

面向边缘计算应用的宽度孪生网络

李逸楷¹ 张通^{1,2} 陈俊龙^{1,2}

摘要 边缘计算是将计算、存储、通信等任务分配到网络边缘的计算模式。它强调在用户终端附近执行数据处理过程,以达到降低延迟,减少能耗,保护用户隐私等目的。然而网络边缘的计算、存储、能源资源有限,这给边缘计算应用的推广带来了新的挑战。随着边缘智能的兴起,人们更希望将边缘计算应用与人工智能技术结合起来,为我们的生活带来更多的便利。许多人工智能方法,如传统的深度学习方法,需要消耗大量的计算、存储资源,并且伴随着巨大的时间开销。这不利于强调低延迟的边缘计算应用的推广。为了解决这个问题,我们提出将宽度学习系统(Broad learning system, BLS)等浅层网络方法应用到边缘计算应用领域,并且设计了一种宽度孪生网络算法。我们将宽度学习系统与孪生网络结合起来用于解决分类问题。实验结果表明我们的方法能够在取得与传统深度学习方法相似精度的情况下降低时间和资源开销,从而更好地提高边缘计算应用的性能。

关键词 宽度学习系统, 边缘计算, 孪生网络, 浅层网络, 边缘智能

引用格式 李逸楷, 张通, 陈俊龙. 面向边缘计算应用的宽度孪生网络. 自动化学报, 2020, 46(10): 2060–2071

DOI 10.16383/j.aas.c200555

Broad Siamese Network for Edge Computing Applications

LI Yi-Kai¹ ZHANG Tong^{1,2} CHEN C. L. Philip^{1,2}

Abstract Edge computing is a paradigm that allocates computing, storage, communication tasks to the edge of network. It emphasizes that data processing should be placed in the proximity of user terminals in order to reduce latency and energy consumption while protecting user privacy. However, resources for computation, storage and power at the edge of network are limited, which brings new challenges to edge computing applications. With the emergence of edge intelligence, people prefer to combine edge computing applications with artificial intelligence techniques to bring more convenience to our lives. Many artificial intelligence methods such as traditional deep learning methods, need to consume a lot of computation and storage resources with a huge time consumption. It is not conducive to the popularity of edge computing applications, which always require low latency. In order to resolve such a problem, we propose that shallow network methods such as broad learning system can be applied in edge computing applications and design a broad Siamese network algorithm. We combine broad learning system (BLS) with Siamese network for classification tasks. Experimental results show that our method can reduce the cost of time and resources, while achieving a similar result as deep learning methods, consequently improving the performance of edge computing applications.

Key words Broad learning system (BLS), edge computing, Siamese network, shallow network, edge intelligence

Citation Li Yi-Kai, Zhang Tong, Chen Jun-Long. Broad Siamese network for edge computing applications. *Acta Automatica Sinica*, 2020, 46(10): 2060–2071

1 背景知识

收稿日期 2020-07-15 录用日期 2020-08-27

Manuscript received July 15, 2020; accepted August 27, 2020

国家自然科学基金(61702195, 61751202, U1813203, U1801262, 61751205), 国家科技部重点专项(2019YFA0706200, 2019YFB1703600), 广州市重大科技项目(202007030006)资助

Supported by National Natural Science Foundation of China (61702195, 61751202, U1813203, U1801262, 61751205), National Key Research and Development Program of China (2019YFA0706200, 2019YFB1703600), Science and Technology Major Project of Guangzhou (202007030006)

本文责任编辑 杨涛

Recommended by Associate Editor YANG Tao

1. 华南理工大学计算机科学与工程学院 广州 510006 2. 琶洲实验室 广州 510355

1. School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006 2. Pazhou Laboratory, Guangzhou 510335

本节将对宽度学习系统(Broad learning system, BLS)和孪生网络的有关背景知识进行简要的介绍。我们将分别从方法介绍和应用研究两个角度展开描述,对算法的有关设计思路以及实际应用进行介绍。

1.1 宽度学习系统

不同于深度学习方法,宽度学习系统^[1]提供了一种新的学习架构。它不是通过增加网络深度的方式来进行学习,而是以随机的方式将输入样本特征横向扩展到宽度空间。宽度学习系统网络主要包括两类结点,即特征节点和增强节点。输入宽度学习系统网络的样本数据需要进行处理转化为特征节点和增强节点才可以进行进一步的学习。其中,特征节点

由输入样本特征以随机的方式映射产生, 增强节点则是从特征节点以随机方式扩展得到. 随后系统将特征节点和增强节点拼接在一起, 通过联合求解得到最终的输出. 不同于以往的深度学习方法, 宽度学习系统需要学习的参数相对较少. 因为特征节点和增强节点的有关参数, 如结点权重和偏差只初始化一次, 在训练过程中只需学习由特征结点和增强结点连接到输出层的权值. 因此, 与深层网络方法相比, 宽度学习系统网络结构相对简单, 所需要学习的参数相对较少, 可以更加快速地对研究内容进行建模, 缩短了响应处理的时间. 图 1 给出了一种典型的宽度学习系统的网络结构.

1.1.1 方法介绍

宽度学习系统是在随机向量函数链接神经网络 (Random vector functional-link neural network, RVFLNN) 的基础上构建起来的. 传统的随机向量函数链接神经网络 (RVFLNN) 直接采用样本数据和生成的增强结点作为网络的输入信息. 而宽度学习系统则是先将输入样本特征映射为特征结点, 由特征结点生成增强结点, 再将特征结点和增强结点一起作为网络的输入信息. 宽度学习系统的构建可以大致分为三个步骤. 首先, 通过预先随机设定的映射矩阵将输入样本特征映射为特征结点. 随后, 生成的特征结点采用随机设定的权重和偏差进一步映射为增强结点. 值得注意的是, 生成特征结点和增强结点的权重和偏差值都是随机初始化的, 在后续的训练过程中不需要做进一步的学习调整. 同时特征结点和增强结点位于网络的同一层次, 这样就从宽度的层面拓展了网络的结构. 最后, 我们将特征结点和增强结点与输出结点连接起来, 并且采用求解伪逆的方法来求解连接的权重. 在这个过程中, 我们规定

特征结点以及增强结点与输出结点之间的连接只设置权重, 没有偏差值. 由于不需要像深度学习方法一样通过逐轮的迭代来训练相关的参数, 并且所需要的学习训练的参数相对较少, 宽度学习系统的训练过程相对较快, 从而减少了时间开销.

以分类任务为例子, 宽度学习系统的步骤介绍如下. 我们首先假设输入的样本数据为 $\{X, Y\} \in \mathbf{R}^{N \times (M+C)}$. 其中 X 是样本特征数据, Y 为样本标签, N 代表样本的数量, M 表示输入样本特征的维度, C 表示类别的数量. 在将输入样本特征转化为特征结点之前, 我们还需要定义参数 n 和参数 p . 其中 n 表示特征结点的组数, p 表示每组特征结点有 p 个结点. 将输入样本特征转化为特征结点的过程如式 (1) 所示.

$$Z_i = \varphi_i(XW_{zi} + \beta_{zi}), \quad i = 1, \dots, n \quad (1)$$

其中, $X \in \mathbf{R}^{N \times M}$, $W_{zi} \in \mathbf{R}^{M \times p}$, $\beta_{zi} \in \mathbf{R}^{1 \times p}$. 并且 W_{zi} , β_{zi} 以随机的方式生成, 分别代表用于产生第 i 组特征结点的权重和偏差. Z_i 表示第 i 组生成的特征结点. 我们将所有的 n 组特征结点拼接在一起得到最终的特征结点集合 Z^n , 其中 $Z^n = [Z_1, \dots, Z_n]$.

