

基于 Autoencoder 网络的数据降维和重构

胡昭华^{①②} 宋耀良^①

^①(南京理工大学电子工程与光电技术学院 南京 210094)

^②(南京信息工程大学电子与信息工程学院 南京 210044)

摘要: 在机器学习, 模式识别以及数据挖掘等诸多研究领域中, 往往会面临着“维数灾难”问题。因此, 特征数据的降维方法, 即将高维的特征数据如何进行简化投射到低维空间中再进行处理, 成为当前数据驱动的计算方法研究热点之一。该文引入一种特殊的非线性降维方法, 称为自编码 (Autoencoder) 神经网络, 该方法采用 CRBM (Continuous Restricted Boltzmann Machine) 的网络结构, 通过训练具有多个中间层的双向深层神经网络将高维数据转换成低维嵌套并继而重构高维数据。特别地, 自编码网络提供了高维数据空间和低维嵌套结构的双向映射, 有效解决了大多数非线性降维方法所不具备的逆向映射问题。将 Autoencoder 用于人工数据和真实图像数据的实验表明, Autoencoder 不仅能发现嵌入在高维数据中的非线性低维结构, 也能有效地从低维结构中恢复原始高维数据。

关键词: 自编码网络; 高维数据; 降维; 重构

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2009)05-1189-04

Dimensionality Reduction and Reconstruction of Data Based on Autoencoder Network

Hu Zhao-hua^{①②} Song Yao-liang^①

^①(School of Electronic Engineering and Optoelectronic Technology, Nanjing University of Science and Technology, Nanjing 210094, China)

^②(College of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: The curse of dimensionality is a central difficulty in many fields such as machine learning, pattern recognition and data mining etc. The dimensionality reduction method of characteristic data is one of the current research hotspots in data-driven calculation methods, which high-dimensional data is mapped into a low-dimensional space. In this paper, a special nonlinear dimensionality reduction method called “Autoencoder” is introduced, which uses Continuous Restricted Boltzmann Machine (CRBM) and converts high-dimensional data to low-dimensional codes by training a neural network with multiple hidden layers. In particular, the “autoencoder” provides such a bi-directional mapping between the high-dimensional data space and the low-dimensional manifold space and is therefore able to overcome the inherited deficiency of most nonlinear dimensionality reduction methods that do not have an inverse mapping. The experiments on synthetic datasets and true image data show that the “autoencoder” network not only can find the embedded manifold of high-dimensional datasets but also reconstruct exactly the original high-dimension datasets from low-dimensional structure.

Key words: Autoencoder network; High-dimensional data; Dimensionality reduction; Reconstruction

1 引言

数据挖掘指的是从大量的数据中提取隐含的、事先未知的、潜在有用的知识的技术, 是目前国际上数据库和信息决策领域最前沿的研究方向之一。随着计算机技术、多媒体技术的发展, 在实际应用中经常会碰到高维数据, 如文档音频数据、交易数据以及多媒体数据等。由于这种数据存在的普

遍性, 使得对高维数据挖掘的研究有着非常重要的意义。但由于“维数灾难”的影响, 也使得高维数据挖掘变得异常困难, 必须采用一些特殊的手段进行处理。

处理高维数据的方式之一是对之进行降维, 通过学习以获得低维的隐变量模型。这些高维数据通常包含许多冗余, 其本质维数 (intrinsic dimension) 往往比原始的数据维数 (data dimension) 要小得多, 因此高维数据的处理问题可以归结为通过相关的降维方法减少一些不太相关的数据而降低它的维数, 然后用低维数据的处理办法进行处理。传统的

降维技术可以分为线性和非线性两类。线性降维方法如主成分分析(principal component analysis)、独立分量分析和因子分析(factor analysis)在高维数据集具有线性结构和高斯分布时能有好的效果。当数据集在高维空间呈现高度扭曲时, 这些方法则难以发现嵌入在数据集中的非线性结构以及恢复内在的结构。根据先前的研究表明, 大多数情况特征数据是分布在弯曲的流形空间中, 因此研究非线性的降维方法势在必行。自 2000 年在 Science 上陆续发表了两篇被称为等度规映射(ISOMAP)^[1]和局部线性嵌入算法(LLE)^[2]的有关非线性降维的开创性论文以来, 流形学习方法得到了广泛的应用。流形学习通常假定数据集具有本质维数, 从而可以通过将数据集降维至低维空间来避免维数灾难问题, 并发现隐含在数据中的内在物理意义。然而, 现有大多流形学习方法^[3,4]的主要缺陷在于它们只能学习已知数据集的潜在低维结构, 并不能给出高维空间中数据点到低维空间的确切映射, 同样也无法给出相应的逆映射。从而这些非双向映射方法只适用于原始的训练数据集而不能有效运用于新出现的数据集。因此, 采用这类降维方法必须联合运用其他复杂的技术来改进这种缺陷。

本文引入一种新的非线性降维方法——自编码神经网络(autoencoder network)方法^[5,6]。该方法采用自适应、多层编码(encoder)网络将高维原始数据转换成低维嵌套, 通过训练具有多个中间层的双向深层神经网络将高维数据转换成低维嵌套并利用类似的解码(decoder)网络从低维嵌套中重构高维数据。自编码网络给出了输入的高维数据与低维嵌套之间的双向映射, 从而克服了大多数非线性降维方法所不具备的逆映射问题。将自编码网络用于人工数据和真实图像数据的实验表明, 自编码网络具有很好的降维和重构功能。

