Creative Education Studies 创新教育研究, 2018, 6(4), 299-305
Published Online August 2018 in Hans. http://www.hanspub.org/journal/ces
https://doi.org/10.12677/ces.2018.64047

Hans汉斯

A MOOC-Oriented Emotion Recognition System

Jiangqin Xu[1], Yongfeng Zhang[2]

[1]School of Foreign Language, Xiamen University, Xiamen Fujian
[2]Yabo-Fengtian Technology Co. Ltd., Xiamen Fujian
Email: vanessaxjq@163.com

Received: Jul. 24th, 2018; accepted: Aug. 6th, 2018; published: Aug. 13th, 2018

Abstract

Accurate and rapid detect changes of learner's emotional state is of great importance to improve the teaching quality of Massive Open Online Courses (MOOC). However, the emotion recognition tools for MOOC must solve the two key issues: robustness and real-time. In this study, we proposed a deep learning approach which is based on the Computer Unified Device Architecture (CUDA) technology, called CUDA-DeSTIN, to quickly and accurately identify the learner's facial emotional state. We tested our method using AR data with different noise, and compared the results with other deep learning methods. The experimental results prove the effectiveness of our method.

Keywords

MOOC, Emotion Recognition, Artificial Intelligence

面向慕课的情绪识别系统

徐姜琴[1]，张永锋[2]

[1]厦门大学外国语学院，福建 厦门
[2]厦门亚伯锋天科技有限公司，福建 厦门
Email: vanessaxjq@163.com

收稿日期：2018年7月24日；录用日期：2018年8月6日；发布日期：2018年8月13日

摘 要

准确而快速地发现学习者的情绪变化，对提高慕课的教学质量具有极为重要的价值。然而面向慕课的情绪识别工具必须解决鲁棒性和实时性这两个关键问题。在这项研究中，我们提出了使用Computer Uni-

文章引用: 徐姜琴, 张永锋. 面向慕课的情绪识别系统[J]. 创新教育研究, 2018, 6(4): 299-305.
DOI: 10.12677/ces.2018.64047

fied Device Architecture (CUDA)技术来对深度时空推理网络进行加速，从而快速而准确的识别学习者的面部情绪状态。我们利用添加不同噪声的AR数据来测试了我们的方法，并将结果同其他深度学习方法进行了对比。实验结果证明了我们的方法的有效性。

关键词

慕课，情绪识别，人工智能

1. 引言

近年来，随着大规模开放网络课程(Massive Open Online Courses, MOOC，以下简称慕课) [1]和其它类型的 E-learning 方法的兴起，越来越多的知识交互活动，借由这些新的形式开展。如何借助新的技术手段，来提高相应的教学活动质量，具有重要的研究价值，同时也极具挑战。较传统的教学手段而言[2]，这类新兴的技术手段有很多优势，但与此同时也存在着一些新特性，需要加以认真考虑[3]。例如，授课教师往往与学生不在同一地点，从而使得教授者难以及时发现而如何加强教学活动中的交互质量，是评价慕课是否真正成功的重要依据。

众多已有的研究已经表明，情绪对学习的效果[4] [5]有着至关重要的影响。这也意味着，学习者的情绪会严重影响其知识掌握程度和总体学习目标。因此，对于以慕课为代表的 E-Learning 来说，有效教学活动的前提条件之一是教师能够快速响应学生的情绪变化。然而，如何在复杂的学习环境中准确、快速地理解学生的情感互动仍然是一项非常具有挑战性的任务。

利用人工智能技术来提高教学质量的需求吸引了相当多的研究者的兴趣[6] [7]。在许多尝试中，使用新的模式识别技术来检测学习者情绪变化的方法被认为是一种有效的方法。例如，在[6]中，我们提出了一种基于深度学习的方法来检测学习者面部表情，从而判断其情绪的变化。在[7]中，作者提出了一种基于 AdaBoost 的人眼状态检测方法，并使用该方法来判断学习者的情绪状态。AutoTutor [8]研究项目希望检测和利用学习者的情绪，以加强学习和教学过程。在[9]中，构造了一个神经网络体系结构，能够处理语音中的面部特征、韵律和词汇内容的融合。Yau 等人在其论文[10]中提出了一种鲁棒的人脸表情识别方法。