类似地我们可以规定增强结点的组数为 m , 每一组增强结点的个数为 q . 则由特征结点生成增强结点的过程如式 (2) 所示.

$$H_j = \xi_j(Z^n W_{hj} + \beta_{hj}), \quad j = 1, \dots, m \quad (2)$$

其中, $Z^n \in \mathbf{R}^{N \times np}$, $W_{hj} \in \mathbf{R}^{np \times q}$, $\beta_{hj} \in \mathbf{R}^{1 \times q}$. 并且 W_{hj} , β_{hj} 以随机的方式生成, 分别代表用于产生第 j 组特征结点的权重和偏差. H_j 表示第 j 组生成的特征结点. 我们将所有的 m 组特征结点拼接在一起得到最终的特征结点集合 H_m , 其中 $H_m = [H_1, \dots, H_m]$.

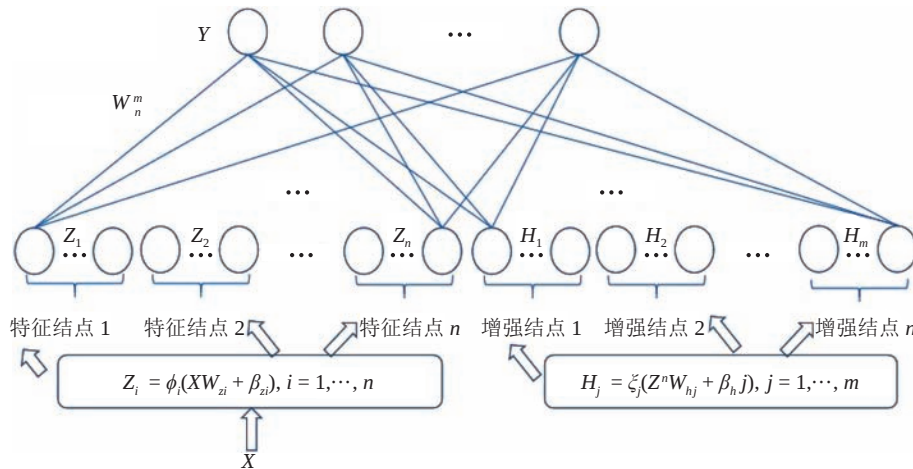


图 1 一种典型的宽度学习系统网络结构

Fig.1 A typical network structure of broad learning system (BLS)

在生成增强节点之后, 我们还需要将特征结点和增强结点拼接在一起作为输入层, 经过进一步计算得到输出的结果. 我们设连接输入层与输出层之间的权重为 W_n^m . 由输入层映射到输出层的过程如式 (3) 所示.

$$Y = [Z^n, H^m] W_n^m \quad (3)$$

$$W_n^m = [Z^n, H^m]^+ Y \quad (4)$$

其中, $W_n^m \in \mathbf{R}^{(np+mq) \times C}$. 我们可以根据式 (4) 来计算权重 W_n^m . $[Z^n, H^m]$ 的伪逆 $[Z^n, H^m]^+$ 可以通过岭回归算法计算得到.

1.1.2 应用研究

自从宽度学习系统被提出来以后, 人们又进一步做了许多相关的研究. 宽度学习系统和深度学习方法相比需要训练的参数较少, 网络结构相对简单, 所需要的时间开销也相对较小, 能够更好地适应实际应用的需求. 近年来在宽度学习系统的应用研究领域, 人们也取得了许多新的成果.

Peng 等^[2] 提出将宽度学习网络应用于网络流量分析领域. 他们设计了一种基于雾计算的架构并在树莓派上进行了实验. 实验结果表明采用宽度学习能够在准确识别网络流量数据的情况下进一步提高训练速度. Zhang 等^[3] 将宽度学习方法应用于人脸面部表情识别领域. 他们在拓展 Cohn-Kanad (CK+) 数据集上进行了实验, 证明了宽度学习方法在面部表情识别领域的有效性. Gao 等^[4] 提出了一种基于事件的目标检测方法. 他们将宽度学习网络应用于提高事件摄像机的在线训练及推理能力, 并设计了一个梯度下降方法来训练相关的网络参数. Liu 等^[5] 将宽度学习方法应用于半监督分类问题领域. 他们设计了一种流形正则化方法来探索潜在的数据分布, 从而协助提高分类准确度. 这种方法适合于样本数据随时间动态更新的应用领域. 与相关方法的对比实验结果表明 Liu 等提出的方法能够有效提高分类精度, 同时减少学习时间开销和存储空间消耗. Wang 等^[6] 将宽度学习系统应用于脑电图情绪识别领域. 他们将宽度学习系统与图卷积神经网络结合起来, 提出了一种新的脑电信号处理系统, 即宽度动态图学习系统. 实验结果表明该方法能够有效提高情绪识别的准确度. Chu 等^[7] 提出了一种加权宽度学习网络, 用于处理工业过程中的噪声和异常值. 他们对正常样本和异常样本分别赋予不同的权重来调整不同样本对建模的影响. 来自公共数据集和实际应用的实验结果表明这种方法具有较好的泛化性和鲁棒性.

1.2 孪生网络

孪生网络的概念由 Bromley 等^[8] 提出, 并被使用于签名验证领域, 用来验证两个签名笔迹是不是同一个人所书写的. 随着人们研究的深入, 孪生网络的应用范围也得到了进一步的拓广. 例如, 人们可以使用孪生网络来比较文本之间的语义相似度, 或者进行视觉跟踪任务. 虽然孪生网络的应用越来越多样化, 但是这些应用的基本思想是一致的. 那就是通过子网络将输入样本的特征信息映射为新的输出向量, 并且通过比较输出向量之间的相似程度来判断输入样本之间的近似程度. 在构建孪生网络的时候, 人们可以使用相同类型的子网络, 也可以根据实际应用的需求使用不同类型的子网络来构建孪生网络. 子网络之间可以共享权重, 也可以根据实际情况采用不同的权重. 这些子网络可以被视为是用于处理输入样本信息的映射模型, 负责将样本特征映射为指定维度的输出向量. 而训练的目标则是希望同类型样本的输出向量差异尽可能小, 不同类型样本的输出向量差异尽可能大. 近年来人们又对孪生网络的方法做了进一步的研究, 例如通过采用新的距离度量方法来比较输出向量之间的相似性, 或者将正负样本与基准样本同时输入来提高识别的效果. 为了更好地介绍孪生网络, 在方法介绍部分, 我们选择了其中一种典型的孪生网络结构来进行介绍.

1.2.1 方法介绍

我们选择了 Nair 等^[9] 提出的网络结构来进行介绍. 他们将孪生网络应用于人脸验证领域, 为后续人们将孪生网络应用于解决实际问题提供了更多的参考. 同时他们提出的孪生网络结构也具有一定的代表性, 能够反映孪生网络方法的基本思想. 在本节内容中, 我们将孪生网络分为特征提取, 相似度度量, 决策判别三个部分进行介绍.

在特征提取过程中, 人们需要将输入样本特征映射为指定的输出向量. 这个过程可以视为是一个降维过程, 将相对复杂的样本数据进行简化, 从而有利于后续对样本信息进行进一步处理. 特征提取过程的另一个作用在于抽取有效的样本信息, 通过训练使得输出向量能够更好地反映样本的类别信息, 从而提高最终识别的准确程度. 为了完成特征提取的任务, 人们可以采用卷积神经网络, 长短时记忆网络作为子网络来进行训练. 在原始版本的孪生网络结构中, 人们倾向于使用两个相同类型的网络作为子网络来进行训练. 子网络之间通常共享权重信息, 这样能够使得同一类别的样本生成的输出向量尽可能接近, 而不同类别样本生成的输出向量尽可能有所差异. 当然, 在实际应用中, 人们根据实际情况的需要也可以采用不同类型的网络作为子网络来进行

训练. 同时人们也对共享权值的策略进行了研究, 在一些特殊情况下不共享权重, 而是为不同子网络赋予不同的权重, 可以进一步提高算法的性能^[10].

