



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331, CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目： 两级邻域采样的孪生网络在流形学习中的应用
作者： 徐承志， 万方
网络首发日期： 2020-10-21
引用格式： 徐承志， 万方. 两级邻域采样的孪生网络在流形学习中的应用. 计算机工程与应用. <https://kns.cnki.net/kcms/detail/11.2127.TP.20201021.1644.012.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

两级邻域采样的孪生网络在流形学习中的应用

徐承志, 万 方

湖北工业大学 计算机学院, 武汉 430068

摘 要: 流形学习是一类特殊的非线性求解问题, 即从高维采样数据中恢复低维流形结构, 以达到维数简约的目的, 是模式识别与数据可视化中的重要方法。流形学习存在许多基于局部线性假设的数值解法, 即显示地定义局部线性映射模型再进行全局优化, 这些方法对于流形的形状、采样的方式都比较敏感。另一种非线性求解工具, 神经网络, 因为不依赖于具体数学模型, 理论上具有较好的鲁棒性, 但是流形学习的特殊非线性, 使得传统神经网络很难达到满意的效果。针对上述问题, 该文改进了一种同质双通道神经网络——孪生网络, 并应用于流形学习。针对孪生网络的两条通道, 设计了三重结构, 即升维层、过滤层和降维层, 同时基于两级邻域的概念, 提出了包含正、负样本对的损失函数, 再经过“样本对”的训练, 实现了邻近数据的空间关系在降维后依然得以保持。通过将孪生网络用于仿真数据 (Swiss roll) 的降维, 并与传统方法进行比较, 发现孪生网络可以更真实的还原高维流形的内在结构。同时, 将孪生网络用于真实数据 (handwritten digits) 的二维可视化, 并与传统方法进行比较, 发现孪生网络聚类效果同样明显, 并且类别分布更为均匀, 边界更易辨识。

关键词: 流形学习; 孪生网络; 两级邻域; 样本对训练

文献标志码: A 中图分类号: TP391 doi: 10.3778/j.issn.1002-8331.2007-0035

徐承志, 万方. 两级邻域采样的孪生网络在流形学习中的应用. 计算机工程与应用

XU Chengzhi, WAN Fang. Application of Siamese Network with Two-Level Neighborhood Sampling in Manifold Learning. Computer Engineering and Applications

Application of Siamese Network with Two-Level Neighborhood Sampling in Manifold Learning

XU Chengzhi, WAN Fang

School of Computer Science, Hubei University of Technology, Wuhan 430068, China

Abstract: Manifold learning is a special kind of nonlinear problem that is to recover the low-dimensional manifold structure from high-dimensional sampled data to achieve the purpose of dimensionality reduction. It is an important method in pattern recognition and data visualization. There are many numerical methods for manifold learning based on local linear assumptions, that is, explicitly defining the local linear mapping model and then performing global optimization. These methods are sensitive to the shape of the manifold and the way of sampling. Another nonlinear tool, neural network, is theoretically robust because it does not rely on the specific mathematical models. However, due to the special nonlinearity of manifold learning, the traditional neural network is difficult to achieve satisfactory results. In order to solve these problems, this paper improves a homogeneous dual channel neural network, Siamese network, and applies it to manifold learning. For the two channels of Siamese network, a triple

基金项目: 湖北省教育厅科学技术研究计划项目(No.B2017048, No.B2017046); 湖北工业大学科研启动基金项目(No.BSQD12031)。

作者简介: 徐承志(1976—), 男, 博士, 讲师, CCF 会员, 研究领域为流形学习, 视觉 SLAM, E-mail: xcz911@hbut.edu.cn; 万方(1976—), 男, 博士, 讲师, 研究领域为计算机视觉。

structure is designed, namely dimension increasing layer, filtering layer and dimension reducing layer. At the same time, based on the concept of two-level neighborhood, a loss function including positive and negative sample pairs is proposed. After the training of "sample pair", the spatial relationship of adjacent data is maintained after dimension reduction. By using the Siamese network to reduce the dimension of the simulation data (Swiss roll), and compared with the traditional methods, it is found that the Siamese network can more truly restore the internal structure of high-dimensional manifold. At the same time, the Siamese network is used for the two-dimensional visualization of real data (handwritten digits), and compared with the traditional methods, it is found that the clustering effect of Siamese network is also obvious, and the classification distribution is more uniform, and the boundary is easier to identify.