深度学习[11]是近年来在人工智能领域受到极大关注的创新性技术手段，这类方法较其他模式识别方法而言，具有的最大优点是能够自动抽取出非常高质量的特征，而高质量的特征是一个成功的情感识别系统的重要组成部分。

依据我们已有的研究成果，EM-DeSTIN [12]方法在处理噪声图像时非常有效，而在 E-Learning 系统中获得的图像通常包含噪声，因此，使用 EM-DeSTIN 方法能够帮助我们准确地识别特定场景下参与者的情绪。但是，在交互学习的过程中，我们往往要求情绪识别系统具有较好的实时性，换句话说，现实世界中的 E-Learning 要求系统必须能够较快的识别出学习者的情绪变化。而现有的深度学习方法，尤其是深度时空推理网络，往往需要很长的训练时间。

因此，在本文中，我们利用了一种基于 CUDA 架构的深度时空推理网络[13]，CUDA-DeSTIN，来构建面向 E-Learning 环境的情感识别系统。该系统的特点在于，它不仅能够在高噪声的环境中工作，同时其学习速度较现有的方法也有了较大的提升，更加适合慕课等新兴的教学方法。

本文的其余部分整理如下。在第二节中，我们将简要介绍深空推理网络(DeSTIN)和 CUDA 技术，第三节我们介绍基于 CUDA 架构的深度时空推理网络 CUDA-DeSTIN。在第四节中，我们将介绍实验内容以及实验结果。最后，在第五节中，我们将给出了一个简短的结论和未来的研究方向。

2. 背景知识

2.1. 深度时空推理网络

深度时空推理网络(Deep Spatio-Temporal Inference Network, DeSTIN) [14]是一种尝试模仿人类视觉和听觉皮层的概念结构和动力学模型。DeSTIN 的特殊优势在于它具有较为出色的时间数据处理能力，这种模型能够从较少量的训练实例中学习时空特征，并与外部认知软件系统进行交互。而这种能力对于，面向慕课的系统而言，尤其重要。

如图 1 所示，在宏观层面上，DeSTIN 是一种分层结构的模型。其每一层(Layer)都被分为若干个 2×2 的区域。第 $N-1$ 层的某个 2×2 的区域会作为第 N 层的一个节点(Node)的输入，而第 N 层的四个节点又会与第 $N+1$ 层的一个特定节点相连。在 DeSTIN 网络的最底层，原始的图像数据会直接输入。可以将 DeSTIN 网络中的每一层都理解为原始输入数据在特定层次上的抽象。在中间层每一节点都包着数量聚类中心(Centrino)，而节点可以通过无监督学习产生与到本层的观察值特征。

在这样的结构中，每个节点输出在其所在层的信念状态，这些信念状态会包含输入数据中存在的空间和时间规律。该系统顶层的输出可以作为有监督学习算法(如神经网络，支持向量机(Support Vector Machine))的输入，从而进行有效的模式分类。

在图 1 所示的例子中，数字“9”被直接输入第四层(Layer 4)，第四层是该系统的底层，包含 $64 = 8 \times 8$ 个节点，该层接受原始图像的像素作为输入，每个节点对应一个的 4×4 像素区域，每个节点对应的