相似度度量用于比较不同样本输出向量之间的相似程度, 从而为最终的决策判别提供依据. 由于输入样本特征已经被转化为指定维度的输出向量, 我们可以根据实际情况的需要采用不同的度量方法来进行衡量, 例如采用欧几里得距离或者余弦距离. 相似度度量的结果也将作为子网络进一步训练的依据, 用于进一步优化网络的结构.

决策判别则是根据相似度度量的结构生成最后的结论. 我们可以通过设置阈值的方式来对相似度度量的结果来进行划分, 也可以将相似度度量的结果转化为判别概率, 代表判别结果属于某一个类别的概率. 图 2 展示了一种典型的孪生网络结构, 其中的子网络采用相同的网络结构, 并且共享了权重信息.

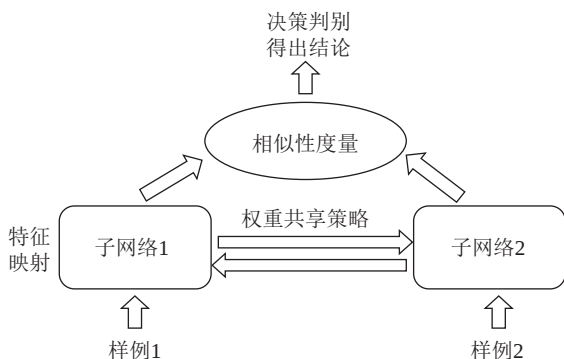


图 2 一种典型的孪生网络结构

Fig.2 A typical network structure of Siamese network

1.2.2 应用研究

孪生网络在自然语言处理和图像视觉领域都有着广泛的应用, 一些典型的例子列举如下. Baraldi 等^[11]设计了一种应用于广播视频自动场景分割的深度孪生网络算法. 他们提出了一个新的性能度量方法来进行进一步提高场景分割的准确程度. 作者同时将视觉与文本数据进行了融合. 与该领域相关算法的对比实验进一步证明了这种算法的有效性. Melekhov 等^[12]将孪生网络应用于图像匹配领域. 他们的成果可以被进一步应用于图像检索、三维重建等计算机视觉应用领域. Bertinetto 等^[13]提出了一种全卷积孪生网络算法用来进行视频目标追踪. 他们将全卷积神经网络用于构建孪生网络. 并且提出了一种新的离线学习策略. 实验结果表明新提出的方法能够有效地提高目标追踪的准确率. Zeghidour 等^[14]将孪生网络应用于比较说话人和语音之间的相似

性. 他们设计了一种联合嵌入架构来提高将说话人身份信息从语音数据中提取出来的准确度. 实验结果表明这种方法在类似的弱监督方法的基础上有了进一步的提高. Neculoiu 等^[15]研究了如何比较文本内容相似性的问题. 他们采用长短时记忆神经网络来搭建孪生网络. 实验结果表明这种方法能够有效地识别可变长字符串之间的相似性. Rahul 等^[16]将孪生网络应用于水下环境的目标追踪问题. 他们考虑了水下环境光照变化、背景干扰对目标追踪的影响. 有关的实验结果表明作者提出的方法能够更好地克服周围环境的干扰, 进一步提高目标追踪的准确性.

2 宽度孪生网络

在本节的内容中, 我们将对本文提出的宽度孪生网络算法进行介绍. 我们展示了算法的设计思路, 并将有关的内容将分为特征映射, 相似度度量和决策判别三个部分展开描述.

2.1 特征映射

传统的孪生网络算法在训练的时候需要构造成对的相似数据和不相似数据进行训练, 使得相似样本的输出向量尽可能接近, 不相似样本的输出向量差异尽可能明显. 当训练样本规模增大的时候, 所需要构造的相似对和非相似对数据规模也会显著增大. 如式 (5) 所示, 在完全图中 (即任意两个样本之间都用一条边来描述它们之间的相似关系), 样本数量 $nSam$ 和构造所有相似对信息的时间复杂度 $T(nSam)$ 之间满足以下关系:

$$T(nSam) = nSam(nSam - 1)/2 = O(nSam^2) \quad (5)$$

其中, 样本数量 $nSam$ 和构造所有相似对的时间复杂度 $T(nSam)$ 之间近似满足平方关系. 由于实际应用中的样本数据通常没有直接标注相似对信息, 因此我们往往在训练之前需要人为先构造相似对信息用于训练. 当样本数量 $nSam$ 较大的时候, 如果为任意两个样本数据构造相似对信息进行训练, 所需要的时间开销和计算资源开销将会明显增加. 这与边缘计算应用要求降低处理延迟, 并且满足终端设备资源受限情况的要求是不相匹配的. 虽然可以通过采样的方式, 只选取部分样本来构造相似对信息进行训练. 但这种方法在训练的时候会丢失部分样本信息, 从而影响最终训练的效果. 显然, 选择怎样的采样策略进行训练将会影响算法最终的性能表现. 人们需要在面对实际问题的时候额外考虑选取合适的采样策略来进行训练. 为了解决这一问题, 我们提

出了一种特征映射策略来进行训练.

我们首先分析了传统孪生网络方法在训练的时候需要加入相似对信息训练的原因. 通常用孪生网络进行训练的时候, 子网络只是提取了样本数据的一些特征用于进一步的处理判断. 这些特征并不直接与类别信息相关. 以面部表情识别为例, 子网络所提取的特征可能只是面部某个器官的相关数据, 需要我们借助相似对信息来辅助调整子网络的有关参数, 使得同类别样本数据提取的特征差异尽可能小, 不同类别样本数据提取的特征差异尽可能明显.

为了减少对相似对信息的依赖, 使得在没有构造相似对信息的情况下也能够对孪生网络进行训练. 我们提出在映射特征的时候加入类别信息, 通过采用适当的特征映射策略, 使得同一类别样本映射得到的特征尽可能接近, 而不同类别样本映射得到的特征差异尽可能明显.

例如, 我们可以采用独热 (One-hot) 编码的方式来构造映射后的输出特征向量. 在独热编码中, 假设样本数据的类别数量为 C , 则每一个样本在编码后将会得到一个长度为 C 的输出向量 T , T 中的每一位代表样本属于对应类别的概率. 在理想情况下, 我们在训练的时候希望向量 T 只有一位的数值为 1, 代表样本属于对应类别的概率为 1; 向量 T 中其他位置的数值为 0, 代表样本属于其他类别的概率为 0. 如图 3(a) 所示, 假设两个样本为 T_1 和 T_2 , 它们之间的欧几里得距离为 Eu , T_1 和 T_2 的长度为 tn , 并且用 $T_1(i)$ 表示 T_1 中的第 i 个元素 $i = 1, 2, \dots, tn$. 通过计算我们可以得到, 当两个样本属于同个类别的时候, $Eu=0$, 当两个样本不属于同个类别的时候, $Eu=1$, 满足孪生网络希望能类别样本的输出向量之间差异尽可能小, 不同样本输出向量之间差异尽可能大的要求. 如果在相似度度量阶段我们采用其他的度量指标, 例如曼哈顿距离, 我们依然能够得到相似的结论, 我们用 Ma 来表示两个样本 T_1 和 T_2 之间的曼哈顿距离. 有关的说明过程如图 3(b) 所示.