Key words: manifold learning; siamese network; two-level neighborhood; sample pair training

1 引言

随着多媒体技术的发展和大数据技术的兴起,数据维度呈爆炸性增长,使得机器学习、图像处理、模式识别等研究领域的数据分析变得越来越困难。为消除上述问题中的维度灾难,降维操作成为数据预处理中的重要手段。流形学习(Manifold Learning)是一类特殊的降维问题^[1],自2000年科学杂志《Science》中被首次提出以来,已成为信息科学领域的研究热点^[2,3]。流形学习的观点认为:从局部看,数据具有欧氏空间的性质;从全局看,数据是嵌入在一个更高纬度的流形体上。流形学习要消除冗余的纬度,用较低纬度来表示数据(降维操作)。降维的依据是,局部欧氏空间的性质在低维空间中得以保留。保留欧氏空间性质的一个主要方式就是保持数据间的欧氏距离不变,或者说保持数据间的相对位置不变。

鉴于流形结构的高维嵌入特性,传统的线性降维方法,如PCA算法^[4]、LDA算法^[5]等,并不适合流形学习,而一些基于局部线性假设的非线性数值解法则能很好处理这类特殊的降维问题,如基于重建邻域线性关系的LLE算法^[6]、基于测地距离的Isomap算法^[7]、基于重建邻接图的LE算法^[8]、基于切空间的LTSA算法^[9]、以及它们的众多改进算法^[10]。这些算法的理论基础都是基于矩阵数值分析与优化的,但随着深度学习在人工智能领域的兴起,利用神经网络实现降维也成为一种有潜力的研究方向。目前,具有降维效果的神经网络有,用于聚

类和分类问题的自组织映射网络SOM^[11],用于词嵌入的跳字模型Skip-gram和连续词袋模型CBOW^[12],用于特征提取的有损压缩网络自编码器Autoencoder^[13],以及基于共享权值的孪生网络Siamese Network^[14]。上述神经网络通过不同的方式实现了传统领域内的降维需求,并取得了不错的效果。但是因为流形结构的特殊非线性,鲜有神经网络直接应用于流形学习,本文尝试将孪生神经网络与流形学习相结合,希望为流形学习提供一种新的解决方案。

2 孪生网络

传统的基于逻辑回归的分类网络,适用于存在大量标注样本的情况,但在某些情况下,比如类别数量过多或类别数量不定,且每个类别的样本数量又相对较少时,传统分类网络就不那么有效了,比如判定两幅图片是否属于同一类物品,或两张人脸是否属于同一个人等。为了解决上述问题,YannLecun等人^[15]提出了孪生网络模型,用于计算样本间的差异度。孪生网络将数据从样本空间映射到目标空间,同类数据在目标空间中的差异度较小,反之差异度较大。通过计算数据间的差异度,孪生网络可以判断数据是否属于同一类别。一般情况下,孪生网络的输出维度(目标空间维度)要小于输入维度(样本空间维度),所以从这种角度看,孪生网络实现了数据降维。如图1所示,典型的孪生网络包含两条通道,并且两条通道的网络结构相同、权值相同,故称谓孪生网络。

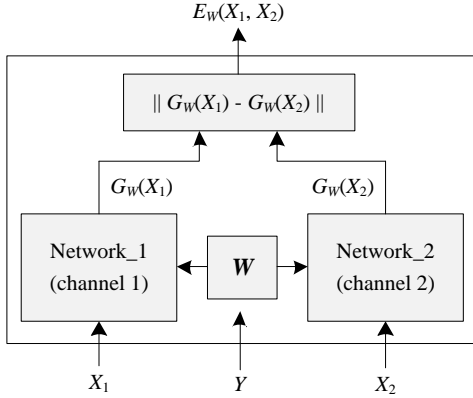


图 1 孪生网络
Fig.1 Siamese network

图 1 中的两个子网 Network_1 和 Network_2 即为两条通道，它们共享权值，即是两个完全相同的子网。两条通道各自接收样本空间中的一对数据 X_1 和 X_2 ，并分别映射为目标空间的 $G_W(X_1)$ 和 $G_W(X_2)$ 。网络输出的能量值 $E_W(X_1, X_2)$ 是一个非负数，等于 $\|G_W(X_1) - G_W(X_2)\|$ ，相当于两个数据在目标空间中的差异度。孪生网络的另一个输入 Y 代表 X_1 与 X_2 这对数据的标签，如果两个数据属于同类，则 Y 为 0，并希望 E_W 尽量小，否则 Y 为 1，并希望 E_W 尽量大，当两个数据完全相同时， E_W 自然为最小值 0。为了让同类型数据的差异度小，异类型数据的差异度大，孪生网络的损失函数 $Loss$ 定义如公式 (1) 和 (2) 所示：

$$L(W, (X_1^i, X_2^i, Y^i)) = (1 - Y^i)L_G(E_W(X_1^i, X_2^i)) + Y^i L_I(E_W(X_1^i, X_2^i)) \quad (1)$$

$$Loss(W) = \sum_{i=1}^n L(W, (X_1^i, X_2^i, Y^i)) \quad (2)$$

其中， n 是训练的样本总数， (X_1^i, X_2^i, Y^i) 是第 i 个数据对， L_G 负责计算同类样本的损失函数， L_I 负责计算异类样本的损失函数。为了达到同类样本的差异度小，异类样本的差异度大的效果，需要将 L_G 设计成单调递增函数， L_I 设计成单调递减函数。