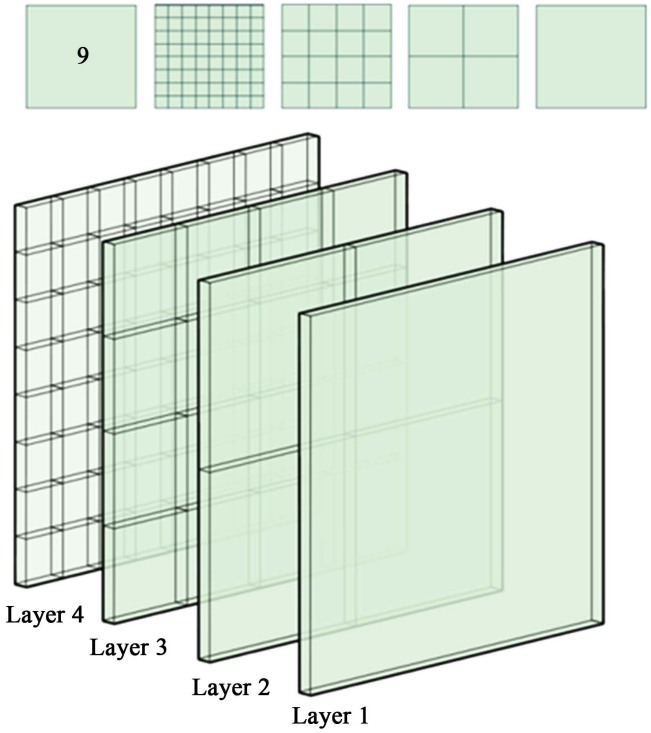

Figure 1. DeSTIN structure diagram
图 1. DeSTIN 结构示意图

像素区域不重叠，每个 4 × 4 的像素区域按照行优先的方式存储成一个 16 维的列向量，这个 16 维的列向量会作为该层输入数据。第四层的上一层是第三层(Layer 3)，第三层包含 16 = 4 × 4 个节点，每一个节点接受第四层相对应的 4 = 2 × 2 个节点的输出作为输入，每个节点的对应的区域也是不重叠的。第三层的上一层是第二层(Layer 2)，第二层包含 4 = 2 × 2 个节点，每一个节点接受第三层相对应的 4 = 2 × 2 个节点的输出作为输入，同样每个节点的对应的区域也是不重叠的。第一层(Layer 1)是该系统的顶层，它包含 1 = 1 × 1 个节点，该节点接受第二层 4 = 2 × 2 个节点的输出作为输入。该节点输出也就是 DESTIN 的输出是一个 10 维的向量，该向量可以作为有监督学习算法(如神经网络，支持向量机)的输入，从而进行有效的模式分类。

2.2. CUDA：可伸缩并行编程模型

CUDA (Computer Unified Device Architecture) [15]是英伟达(NVIDIA)公司开发的一种并行计算架构[15]。这种架构的主要出发点在于，随着技术不断发展，GPU (Graphic Processing Unit)变的越来越强大，对于某些特定的计算任务，其能力已经远远超越了通用的 CPU。在此情形下，CPU 与 GPU “协同处理”模式能够为我们带来更加高效的计算能力。

基于此，英伟达(NVIDIA)发明了 CUDA 并行计算架构。该架构通过利用 GPU 超强的处理能力，可以大幅度提升计算性能。CUDA 采用 C 语言作为编程语言提供大量的高性能计算指令开发能力，使开发者能够在 GPU 的强大计算能力的基础上建立起一种效率更高的密集数据计算解决方案。

在 CUDA 的架构下，一个程序分为两个部份：宿主(host)端和设备(device)端[15]。宿主(Host)端是指在 CPU 上执行的部份，而设备(device)端则是在显示芯片上执行的部份。设备(device)端的程序又称为“kernel (核心)”。通常宿主(host)端程序会将数据准备好后，复制到显卡的内存中，再由显示芯片执行设备(device)端程序，完成后再由宿主(host)端程序将结果从显卡的内存中取回。不同类型的代码由于其运行的物理位置不同，能够访问到的资源不同，因此对应的运行期组件也分为公共组件、宿主组件和设备组件三个部分。更多有关 CUDA 的细节，读者可以参阅[15]。