$$\begin{aligned} &\text{样本1: } 000 \cdots 1 \cdots 00 \quad i=j: \\ &\quad \text{第 } i \text{ 位为 } 1 \quad Eu = \sqrt{\sum_{i=1}^m (T_1(i) - T_2(i))^2} = 0 \\ &\text{样本2: } 000 \cdots 1 \cdots 00 \quad i \neq j: \\ &\quad \text{第 } j \text{ 位为 } 1 \quad Eu = \sqrt{\sum_{i=1}^m (T_1(i) - T_2(i))^2} = 0 \end{aligned}$$

(a) 欧几里得距离
(a) Euclidean distance metric

$$\begin{aligned} &\text{样本1: } 000 \cdots 1 \cdots 00 \quad i=j: \\ &\quad \text{第 } i \text{ 位为 } 1 \quad Ma = \sum_{i=1}^m |T_1(i) - T_2(i)| = 0 \\ &\text{样本2: } 000 \cdots 1 \cdots 00 \quad i \neq j: \\ &\quad \text{第 } j \text{ 位为 } 1 \quad Ma = \sum_{i=1}^m |T_1(i) - T_2(i)| = 0 \end{aligned}$$

(b) 曼哈顿距离
(b) Manhattan distance metric

图 3 采用独热编码的相似性度量

Fig.3 Similarity metrics of one-hot coding

上述特征映射的过程虽然直观上来看是包含了样本的类别信息, 但其实这些类别信息也是在样本特征向量的基础上经过模型处理才得出的, 因此特征映射的结果也能够一定程度上反映样本特征信息的特点. 如图 4 所示, 在图 4 中, 样本 1 和样本 2 的独热编码中数值最高位是不相同的, 但是它们在不同位置上的总体数值分布却有相似的地方, 这说明样本 1 和样本 2 的特征分布有相似的地方, 在一定程度上也反映出了样本的特征信息. 在进行相似度计算的时候, 我们不仅仅只比较数值最高位的情况, 还应该考虑其他位置上的数值. 在本文中, 为了简化相似度计算过程以及方便说明, 我们只考虑了映射结果取值最高两位的情况, 具体的相似度计算过程在下一节中将进行计算. 未来我们还可以对上述相似度计算方案做进一步的改进, 例如综合地考虑更多位置上的数值, 以进一步提高分类的准确度. 当然在特征映射的时候, 我们可以对映射的策略做进一步的改进. 例如在独热编码的基础上加入更

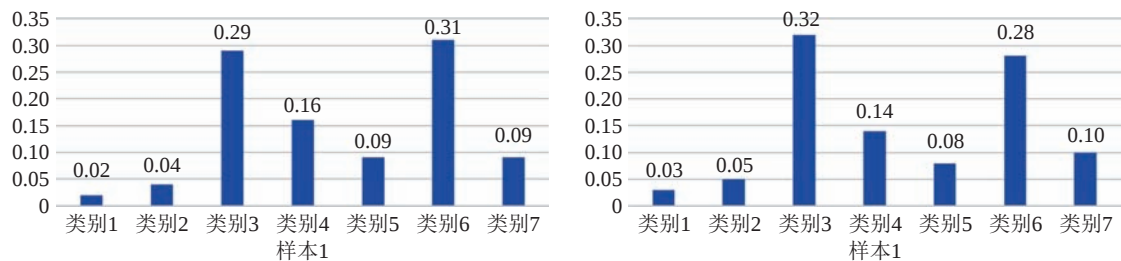


图 4 特征映射结果分析

Fig.4 An analysis of feature mapping results

多的特征信息, 包括与表情相关的人脸活动单元特征, 以进一步丰富输出向量所代表的内容. 一种可能的结合独热编码与额外样本特征信息的混合特征映射方案如图 5 所示.

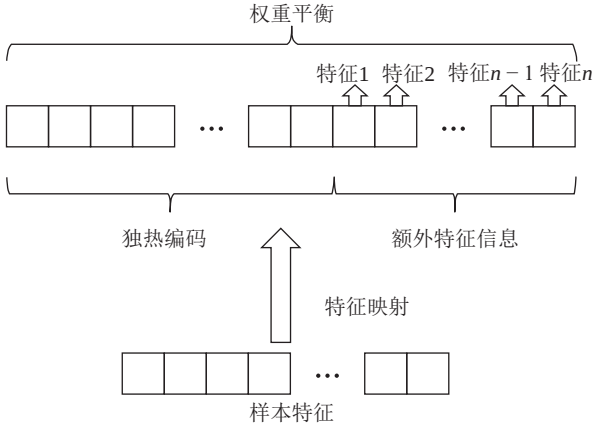


图 5 一种可能的混合特征映射方案

Fig.5 A possible hybrid feature mapping scheme

在图 4 中, 我们可以为不同位的编码赋予不同的权重, 以进一步提高混合映射编码对样本信息的表达能力. 在本文的实验中, 为了方便进行说明展示, 我们采用了最简明的独热编码方案作为输出向量的映射策略. 我们将在后续的实验研究内容中, 进一步介绍有关的实验结果.

在对输出向量的映射策略进行介绍之后, 如图 6 所示, 我们展示了宽度孪生网络的特征映射过程. 我们采用宽度学习系统来构建孪生网络算法中的子网络, 并且采用了共享权重的策略来调整来调整子网络的权重信息. 采用共享权重的策略可以进一步降

低训练时所需要的时间和存储空间开销, 从而进一步满足边缘计算应用的需求.

在图 5 中, X 表示样本特征数据, Y 为样本标签. 参数 n 表示特征结点的组数, p 表示每组特征结点有 p 个结点, e 表示增强节点的数量. 在生成特征结点的时候, 由输入样本特征根据随机设定的权重和偏差值进行映射, 得到 n 组特征结点 $Z_1, \dots, Z_n$. 将所有特征结点拼接在一起得到最终的特征结点 $Z^n, Z^n = [Z_1, \dots, Z_n]$. 在生成增强结点的时候, 我们将 Z^n 作为一个整体, 同样通过随机设定的权重和偏差值进行映射, 得到增强结点集合 H . 我们将特征结点和增强结点拼接在一起作为输入层 $[Z^n, H]$. 随后我们定义连接输出向量 Y 和输入层 $[Z^n, H]$ 之间的权重矩阵为 W_{mn} , 其中 Y 采用独热编码方式进行编码. 值得注意的是, 连接输入样本与特征结点之间的系数矩阵是随机设定的. 而连接特征结点和增强结点之间的系数矩阵是正交规范化后生成的随机矩阵. 连接输入层和输出向量之间的只有权重矩阵 W_{mn} , 不设置偏差值. 在训练的时候, 我们可以通过求伪逆的方法来求解 W_{mn} . 已知 $Y = [Z^n, H^m]W_{mn}$, 可得 $W_{mn} = [Z^n, H^m]^+ Y$. 其中, $[Z^n, H^m]^+$ 是输入层 $[Z^n, H^m]$ 的伪逆, 可以通过岭回归的方法得到^[9].