孪生网络的通道结构可以根据问题的需要进行不同设计，比如针对回归问题可以选择多层感知机 (MLP)^[16]、针对图像问题可以选择卷积神经网络 (CNN)^[17]、针对时序问题又可以选择循环神经网络 (RNN)^[18] 等。同时，差异度的计算方法也可以灵活选择，比如余弦相似度、曼哈顿距离、欧氏

距离、汉明距离等。最后，损失函数也可以自由设计，只要保证优化方向符合预期目标就行，比如 YannLecun^[15] 就将 L_G 定义为的 E_W 二次项函数（递增函数），把 L_I 定义为的 E_W 的负指数幂函数（递减函数）。

3 孪生网络与流形学习

孪生网络可以实现从样本空间到目标空间的降维，但是孪生网络是有监督学习，需要样本标签来引导网络训练，而传统的流形学习是无监督学习，为了将孪生网络应用于流形学习，需要调整网络的使用方法。

3.1 损失函数与子网设计

如众多传统算法认为的，从局部看，流形是线性嵌入在高维空间中的低维结构，通过找出所有局部映射关系，并附加人为约束，最后整体优化，即可实现流形降维，如 LLE 算法设定邻域样本的线性组合为当前样本，并以 $\sum W_{ij}=1$ 为约束进行整合^[6]；LE 算法在降维过程中保持邻域内的邻接图关系，并以 $Y^T D Y = I$ 为约束进行整合^[8]；LTSA 是对局部空间进行 PCA 映射，并以 $T T^T = I$ 为约束进行整合^[9]；而 LMDS 则是对局部空间进行 MDS 映射，并以 $Y Y^T = I$ 为约束进行整合^[19]。理论上神经网络具有任意复杂的模式分类能力和优良的多维函数映射能力，如果神经网络能够蕴含足够多的局部映射关系，并且实现不同的输入样本激活不同的映射关系，就可以将神经网络引入行流形学习。

为了将孪生网络引入流形学习，需要重新设计孪生网络的损失函数。将样本间的空间关系用欧氏距离来表示，并希望映射到低维空间后，样本间尽量保持原空间里的欧氏距离。假设样本所在的原始空间是 R ，降维后的子空间是 P ，采用图 1 中的孪生网络结构，则样本在原始空间内的欧氏距离表示为 $d_R = \|X_1 - X_2\|$ ，在子空间内的欧氏距离表示为 $d_P = \|G_W(X_1) - G_W(X_2)\|$ 。网络训练的目标是希望 d_R 与 d_P 的差值尽量小，即降维后空间关系尽量保持，则可以设计网络的损失函数 $Loss_1$ ，如公式 (3)、(4)、(5) 所示：

$$d_R(X_1^i, X_2^i) = \|X_1^i - X_2^i\| \quad (3)$$

$$d_P(X_1^i, X_2^i) = \|G_W(X_1^i) - G_W(X_2^i)\| \quad (4)$$

$$Loss_1(W) = \sum_{i=1}^n |d_R(X_1^i, X_2^i) - d_P(X_1^i, X_2^i)| \quad (5)$$

上述网络训练思路与多维缩放算法 (MDS) 非常相似^[20], 不同的是, MDS 要保证所有数据点在低维空间中的距离等于在高维空间中的距离, 而流形学习只需要保证邻域内的样本在低维空间内的距离保持不变。因此在构建训练集时, 只需要选择邻域内数据来构建训练样本。确定邻域的方法有两种, 第一种是 k 邻近法^[21], 即选择距离当前样本最近的 k 个样本作为邻域, 第二种是 ε 邻域法^[22], 即选择与当前样本的欧氏距离在 ε 以内的样本作为邻域。传统流形学习算法通常选择第一种方法来创建邻域, 因为第一种方法能够产生固定数目的邻域样本, 非常适合进行矩阵运算与分析。本文则采用第二种方法, 训练集对邻域内样本数目没有限制。确定邻域产生方法后, 每个数据与其邻域数据形成一组样本对, 汇总所有数据的样本对, 并去掉重复后, 就形成了孪生网络的训练集。

后续工作是确定孪生网络中的通道结构, 即子网结构。孪生网络输入的是高维样本, 输出的是低维数据, 仅从某侧通道来看, 它是一个解决多元多值回归问题的模型, 对于简单应用可以直接选择多层感知机 (MLP) 作为子网结构, 如图 2 所示。