3. CUDA-DESTIN

在这一节的前半部分，我们简要的介绍我们所提出的 CUDA-DeSTIN [13]的技术细节。然后，我们讨论，如何利用这种方法来进行面部表情的情绪识别。

3.1. CUDA-DeSTIN

简要地将，CUDA-DeSTIN 就是利用 CUDA 的架构下，重新实现了 DeSTIN，从而使得训练速度得到了极大的提升。

有几点需要说明的是，在 CUDA-DeSTIN 中，聚类中心的数据存储到全局内存(global memory)中，每一个线程在计算相似度时会用到该数据。DeSTIN 中的一个节点(node)对应一个区块(block)，每一个区块有一个共享(shared memory)，其大小为 16 kilobytes。第四层(Layer 4)中的每个节点的输入大小为 16 × 4 = 64 bytes，第三层(Layer 3)中的每个节点的输入大小为 100 × 4 = 400 bytes，第二层(Layer 2)中的每个节点的输入大小为 64 × 4 = 256 bytes，第一层(Layer 1)中的每个 node 的输入大小为 48 × 4 = 192 bytes。

每个区块(block)中的每一线程(thread)等同于 CUDA-DeSIN 中的一个聚类中心(centroid)。线程的数据，也就是聚类中心的每一维数据存储在全局变量(global memory)中，输入数据存储在区块(block)的共享内存(shared memory)中。然后计算相似度，并将结果存储到共享内存(shared memory)中。所有的线程计算完成之后，进入排序阶段。

在 CUDA-DeSTIN 中，采用的是快速排序(Quicksort)，这种方法是对冒泡排序的一种改进。由 C. A. R. Hoare 在 1962 年提出。最后优胜的聚类中心(winning centroid)会存储在全局内存(global memory)中，在更新优胜的聚类中心(winning centroid)时，该数据会被再次使用。有关 CUDA-DeSTIN 的更多细节可以参考文献[13]。

3.2. CUDA-DeSTIN 的情绪识别方法

利用 CUDA-DeSTIN 来进行基于面部图像的情绪识别，大体可以分为训练阶段和识别阶段。训练 CUDA-DeSTIN 的过程可以分为如下步骤：

1) CUDA-DeSTIN 中的每一层(除最底层外)，从下一层获得其输出的 belief 作为观测值。

2) 通过聚类中心更新算法得到新的聚类中心，然后依据更新后的聚类中心，计算概率分布向量 Belief，并将此概率分布向量作为输出，向上传输到父层作为父层的观察值。

3) 在进行在线类聚的同时，CUDA-DeSTIN 也会将父层类聚的赢者信息反馈到本层，更新本层的 PSSA 表，从而起到父层对子层的指导作用。

4) CUDA-DeSTIN 中的每一层都按照此流程进行，在金字塔顶端的 belief 就是输入图像的关键特征。

5) 利用最上层的 belief 作为一种有标签的训练数据，训练支持向量机(Support Vector Machine, SVM)，从而实现对相应的分类器的训练过程。

当完成了训练过程之后，我们就得到了一个能够识别不同面部情绪的分类器(一个训练好的 SVM 和一个训练好的 CUDA-DeSTIN 模型)。可以利用以下的步骤来实现识别过程：

1) 将要识别的面部图片输入到 CUDA-DeSTIN 中，由此获得一组 Belief 作为观测值。

2) 根据训练得到的类聚中心与 PSSA (不需要进行类聚中心的更新)，计算概率分布向量 Belief，并将 Belief 传输到父层作为父层的观察值。

3) 按照上述的过程，在训练好的 CUDA-DeSTIN 的顶部，输出的 Belief 就是待识别图片的特征。

4) 将第三步得到的特征，送入已训练好的分类器 SVM，实现情绪识别。

4. 实验

我们利用了 AR 人脸数据集[1]的一个子集来进行实验。该子集由 100 个人组成，每个人有 26 个灰度图像，分别是 165 × 120 像素和 24 位深度。这些图像可根据拍摄日期分成两组。我们以一组作为训练样本，另一组作为测试组。数据集的样本显示在图 2 中。

图 2(a)中的图 AR 样本没有特定的情绪倾向；图 2(b)表示微笑；图 2(c)是生气，而是愤怒。为了模拟在慕课环境中，视频图像质量往往不高的情况，高斯噪声和盐和椒盐噪声被添加到训练样本中。对于高斯噪声，均值为 0，方差分别设为 0.01，0.06 和 0.1。对于椒盐噪音，噪音密度分别设定为 0.01，0.1 和 0.5。