2.2 相似度度量

在相似度度量阶段, 我们需要对特征映射阶段得到的输出向量进行度量, 以进一步得到它们之间的相似程度. 我们首先对输出向量的结构进行分析, 以进一步确定适合的相似度度量方案. 如图 7 所示, 在独热编码中, 输出向量 T 中的每一位可以代表样本

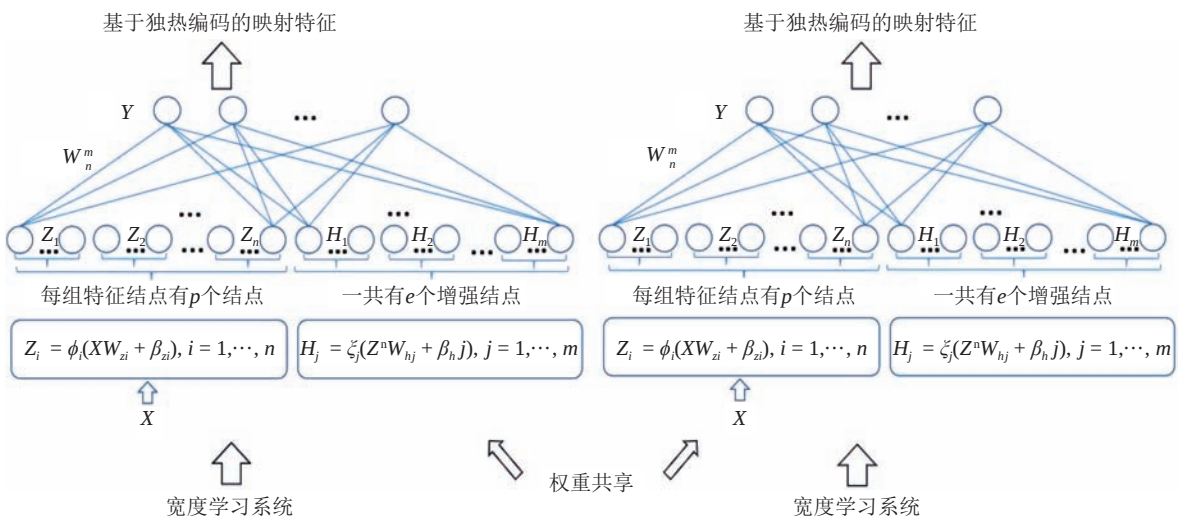


图 6 宽度孪生网络特征映射过程

Fig.6 Feature mapping of broad Siamese network

属于该类别的概率. 在通常情况下, 我们只用取值最高的一位来代表样本的类别信息. 然而如图 1 所描述, 编码中其他位置上的数值也在一定程度上包含了样本的信息. 在进行相似度计算的时候我们应该适当考虑其他位置上的数值, 以便得到更加准确的结果. 在常用的欧几里得距离公式中, 我们需要考虑每一位的数值对相似度度量结果的影响, 这在一定程度上影响了最终相似度的判定结果. 但这种做法容易受到极端值的影响. 考虑到编码中数值较高的几位往往与样本的关系更加密切, 我们设计了一种新的相似度度量方法. 在这种方法中, 我们只比较输出向量 T 中取值较高的几位数值来得到最终的相似度度量结果. 我们首先假设需要比较相似度的两个输出向量分别为输出向量 T_1 和 T_2 , 并且假设输出向量 T_1 和 T_2 中取值最大的位置分别为第 M_1 和 M_2 位, 对应位置的取值分别为 $T_1(M_1)$, $T_2(M_2)$. 如果 $M_1 = M_2$, 我们规定最终的相似度度量结果

$$D = \sqrt{(T_1(M_1) - T_2(M_2))^2} \quad (6)$$

如果 $M_1 \neq M_2$, 我们规定

$$D = \sqrt{(T_1(M_1) - T_2(M_1))^2 + (T_1(M_2) - T_2(M_2))^2} \quad (7)$$

得到的 D 即是输出向量 T_1 和 T_2 之间的相似度, 用于决策判别阶段得到最终的判别结果.

2.3 决策判别

在决策判别阶段, 我们采用设定阈值的方式来得到最终的判别结果. 对于输出向量 T_1 和 T_2 之间的相似度 D , 我们设定判别阈值为 ϕ . 如果 $D > \phi$, 我们认为输出向 T_1 和 T_2 对应的样本属于不同的类别. 如果 $D \leq \phi$, 我们认为输出向量 T_1 和 T_2 对应的样本属于相同的类别. 我们研究了阈值 ϕ 的设置对最终判别结果准确度的影响, 并绘制了相应的 ROC 曲线图像. 有关的实验结果将在实验研究部分进行进一步的说明.

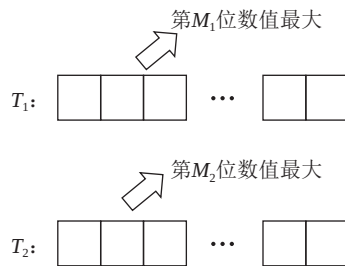


图 7 一种基于独热编码的相似度度量方案

Fig.7 An one-hot based similarity metric

3 实验研究

在本节中, 我们将对有关实验研究的内容进行介绍. 我们将首先介绍实验的总体情况, 随后展示对比实验的结果. 与此同时, 我们也研究了相似性度量指标对宽度孪生网络算法识别准确度的影响. 最后, 我们对实验结果进行了总结分析.

3.1 实验介绍

在本次实验中, 我们采用了 CK+, MNIST 等常见分类数据集来进行有关的实验. 有关数据集信息如表 1 所示. 其中, CK+, JAFFE 数据集是常见的人脸表情分类数据集. MNIST、USPS 是有关于手写体数字识别的数据集. 我们在不同的数据集中采用了不同的参数设置. 宽度孪生网络算法在不同数据集上的参数设置信息如表 2 所示. 为了证明我们提出的宽度孪生网络算法能够在保证分类准确度的同时降低训练时间, 减少内存等计算资源开销. 我们在对比实验中将宽度孪生网络算法与基于深度学习的孪生网络算法进行了比较, 比较的指标包括了分类准确度, 训练时间和内存开销. 与此同时, 我们也研究了相似度度量指标对宽度孪生网络算法性能的影响. 我们在一台配置了 16 GB 内存以及 Intel i7 1.61 GHz 中央处理器的笔记本电脑上进行了上述的实验内容. 最后, 我们对有关的实验内容进行了分析总结.

3.2 对比实验结果

我们在 CK+, JAFFE、MNIST、USPS 等分类数据集上进行了对比实验, 比较的指标包括了分类准确度, 训练时间和内存开销. 考虑到在基于深度学习的孪生网络算法中, 随着网络结构复杂度的加深, 算法的训练时间和内存开销会明显增加. 为了更好地证明我们提出的宽度孪生网络算法能够有效地减少训练时间和内存开销, 我们使用结构相对简单的三层全连接神经网络来构建孪生网络, 并进行了有关的对比实验. 这个全连接网络每一层结点个数的

$M_1 = M_2$, 相似度距离 D 为:

$$D = \sqrt{(T_1(M_1) - T_2(M_2))^2}$$

$M_1 \neq M_2$, 相似度距离 D 为:

$$D = \sqrt{(T_1(M_1) - T_2(M_1))^2 + (T_1(M_2) - T_2(M_2))^2}$$

表 1 实验数据集信息表
Table 1 Table of data set for experiments

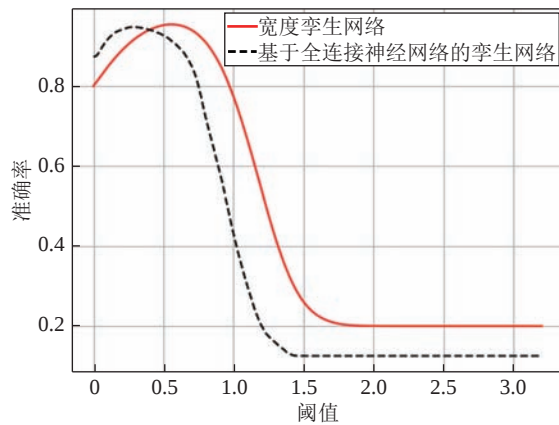
数据集	样本规模	类别数量	各类别样本数量	特征维度
CK+	5876	7	(1022, 233, 868, 546, 1331, 547, 1329)	14400
MNIST	70000	10	每个类别近似 7000 样本	784
JAFPE	213	7	(30, 29, 32, 31, 30, 31, 30)	14400
USPS	20000	10	每个类别 2000 样本	784

表 2 宽度孪生网络参数设置信息
Table 2 Table of parameters for broad Siamese network

数据集	n	p	e
CK+	8	10	9000
MNIST	10	10	500
JAFPE	8	10	9000
USPS	10	10	1500

设置如表 3 所示.