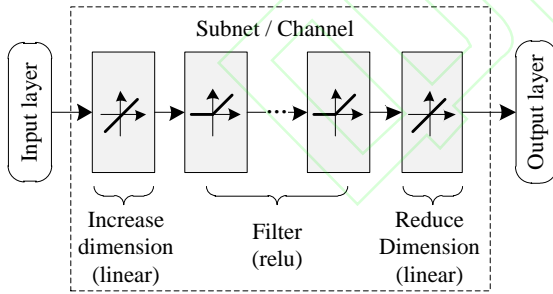


图 2 孪生网络的子网结构

Fig.2 Subnet structure of Siamese network

由于流形学习是非线性降维, 可将子网分为前、中、后三个部分。最前端是包含线性激活函数的升维层 (increase dimension layer), 中间部分是包含非线性激活函数 relu 的过滤层 (filter layer), 后端是包含线性激活函数的降维层 (reduce dimension layer)。升维层目的是为了样本的特性尽量充分地表达出来, 使得训练效果更好。这个过程类似支持向量机 (SVM)^[23]利用核函数将输入数据映射到高维特征空间的效果。如果输入样本的维度本就很

高, 可以考虑省略这一层。过滤层选择激活函数 relu 的原因是, relu 能有效克服梯度消失问题, 并且计算速度和收敛速度都很快。更重要的是, relu 是单向饱和 (负向饱和), 当激励小于 0 时输出为 0, 当激励大于 0 时输出则线性增强。单向饱和的直接效果就是, 过滤层会产生大量饱和值为 0 的“死结点”, 而剩余的“活结点”才能与后续的降维层构成有效输出。整个过程等效于, 过滤层中的“活结点”与后端降维层一同构成了局部线性映射关系。孪生网络的具体算法如图 3 所示。

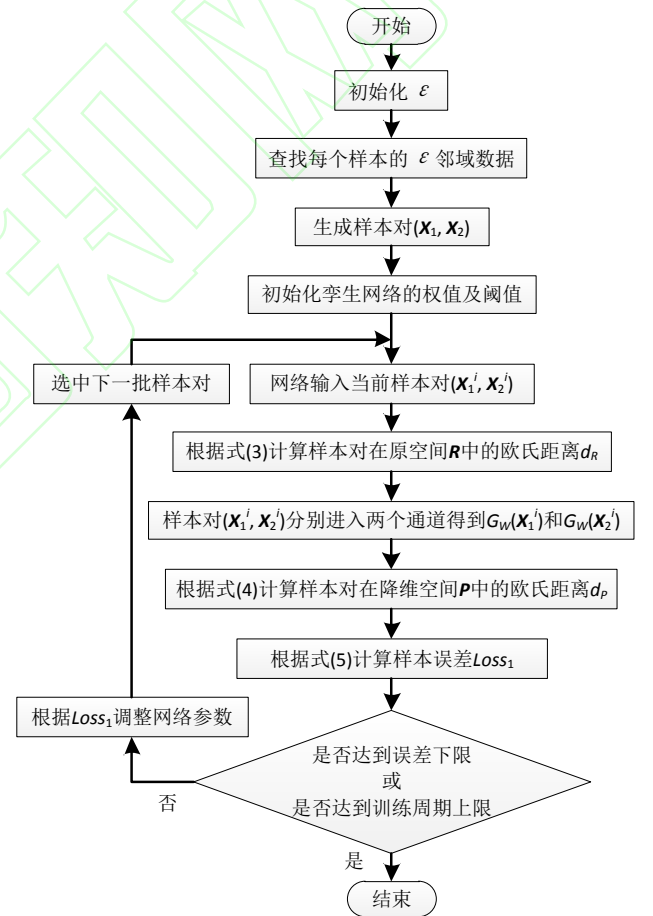


图 3 孪生网络算法流程图

Fig.3 Flow chart of Siamese network

当孪生网络训练到一定的程度后, 可任选两条通道中的一条, 用作降维映射。图 4 呈现了孪生网络对于不同曲率的流形结构的降维效果。由于原始空间只有三维, 所以子网结构包含 1 层线性升维层 (100 个结点), 2 层非线性过滤层 (各 100 个结点) 和 1 层线性降维层 (2 个结点)。

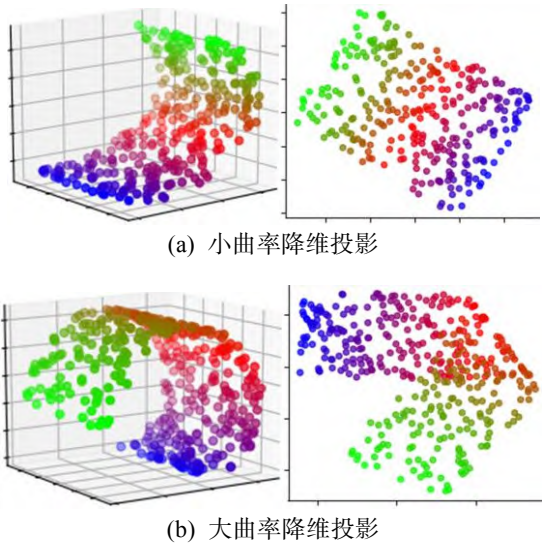


图 4 不同曲率下的流形学习
Fig.4 Manifold learning under different curvature

当流形结构的曲率不大时,如图 4(a)所示,孪生网络的降维效果是良好的。但当曲率增加时,如图 4(b)所示,会出现子空间内重叠的现象。造成这一问题的原因是,设计损失函数时,只考虑了保持邻近样本间的距离,而没有考虑远离样本间的空间关系,当样本在子空间内重叠时,没有对损失函数产生明显的负面影响。为了让孪生网络具有更好的适应性,必须对损失函数做出改进。