在我们的实验中，我们将 CUDA-DeSTIN 和 DeSTIN 的顶层两层输出作为输入特征来训练 SVM，并使用训练的 SVM 对相同的测试样本进行分类分类。为了测试其他深层网络，我们选择了三种不同的方法：深信念网(DBN)，卷积神经网络(CNN)和堆栈自动编码器(SAE)等于。这三种其他方法的源代码是从 DeepLearnToolbox[2] 获得的。

表 1 记录不同噪音环境下的实验结果。实验表明，该算法在处理低到中等噪声水平时具有较强的竞争力，即使在噪声较大的情况下，EM-DeSTIN 也具有良好的性能。

[1] http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html
[2] https://github.com/rasmusbergpalm/DeepLearnToolbox

Figure 2. Samples from the AR Dataset
图 2. AR 数据库中的样本

Table 1. Emotion Recognition of CUDA-DeSTIN, uniform DeSTIN, DBN, CNN and Autoencoder
表 1. 五种不同深度系统的识别正确率

	高斯噪声			椒盐噪声		
	0.01	0.06	0.1	0.01	0.1	0.5
DBN	0.35	0.33	0.30	0.41	0.38	0.32
CNN	**0.43**	0.41	**0.40**	**0.45**	0.42	0.33
SAE	0.41	0.40	0.38	0.41	0.39	**0.72**
DeSTIN	0.40	0.38	0.32	0.40	0.40	0.27
CUDA-DeSTIN	0.41	**0.42**	0.35	0.40	**0.43**	0.27

另外尤其需要说明的是，在 Tesla 上实现了基于 CUDA-DeSTIN 的面部情绪识别，训练时间只有 1200 秒(约 0.3 小时)，而 DeSTIN 算法在 2.4 GHz 的 CPU 上训练需要 50 小时，速度提高了约 166 倍。在识别率在优的情况下，训练速度得到了极大的提升，这一特点，对于在线学习工具而言，具有至关重要的优势。

5. 总结

在本文中，我们提出利用一种 CUDA 加速的深度学习方法，来实现慕课教学中的情感识别。通常情况下，面向慕课的教学工具，必须考虑两个问题。首先，慕课环境下的视频图像往往具有较多的噪声，图像质量不高。其次，情绪识别必须具有较快的响应速度，也就是实时性必须较好。我们的方法的优点是在处理嘈杂的面部表情图片时，它具有较高的识别率。同时，该方法较原有的深度时空推理网络而言，训练速度得到了极大的提高。在未来，我们将进一步研究如何利用其它先进的智能方法[16] [17] [18] [19] [20]来为高质量的慕课教学服务。

致　谢

感谢厦门大学信息科学与技术学院仿脑智能实验室为本研究提供了相关的代码，并协助进行了实验。

基金项目

此研究受到国家自然科学基金(No. 61673328)资助。

参考文献

[1] Daradoumis, T., Bassi, R., Xhafa, F., *et al.* (2013) A Review on Massive E-Learning (MOOC) Design, Delivery and

Assessment. 8*th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing* (3*PGCIC*), Compiegne, 28-30 October 2013, 208-213.

[2] Xu, J.Q., Liu, F. and Jiang, M. (2008) Task-Based Language Teaching: From the Practical Perspective. 2008 *International Conference on Computer Science and Software Engineering*, Hubei, 12-14 December 2008, 1054-1057.

[3] Binali, H.H., Wu, C. and Potdar, V. (2009) A New Significant Area: Emotion Detection in E-Learning Using Opinion Mining Techniques. 3*rd IEEE International Conference on Digital Ecosystems and Technologies*, Istanbul, 1-3 June 2009, 259-264.

[4] Sylwester, R. (1994) How Emotions Affect Learning. *Educational Leadership*, **52**, 60-65.

[5] Bower, G.H. (1992) How Might Emotions Affect Learning? *The Handbook of Emotion and Memory*: *Research and Theory*, **3**, 31.