在准确度方面, 我们比较了算法最终得到的准



(a) USPS 数据集
(a) USPS data set

表 3 对比算法中全连接神经网络结点个数设置信息
Table 3 Table of number about nodes in the fully connected network for comparison

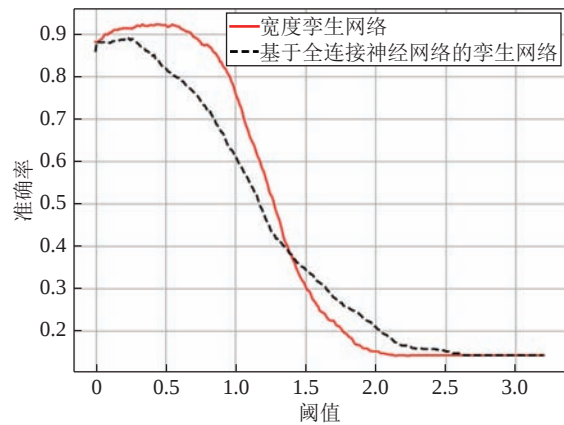
数据集	第一层结点数	第二层结点数	第三层结点数
CK+	512	128	512
MNIST	16	16	16
JAFPE	1024	128	1024
USPS	128	128	128

确率 (如表 4 所示), 并绘制比较了算法准确度随阈值变化的曲线 (图 8) 以及受试者工作特征 (ROC) 曲线 (图 9). 有关训练时间和内存开销的实验结果如表 5 和表 6 所示.

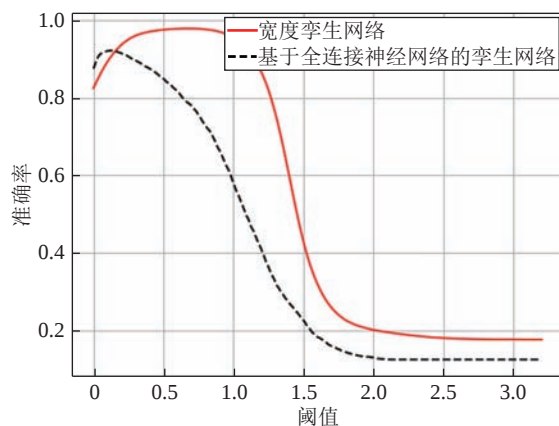
为了进一步地探索相似性度量指标对宽度孪生网络算法性能的影响, 我们在接下来的内容中进一步设计了有关的研究实验, 并给出了相应的实验结果.

3.3 相似性度量指标分析

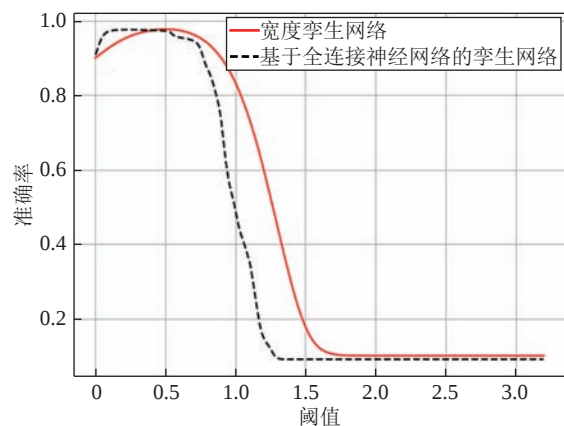
在相似性度量指标的有关研究实验中, 我们比



(b) JAFPE 数据集
(b) JAFPE data set



(c) CK+ 数据集
(c) CK+ data set



(d) MNIST 数据集
(d) MNIST data set

图 8 算法准确度随阈值变化的曲线

Fig.8 Threshold curves of algorithms

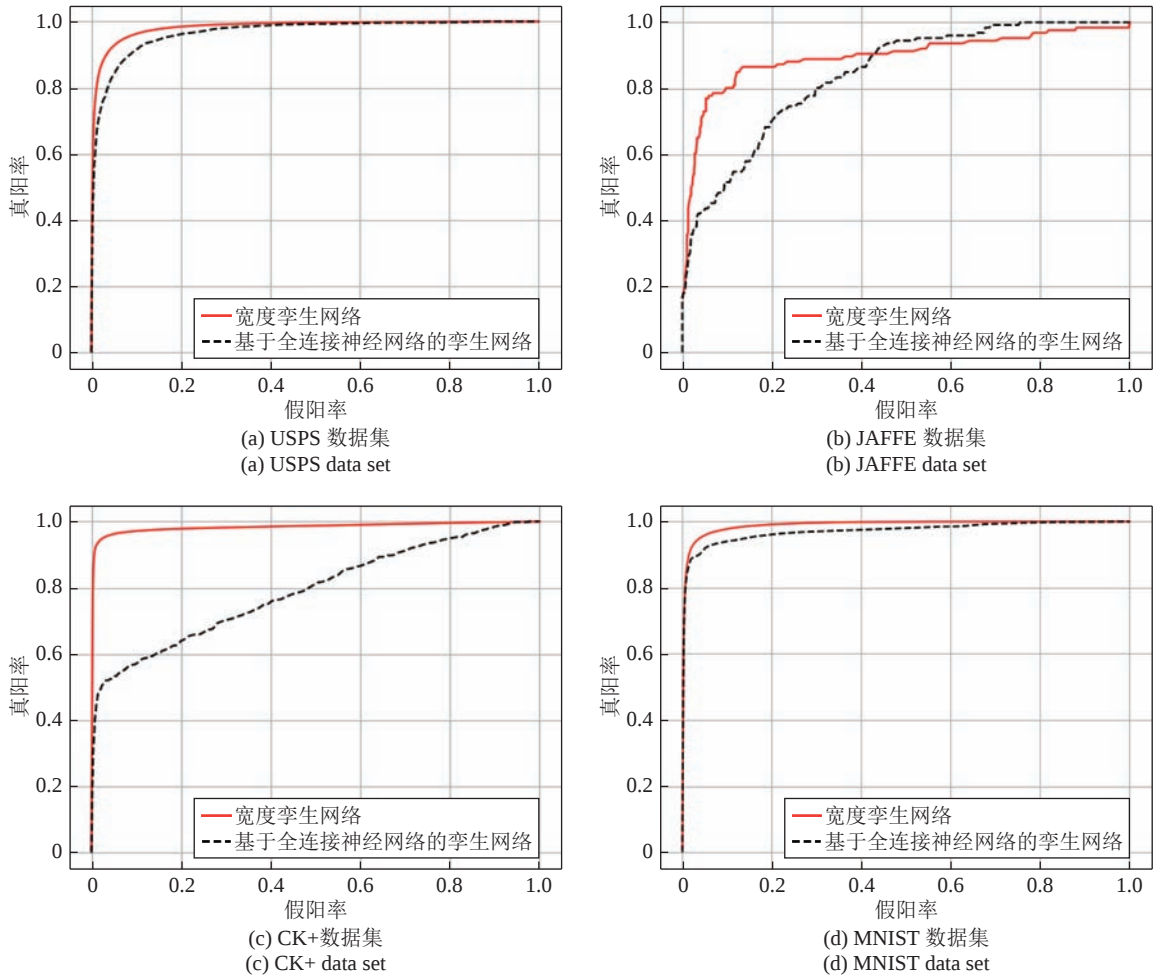


图 9 受试者工作特征 (ROC) 曲线

Fig.9 Receiver operating characteristic (ROC) curves of algorithms

较了欧几里得距离、曼哈顿距离等度量指标与本文提出的相似度度量方法对宽度孪生网络性能的影响。

表 4 准确率实验结果

Table 4 Table of experiment results about accuracy

数据集	宽度孪生网络	基于全连接神经网络的孪生网络
CK+	0.9788738	0.9287094
MNIST	0.9798112	0.9777414
JAFFE	0.9217687	0.9206349
USPS	0.9536075	0.9505025