3.2 两级邻域的损失函数

要避免降维数据在子空间内重叠,除了确保邻域样本在子空间里依然保持距离,还要确保非邻域样本在子空间里也要互相远离。为了在损失函数中体现样本的远离效果,本研究提出了两级邻域的概念,一级邻域的距离是 ε_1 ,二级邻域的距离是 ε_2 ($0 < \varepsilon_1 < \varepsilon_2$),如图 5 所示。为此,需要重新引入样本标签 Y ,样本输入也还原为传统孪生网络的输入形式 (X_1, X_2, Y) 。如果 X_1 与 X_2 距离小于等于 ε_1 , Y 取值为 0,为正样本;如果的 X_1 与 X_2 距离大于 ε_1

且小于等于 ε_2 , Y 取值为 1,为负样本。正样本要保持欧氏距离不变,而负样本要保证欧氏距离大于 ε_1 。从流形结构的特点来看,当样本远离到一定的程度后,比如超过 ε_2 时,样本间的欧氏距离变得没有意义,所以设置 ε_2 值的时候没有必要太大,可以综合考虑样本间的平均距离和最大距离来设定 ε_1 和 ε_2 。

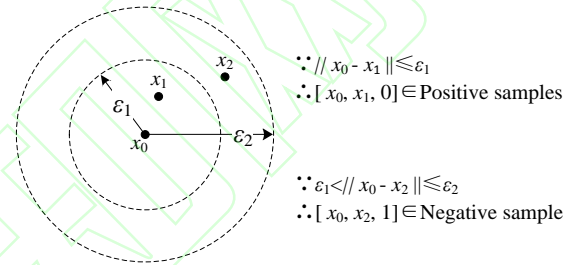


图 5 两级邻域内的正、负样本
Fig.5 Positive and negative samples in two-level neighborhood

为了体现正、负样本各自产生的效果,重新设计损失函数 $Loss_2$,如公式(6)所示:

$$Loss_2(W) = \sum_{i=1}^n \{ (1 - Y^i) |d_R(X_1^i, X_2^i) - d_P(X_1^i, X_2^i)| + \lambda Y^i \text{Max}(0, \varepsilon_1 - d_P(X_1^i, X_2^i)) \} \quad (6)$$

如果第 i 个样本 (X_1^i, X_2^i, Y^i) 是正样本, Y^i 取值为 0,则 $Loss_2$ 的前半项有效,相当于 $Loss_1$ 的效果,即保持邻近样本间的欧氏距离。如果第 i 个样本是负样本, Y^i 取值为 1,则 $Loss_2$ 的后半项有效,此时当 X_1^i 与 X_2^i 在子空间内的距离大于 ε_1 时,符合负样本要远离的特征,对损失函数没有影响,反之当 X_1^i 与 X_2^i 在子空间内的距离小于 ε_1 时,不符合负样本要远离的特征,则产生罚值,且靠得越近惩罚力度越大。 $Loss_2$ 的后半项还带有一个系数 λ ,是为了调节负样本在损失函数中的权重,经验表明,当 λ 的值大于等于 1 时具有的较好的效果。改进后的孪生网络算法如图 6 所示。

