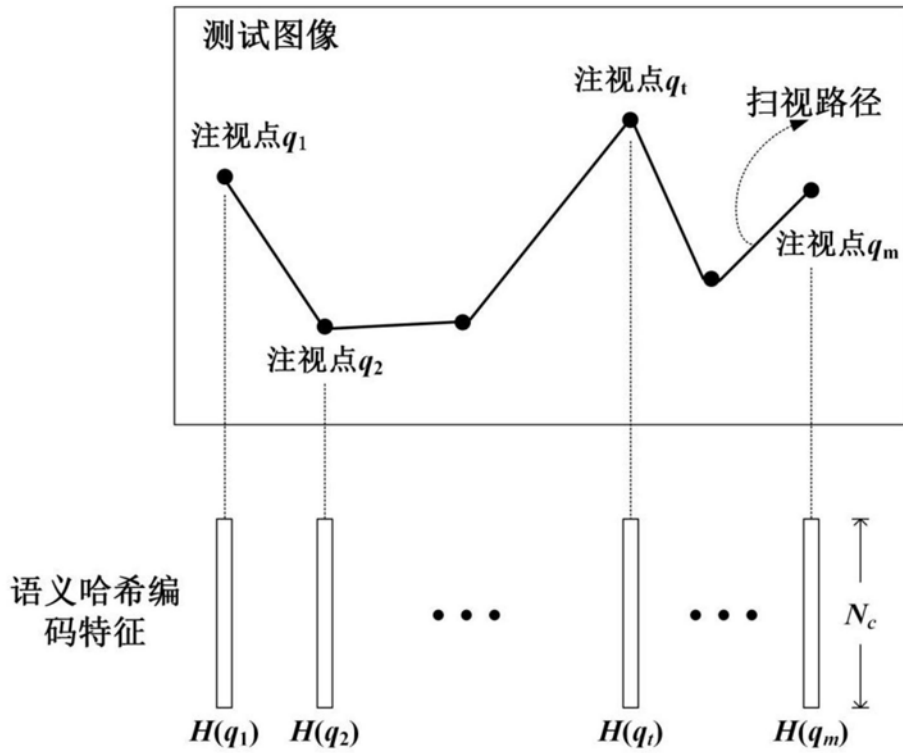


图3

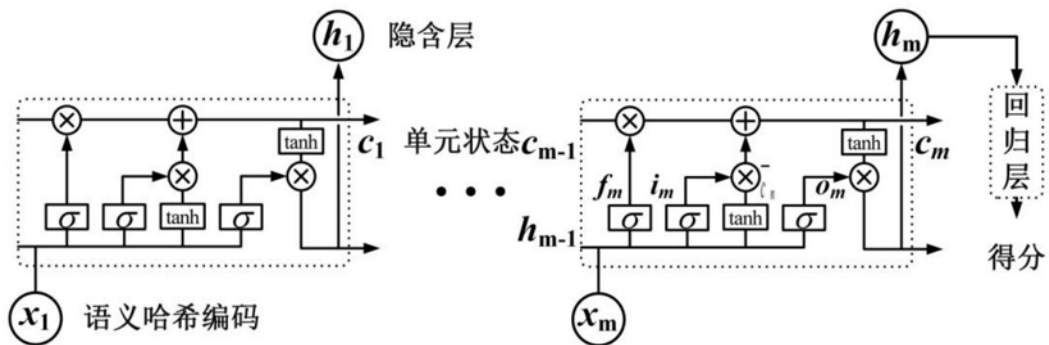


图4

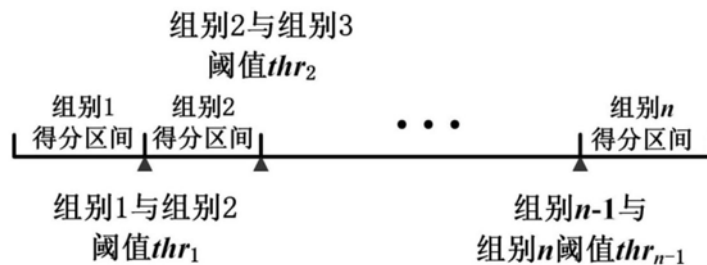


图5