表 5 训练时间实验结果

Table 5 Table of experiment results about training time

数据集	宽度孪生网络	基于全连接神经网络的孪生网络
CK+	94.140997	567.9896214
MNIST	5.6314652	60.0518959
JAFFE	58.4795067	1105.1385579
USPS	3.0834677	29.8392925

表 6 内存开销实验结果

Table 6 Table of experiment results about memory overhead

数据集	宽度孪生网络	基于全连接神经网络的孪生网络
CK+	3.3491211	6.0307884
MNIST	2.4598732	2.0554810
JAFFE	0.2893066	5.6162262
USPS	1.2569504	0.8804893

响。我们按照实验结果绘制了准确率随阈值变化的曲线 (如图 10 所示) 以及受试者工作特征曲线 (如图 11 所示)。

通过观察图 10 和图 11 中的实验结果, 我们可以看到本文新提出的度量方法与传统的欧几里得距离、曼哈顿距离等度量方法相比, 能够有效地进一步提高宽度孪生网络的分类准确度, 具有进一步应用于解决实际问题的潜力。

3.4 结果分析

通过前面的实验内容, 我们可以看到宽度孪生

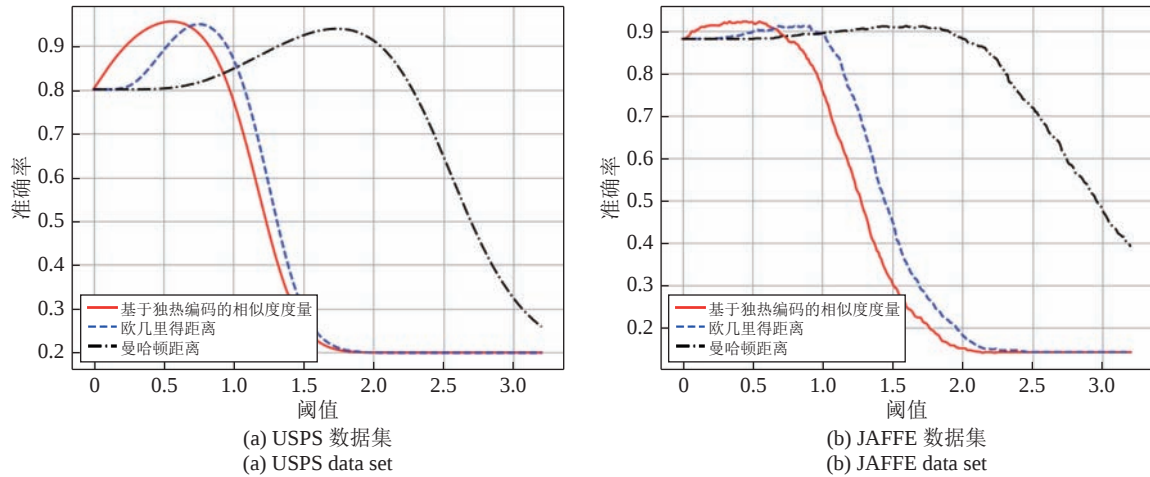


图 10 采用不同相似性度量指标宽度孪生网络算法准确度随阈值变化的曲线

Fig.10 Threshold curves of broad Siamese network with different similarity metrics

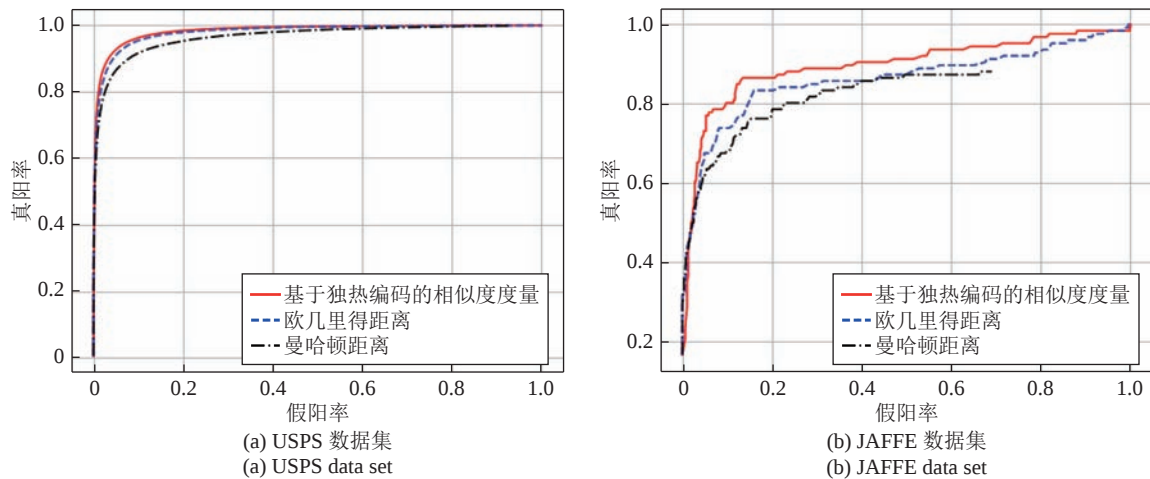


图 11 采用不同相似性度量指标宽度孪生王洛算法的受试者工作特征 (ROC) 曲线

Fig.11 Receiver operating characteristic (ROC) curves of broad Siamese network with different similarity metrics

网络算法相比于基于传统深度学习方法的孪生网络方法,能够在保证分类准确度的前提下降低训练时间和内存资源开销,具有进一步应用于边缘计算应用的潜力.与此同时,我们研究了相似性度量指标对宽度孪生网络算法分类准确度的影响,并且证明了我们提出的相似性度量方法能够进一步提高宽度孪生网络算法的分类准确度,未来可以再做进一步改进以提高算法的分类准确度.由于大多数的边缘计算应用要求减少相应延迟,降低计算资源开销,我们提出未来可以进一步研究以将基于浅层网络的方法应用于提高边缘计算应用的性能.

4 结论

在本文中,我们提出了一种面向边缘计算应用的宽度孪生网络算法.我们考虑了边缘计算应用要

求降低处理延迟,满足边缘设备资源受限的情况,提出可以将宽度学习系统应用于边缘计算应用之中.为了验证本文观点,我们将宽度学习系统与孪生网络相结合,提出了宽度孪生网络算法,并应用于解决分类问题.与传统基于深度学习的孪生网络方法相比,宽度孪生网络算法能够在满足分类准确度要求的情况下,进一步降低训练时间以及内存资源开销.这说明借助宽度学习系统能够有效降低处理延迟,减少计算资源消耗,从而更好地满足边缘计算应用的要求.在文章的最后,我们对未来的研究方向进行了展望.我们可以进一步改进宽度孪生网络方法,用于解决更多的实际分类问题,并应用到边缘计算应用之中.例如,自动驾驶车辆上的交通标识识别以及监控视频中的对象追踪应用.本文算法可以做进一步的改进,以适应这些应用要求低延迟、低资源开

销的特点. 与此同时, 我们也可以进一步探索浅层网络方法如宽度学习系统在边缘计算领域的应用. 借助宽度学习系统等浅层网络方法进一步提高边缘计算应用的性能.