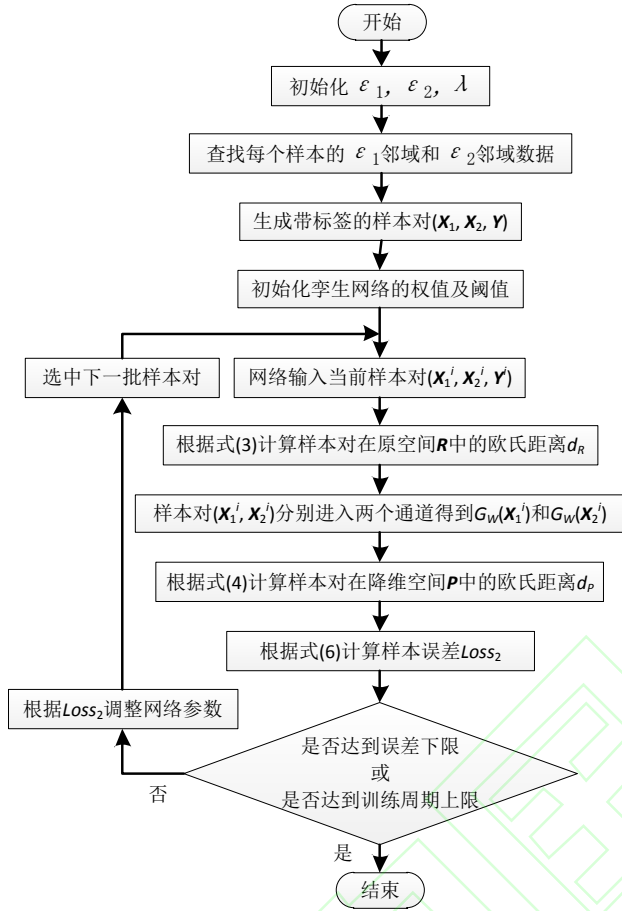


图 6 改进后的孪生网络算法流程图

Fig.6 Flow chart of improved Siamese network

3.3 对比实验

下面将改进后的孪生网络应用到经典的仿真数据 (Swiss roll) 的降维中。由于 Swiss roll 的原始空间仅仅是三维, 为了充分表达样本蕴含的信息, 即降低网络训练难度, 孪生子网的前端设计了 150 个结点的升维层, 将 3 维输入映射到 150 维空间中。子网的中部是三层各包含了 150 个结点的过滤层, 最后是包含 2 个结点的降维层。Swiss roll 采样了 800 个数据, 数据间的平均欧氏距离为 0.68。实验设置一级邻域的 ε_1 设为 3.0, 二级邻域的 ε_2 设为 6.0, $Loss_2$ 中的 λ 设为 100。训练集中正样本 (距离小于等于 ε_1 的数据对) 有 6488 对, 负样本 (距离介于 ε_1 和 ε_2 之间的数据对) 有 17123 对, 网络的学习率为 0.001。经过 150 个周期 (epoch) 的训练, 孪生网络的学习效果与其它传统的流形学习方法进行对比, 如图 7 所示。

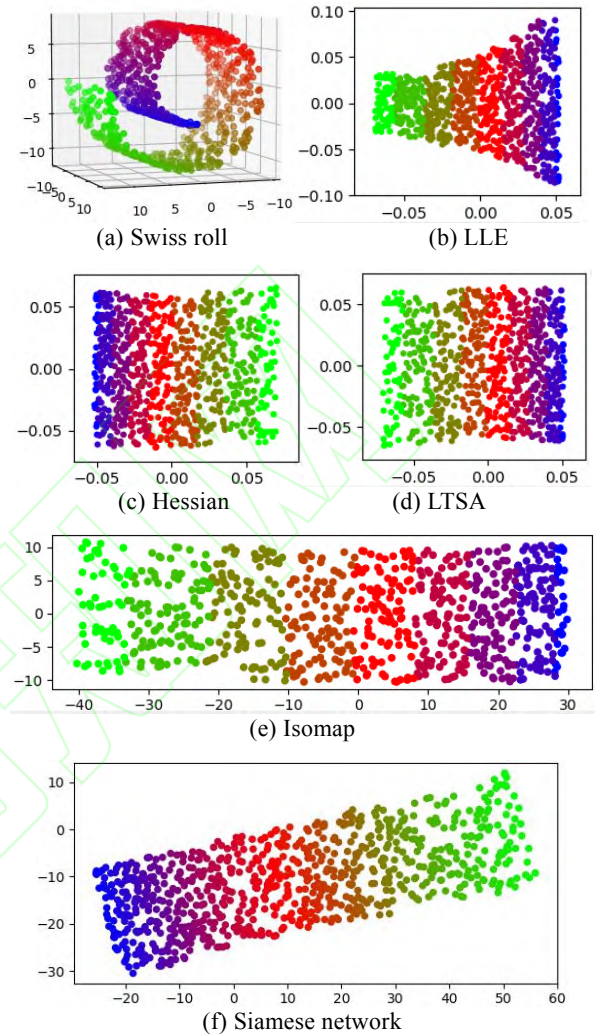


图 7 Swiss roll 降维效果对比

Fig.7 Comparison of dimension reduction effect of Swiss roll

图 7(a)为 Swiss roll 在原始三维空间中形状, 图 7(b~e)分别是采用 LLE、Hessian、LTSA 和 Isomap 方法降维后的效果, 图 7(f)是孪生网的降维效果。通过降维后的比例尺可以看出, Isomap 和孪生网络在降维时, 有效保留了流形结构在高维空间内的真实跨度 (比例), 而其它传统方法则是带人为约束的优化问题 (见 3.1 节), 这种约束造成了各维度间的不成比例的压缩, 如图 7(b~d)所示, LLE、Hessian 和 LTSA 方法将 Swiss roll 压缩在一个长宽大约都为 0.2 的矩形区域内。需要注意, 各维度不成比例的压缩, 将造成样本在子空间内各维度上的疏密程度不能真实反应高维流形的局部空间关系, 最终导致低维数据整体的变形与失真。虽然 Isomap 方法在全局尺度上可以较好的保持数据在空间中的跨度, 但