[6] Xu, J., Huang, Z., Shi, M., *et al*. (2017) Emotion Detection in E-learning Using Expectation-Maximization Deep Spatial-Temporal Inference Network. In: *UK Workshop on Computational Intelligence*, Springer, 245-252.

[7] Shen, L., Wang, M. and Shen, R. (2009) Affective E-Learning: Using "Emotional" Data to Improve Learning in Pervasive Learning Environment. *Journal of Educational Technology & Society*, **12**, 176.

[8] Graesser, A.C., Wiemer-Hastings, K., Wiemer-Hastings, P., *et al*. (1999) AutoTutor: A Simulation of a Human Tutor. *Cognitive Systems Research*, **1**, 35-51. https://doi.org/10.1016/S1389-0417(99)00005-4

[9] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., *et al*. (2001) Emotion Recognition in Human-Computer Interaction. *IEEE Signal Processing Magazine*, **18**, 32-80. https://doi.org/10.1109/79.911197

[10] Shojaeilangari, S., Yau, W.-Y., Nandakumar, K., *et al*. (2015) Robust Representation and Recognition of Facial Emotions using Extreme Sparse Learning. *IEEE Transactions on Image Processing*, **24**, 2140-2152. https://doi.org/10.1109/TIP.2015.2416634

[11] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, **521**, 436-444. https://doi.org/10.1038/nature14539

[12] Jiang, M., Ding, Y., Goertzel, B., *et al*. (2014) Improving Machine Vision via Incorporating Expectation-Maximization into Deep Spatio-Temporal Learning. *International Joint Conference on Neural Networks*, Beijing, 6-11 July 2014, 1804-1811.

[13] 张永锋. 基于 CUDA 架构的深度时空推理网络[D]: [硕士学位论文]. 厦门: 厦门大学, 2012.

[14] Arel, I., Rose, D.C. and Coop, R. (2009) DeSTIN: A Scalable Deep Learning Architecture with Application to High-Dimensional Robust Pattern Recognition. *AAAI Fall Symposium*: *Biologically Inspired Cognitive Architectures*, Arlington, 5-7 November 2009, 11-15.

[15] NVIDIA (2011) NVIDIA CUDA C Programming Guide. Nvidia Corporation, Santa Clara, 120, 8.

[16] Jiang, M., Huang, Z., Qiu, L., *et al*. (2017) Transfer Learning Based Dynamic Multiobjective Optimization Algorithms. *IEEE Transactions on Evolutionary Computation*, 1. https://doi.org/10.1109/TEVC.2017.2771451

[17] Goertzel, B., De Garis, H., Pennachin, C., *et al*. (2010) OpenCogBot: Achieving Generally Intelligent Virtual Agent Control and Humanoid Robotics via Cognitive Synergy. *Proceedings of ICAI*, **10**, 1-12.

[18] Jiang, M., Huang, W., Huang, Z., *et al*. (2017) Integration of Global and Local Metrics for Domain Adaptation Learning via Dimensionality Reduction. *IEEE Transactions on Cybernetics*, **47**, 38-51. https://doi.org/10.1109/TCYB.2015.2502483

[19] Zhang, X., Jiang, M., Zhou, C., *et al*. (2012) Graded BDI Models for Agent Architectures Based on Lukasiewicz Logic and Propositional Dynamic Logic. In: *International Conference on Web Information Systems and Mining*, Springer, Berlin, 439-450.

[20] Jiang, M., Qiu, L., Huang, Z. and Yen, G.G. (2018) Dynamic Multi-Objective Estimation of Distribution Algorithm Based on Domain Adaptation and Nonparametric Estimation. *Information Sciences*, **435**, 203-223. https://doi.org/10.1016/j.ins.2017.12.058

知网检索的两种方式：

1. 打开知网页面 http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD
   下拉列表框选择：[ISSN]，输入期刊 ISSN：2331-799X，即可查询
2. 打开知网首页 http://cnki.net/
   左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：http://www.hanspub.org/Submission.aspx
期刊邮箱：ces@hanspub.org