References

- Chen C L P, Liu Z. Broad learning system: an effective and efficient incremental learning system without the need for deep architecture. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(1): 10–24
- Peng X, Ota K, Dong M. A broad learning-driven network traffic analysis system based on fog computing paradigm. *China Communications*, 2020, 17(2): 1–13
- Zhang T, Liu Z, Wang X, Xing X, Chen C L P, Chen E. Facial expression recognition via broad learning system. In: Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan: IEEE, 2018. 1898–1902
- Gao S, Guo G, Huang H, Cheng X, Chen C L P. An end-to-end broad learning system for event-based object classification. *IEEE Access*, 2020, 8: 45974–45984
- Liu X, Qiu T, Chen C, Ning H, Chen N. An incremental broad learning approach for semi-supervised classification. In: Proceedings of the 2019 IEEE International conference on Dependable, Autonomic and Secure Computing, International conference on Pervasive Intelligence and Computing, International conference on Cloud and Big Data Computing, International conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCoM/CyberSciTech), Fukuoka, Japan: IEEE, 2019. 250–254
- Wang X, Zhang T, Xu X, Chen L, Xing X, Chen C L P. EEG emotion recognition using dynamical graph convolutional neural networks and broad learning system. In: Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain: 2018. 1240–1244
- Chu F, Liang T, Chen C L P, Wang X, Ma X. Weighted broad learning system and its application in nonlinear industrial process modeling. *IEEE Transactions on Neural Networks and Learning Systems*. 2020, 31(8): 3017–3031
- Bromley J, Guyon I, Lecun Y, et al. Signature verification using a Siamese time delay neural network. In: Proceedings of the Advances in Neural Information Processing Systems 6, 7th NIPS Conference, Denver, Colorado, USA: Morgan Kaufmann Publishers Inc., 1993.
- Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10). Omnipress, Madison, WI, USA: 2019. 807–814
- Treible W, Saponaro P, Kambhamettu C. Wildcat: in-the-wild color-and-thermal patch comparison with deep residual pseudo-Siamese networks. In: Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, China: IEEE, 2019. 1307–1311
- Baraldi L, Grana C, Cucchiara R. A deep Siamese network for scene detection in broadcast videos. In: Proceedings of the 23rd ACM International Conference on Multimedia (MM'15). Association for Computing Machinery, New York, NY, USA: 1199–1202
- Melekhov I, Kannala J, Rahtu E. Siamese network features for image matching. In: Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico: 2016. 378–383
- Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking. In: Proceedings of European Conference on Computer Vision (ECCV) Workshops. 2016. 850–865
- Zeghidour N, Synnaeve G, Usunier N, et al. Joint learning of speaker and phonetic similarities with siamese networks. In: Proceedings of Interspeech 2016, 2016. 1295–1299
- Neculoiu P, Versteegh M, Rotaru M. Learning text similarity with Siamese recurrent networks. In: Proceedings of the 1st Workshop on Representation Learning for NLP. 2016. 148–157
- Rahul M V, Ambareesh R, Shobha G. Siamese network for underwater multiple object tracking. In: Proceedings of the 9th International Conference on Machine Learning and Computing (ICMLC 2017). New York, NY, USA: ACM, 2017. 511–516



李逸楷 华南理工大学计算机科学与技术专业博士研究生. 2020 年获得中国广州华南理工大学计算机科学与技术专业学士学位. 主要研究方向为神经网络, 情感计算等机器学习技术及其在边缘计算领域的应用.

(**LI Yi-Kai** Ph. D. candidate in computer science and technology at South China University of Technology. He received his bachelor degree in computer science and technology from South China University of Technology, Guangzhou, China in 2020. His research interest covers machine learning techniques such as neural networks, affective computing and their applications in the field of edge computing.)



张 通 华南理工大学计算机科学与工程学院副教授. 2009 年获得广州中山大学软件工程专业学士学位, 2011 年获得澳门大学应用数学专业硕士学位, 2016 年获得澳门大学软件工程专业博士学位. 主要研究方向为情感计算, 进化计算, 神经网络和其他机器学习技术及其应用. 本文通信作者.

(**ZHANG Tong** Associate professor at the School of Computer Science and Engineering, South China University of Technology, China. He received his bachelor degree in software engineering from Sun Yat-sen University, Guangzhou, China in 2009, and his master degree in applied mathematics from University of Macau, Macau, China in 2011, and the Ph. D. degree in software engineering from the University of Macau, Macau, China in 2016. His research interest covers affective computing, evolutionary computation, neural network, and other machine learning techniques and their applications. Corresponding author of this paper.)



陈俊龙 华南理工大学计算机科学与工程学院特聘讲席教授及院长. IEEE Fellow, 美国科学促进会 AAAS Fellow, IAPR Fellow, 国际系统及控制论科学院 IASCYS 院士, 香港工程师学会 Fellow, 中国自动化学会 Fellow. 欧洲科学院 (AE) 外籍院士、欧洲科

学与艺术学院 (EASA) 院士, 国际系统与控制论科学院 (IASCYS) 院士. 1985 年毕业于美国密歇根州安娜堡市的密歇根大学安娜堡分校, 2016 年获得由母校普渡大学 (1988 年获得博士学位) 颁发的杰出电气和计算机工程师奖. 作为美国工程技术教育认证委员会 (ABET) 的项目评估员, 成功地设计了澳门大学的工程和计算机科学课程, 并通过香港工程师学会 (HKIE) 获得了华盛顿/首尔协议

的认证. 这被认为是他作为澳门大学科技学院前任院长在工程/计算机科学教育方面的最大贡献. 他在 2018 年获 IEEE 系统人机控制论的最高学术奖——IEEE 诺伯特·维纳奖 (Norbert Wiener Award). 2018 和 2019 年连续两年入选科瑞唯安 (Clarivate) 全球高被引科学家. 在 2012 年至 2013 年, 他是 IEEE 系统、人和控制论协会主席, 是 IEEE 期刊《系统、人和控制论: 系统》(2014-2019) 主编. 目前, 他是 IEEE 控制论汇刊的主编, 以及 IEEE 模糊系统汇刊的副主编. 他于 2015 年至 2017 年任国际自动控制联合会 TC9.1 经济与商务系统主席. 主要研究方向为控制论、系统和计算智能.

(CHEN C. L. Philip Chair professor and dean of the College of Computer Science and Engineering, South China University of Technology. He is a Fellow of IEEE, AAAS, IAPR, CAA, and HKIE; a member of Academia Europaea (AE), European Academy of Sciences and Arts (EASA), and International Academy of Systems and Cybernetics Science (IASCYS). Being a Program Evaluator of the Accreditation Board of Engineering and Technology Education (ABET) in the U.S., for computer engineering, electrical engineering, and software engineering programs, he successfully architects the University of Macau's Engineering and

Computer Science programs receiving accreditations from Washington/Seoul Accord through Hong Kong Institute of Engineers (HKIE), of which is considered as his utmost contribution in engineering/computer science education for Macau as the former Dean of the Faculty of Science and Technology. He received IEEE Norbert Wiener Award in 2018 for his contribution in systems and cybernetics, and machine learnings. He is also a highly cited researcher by Clarivate Analytics in 2018 and 2019. Dr. Chen was a recipient of the 2016 Outstanding Electrical and Computer Engineers Award from his alma mater, Purdue University (in 1988), after he graduated from the University of Michigan at Ann Arbor, Ann Arbor, MI, USA in 1985. He was the IEEE Systems, Man, and Cybernetics Society President from 2012 to 2013, the Editor-in-Chief of the IEEE Transactions on Systems, Man, and Cybernetics: Systems (2014-2019), and currently, he is the Editor-in-Chief of the IEEE Transactions on Cybernetics, and an Associate Editor of the IEEE Transactions on Fuzzy Systems. He was the Chair of TC 9.1 Economic and Business Systems of International Federation of Automatic Control from 2015 to 2017. His research interest covers cybernetics, systems, and computational intelligence.)