是因为采用近似测地距离作为计算的依据,所以在局部结构上会存在变形,特别是当数据分布不均或存在采样空洞的情况下,变形将更为明显。与 Isomap 不同,孪生网络是将高维空间中的欧氏距离尽可能地映射到低维空间中,所以对数据分布或采样空洞不敏感,使得 Swiss roll 的流形结构在降维后的空间中得以真实地保留。

4 可视化手写数据集

流形学习主要用于流形结构的降维、特征提取或可视化等目的,本节将孪生网络用于实际数据的可视化处理中。文献[24]分析了字符的旋转和缩放数据集,证明字符是嵌入高维空间内的流形结构,本节对 scikit-learn 手写数字集^[25]进行二维可视化操作。

scikit-learn 手写数字集是 8×8 的图像,样本空间是 64 维,具有较高维度,可以省略升维层,而直接添加过滤层和降维层。孪生网络的子网结构是包含 64 个结点的输入层、两层各包含 200 个结点的过滤层和包含 2 个结点的降维层。从手写数据集中选择 0~5 这六种数字,并对每种数字随机选择 50 个样本,共 300 个原始样本。原始样本归一化后,设置一级邻域的 ε_1 为 2.0,二级邻域的 ε_2 为 3.0,产生 3795 个正样本和 15006 个负样本。 $Loss_2$ 中的 λ 设为 10,网络的学习率为 0.005。经过 250 个周期的训练,孪生网络的二维可视化效果与其它传统方法对比,如图 8 所示。

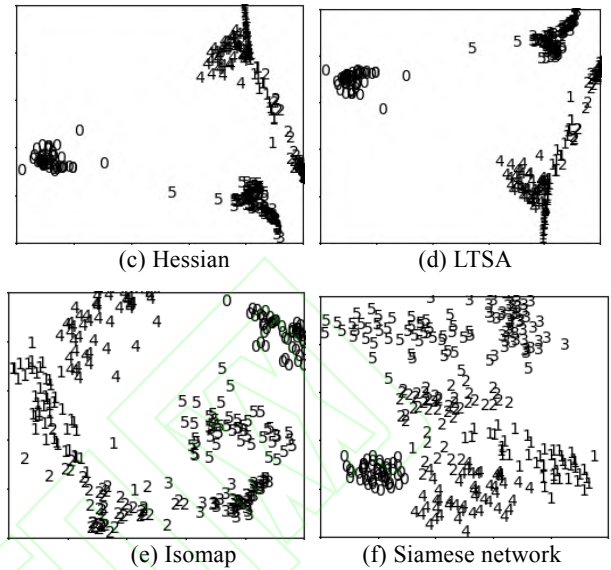
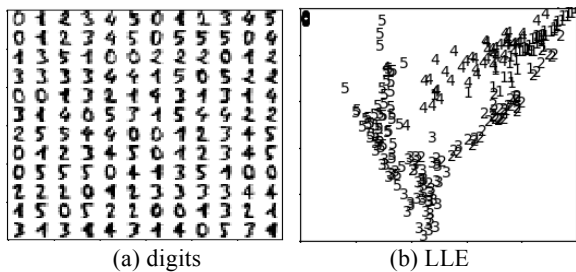


图 8 手写数字集降维可视化效果对比

Fig.8 Comparison of dimension reduction visualization effect of handwritten digit set

从图 8 的实验结果来看,LLE、Hessian 和 LTSA 方法虽然有明显的聚类效果,但是类别的边界较模糊,个别类别有明显的重叠现象。相对的,Isomap 和孪生网络也有明显的聚类效果,类别边界较为清晰。整体来看,经过孪生网络可视化后,数据在空间内的分布更加均匀,没有明显空间压缩的痕迹,空间关系的辨识度较高,这为后续的数据分析打下了良好的基础。

5 结束语

本文为流形学习提供了一种新的研究思路,即通过孪生网络的“样本对”训练方法,实现能够保持原有的空间关系的降维映射。同时,针对孪生网络的通道设计,提出了升维、过滤和降维的三层结构,以及两级邻域采样的概念,并设计了包含正、负样本对的损失函数 $Loss_2$ 。将该方法运用在仿真数据 (Swiss roll) 的流形降维和真实数据 (handwritten digits) 的二维可视化上,都取得了良好的效果。该方法的优点有:(1) 对流形结构没有凸包要求、空洞限制;(2) 对数据的均匀采样要求不高;(3) 降维设计避免了复杂的数值分析和添加额外约束;(4) 降维过程没有维度压缩、还原度高。

同时,用孪生网络进行流形学习也存在局限性与不足:(1) 随着训练集增大、流形结构曲率增加,

网络结构也更加复杂,训练难度也在加大;(2)与传统方法类似,如果数据集中引入过多的新数据,则需要重新进行学习,才能达到较好的效果。针对前一种局限性,可以考虑优化训练集的产生方式,控制训练集的规模,同时根据具体的应用,给出子网结构的优化方案。针对后一种局限性,可以考虑迁移学习,即在原有训练好的网络基础上继续进行新的数据集的训练。

参考文献

- [1] Olson C C, Judd K P, Nichols J M, et al. Manifold learning techniques for unsupervised anomaly detection[J]. *Expert Systems with Applications*, 2018: 374-385.
- [2] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000, 290(5500): 2323-2326.
- [3] Tenenbaum J B. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. *Science*, 2000, 290(5500): 2319-2323.
- [4] Salem N, Hussein S. Data dimensional reduction and principal components analysis[J]. *Procedia Computer Science*, 2019: 292-299.
- [5] Qi Y, Zhang H, Zhang B, et al. Cross-media retrieval based on linear discriminant analysis[J]. *Multimedia Tools and Applications*, 2018, 78(17):1-20.
- [6] Wang J, Wong R K W, Lee T C M. Locally linear embedding with additive noise[J]. *Pattern Recognition Letters*, 2019, 123(MAY):47-52.
- [7] Li H, Trocan M. Sparse reconstruction of ISOMAP representations[J]. *Journal of Intelligent and Fuzzy Systems*, 2019, 37(6): 7519-7536.
- [8] Li B, Li Y, Zhang X, et al. A survey on Laplacian eigenmaps based manifold learning methods[J]. *Neurocomputing*, 2019: 336-351.
- [9] Zhao Y, Fang L, Lv Y, et al. Improved linear local tangent space alignment and its application to pattern recognition[J]. *International Journal of Computer Applications in Technology*, 2017, 56(3):244.
- [10] 崔鹏,张雪婷.基于流形学习的泛化改进的 LTSA 算法[J]. *计算机工程与应用*, 2017, 53(3): 201-204.
Cui P, Zhang X. Generalized improvement of LTSA algorithm based on manifold learning. *Computer Engineering and Applications*, 2017, 53(3): 201-204.
- [11] Pimenta M Z, Comin C H, Rodrigues F A, et al. The impact of Interconnecting Topologies on SOM Neural Networks[C]. *international joint conference on neural network*, 2018: 1-6.
- [12] Zheng Y, Shi Y, Guo K, et al. Enhanced word embedding with multiple prototypes[C]//2017 4th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS). *IEEE*, 2017: 1-5.
- [13] Hinton G E, Salakhutdinov R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.
- [14] Langford Z, Eisenbeiser L, Vondal M. Robust signal classification using siamese networks[C]//*Proceedings of the ACM Workshop on Wireless Security and Machine Learning*, 2019: 1-5.
- [15] Chopra S, Hadsell R, Lecun Y. Learning a similarity metric discriminatively, with application to face verification[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). *IEEE*, 2005: 539-546.
- [16] A A H F, B F M, A M A. Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems[J]. *Petroleum*, 2020, 6(1):80-91.
- [17] Yuan X, Pang E, Lin K, et al. Hierarchical multilabel classifier for gene ontology annotation using multihead and multiend deep CNN model[J]. *IEEJ Transactions on Electrical and Electronic Engineering*, 2020, 15(7): 1057-1064.
- [18] Karbasi D, Nazemi A, Rabiei M. A parametric recurrent neural network scheme for solving a class of fuzzy regression models with some real-world applications[J]. *Soft Computing*, 2020, 24(15):11159-11187.
- [19] Lancewicki T, Aladjem M. Locally multidimensional scaling by creating neighborhoods in diffusion maps[J]. *Neurocomputing*, 2014, 139: 382-396.
- [20] Saeed N, Nam H, Haq M I U, et al. A Survey on Multidimensional Scaling[J]. *Acm Computing Surveys*, 2018, 51(3):1-25.
- [21] Jiang Z, Liu X, Sun M, et al. A Density Peak Clustering Algorithm Based on the K-Nearest Shannon Entropy and Tissue-Like P System[J]. *Mathematical Problems in Engineering*, 2019: 1-13.
- [22] Noh Y K, Zhang B T, Lee D D. Generative Local Metric Learning for Nearest Neighbor Classification[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(1):106-118.
- [23] Lan L, Wang Z, Zhe S, et al. Scaling Up Kernel SVM on Limited Resources: A Low-Rank Linearization Approach[J]. *Neural Networks and Learning Systems, IEEE Transactions on*, 2019, 30(2):369-378.
- [24] Wikipedia. Nonlinear Dimensionality Reduction[EB/OL]. (2020-03-15) http://wikipedia.moesalih.com/Nonlinear_dimensionality_reduction.
- [25] Scikit-learn. Manifold learning on handwritten digits Reduction[EB/OL]. (2020-03-15) https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py.