本文在介绍自编码神经网络和它的结构和工作原理的基础上, 着重研究了利用人工数据和真实图像数据实现自编码网络的降维和重构机理。

2 自编码网络(Autoencoder network)结构和工作原理

2.1 自编码网络结构

已知 $R = \{r_1, \dots, r_n\} \in \mathcal{R}^{D \times n}$ 为包含 n 个数据的高维数据集, 其中 r_i 是第 i 个 D 维数据, $M = \{m_1, \dots, m_n\} \in \mathcal{R}^{d \times n}$ 满足 $d \ll D$, 为一种低维嵌套。我们希望利用自编码网络能找到高维数据集 R 的低维嵌套结构 M 。

自编码网络系统结构如图 1 所示, 整个系统由编码(encoder)和解码(decoder)两个网络构成。编码网络属于降维部分, 作用是将高维原始数据降到具有一定维数的低维嵌套结构上; 解码网络属于重构部分, 可视为编码网络的逆过程, 作用是将低维嵌套上的点还原成高维数据。编码网络与解码网络之间还存在一个交叉部分, 称之为“码字层”(code layer), 它是整个自编码网络的核心, 能够反映具有嵌套结

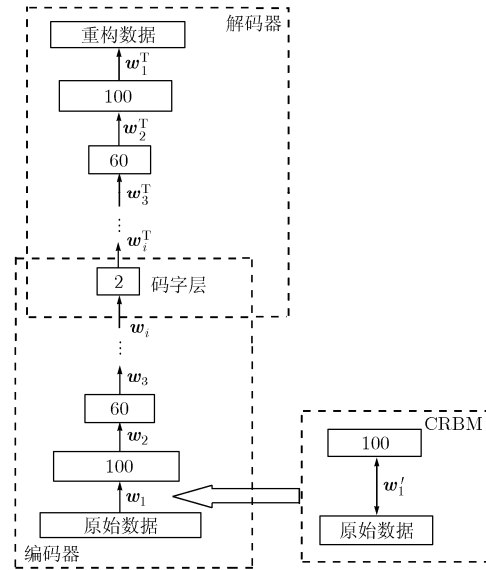


图 1 自编码(autoencoder)网络系统结构图

构的高维数据集的本质规律, 并确定高维数据集的本质维数。

自编码网络的工作原理如下: 首先初始化编码和解码两个网络的权值, 然后按照原始训练数据与重构数据之间误差最小化的原则对自编码网络进行训练。先经过解码网络然后再经过编码网络采用向后传播误差导数的链式法则很容易得到所需的梯度值, 进而将自编码网络的权值调到最佳值。

如果自编码网络的初始权值接近最优解, 运用梯度下降法则能达到很好的训练结果。Hinton 和 Salakhutdinov^[5]使用了一种称为限制玻耳兹曼机(Restricted Boltzmann Machine, RBM)^[7]的两层网络来求取自编码网络的适当初始权值。然而用 RBM 来建模连续数据并不太理想。因而本文引入了限制玻耳兹曼机的连续形式(CRBM)^[8,9], 它是一种连续的随机再生模型, 能够用一种简单、可靠的训练算法来建模连续数据。本文将运用 CRBM 建模连续数据的过程称为“预训练”过程。CRBM 的结构如图 1 右框图所示。

在预训练多层 CRBM 后, 编码和解码网络都将使用经 CRBM 训练得到的权值作为自编码网络的初始权值。对于高维数据集而言, 这是一种逐渐揭示数据集内在低维结构的有效方式。接下来的全局调整过程则使用反向传播算法通过整个自编码网络对权值进行调整从而达到数据集的最佳重构。

2.2 CRBM 工作原理

CRBM 的结构包括一个可视层和一个隐层以及它们层间的连接。图 1 的右边框图显示了 CRBM 的结构框架, 图中原始数据对应 CRBM 的可视单元, 因为它们的状态是可观测的。CRBM 的输出对应的是隐单元。可视单元与隐单元之间由权值矩阵 w 连接。设 v_i 和 h_j 分别描述可视单元 i 和隐单元 j 的状态, 且它们之间的双向权值相等, 即 $w_{ij} = w_{ji}$ 。

为了建模连续数据,CRBM 通过在可视层添加一个零均值的高斯噪声从而引入一个连续随机单元。对于 CRBM 而言,隐层中每个单元的输入都来自于可视层中所有单元的状态,按照各个可视单元对每个隐单元的贡献大小,对它们之间的连接赋以相应的重要权值,即每个隐单元的状态值是所有可视单元的状态值按照对其贡献大小的加权和。为了表示方便,从现在起,以下都将采用同一个标识符 s 来表示可视单元和隐单元的状态,设 s_j 表示输入来自于可视单元状态集 $\{s_i\}$ 的隐单元 j 的输出,则有

$$s_j = \varphi_j \left(\sum_i w_{ij} s_i + \sigma N_j(0,1) \right) \quad (1)$$

式中函数 φ_j 的表达式如下:

$$\varphi_j(x_j) = \theta_L + (\theta_H - \theta_L) \frac{1}{1 + \exp(-a_j x_j)} \quad (2)$$

其中 $N_j(0,1)$ 表示零均值、单位方差的高斯随机变量。常数 σ 和 $N_j(0,1)$ 共同产生了一个噪声输入分量 $n_j = \sigma N_j(0,1)$, 其概率分布为

$$p(n_j) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n_j^2}{2\sigma^2}\right) \quad (3)$$

由式(2)可知, $\varphi_j(x)$ 是渐近线在 θ_L 和 θ_H 处的 sigmoid 函数。参数 a_j 控制着 sigmoid 曲线的斜率,是噪声控制变量,当 a_j 由小变大时,可以完成从无噪声的确定性状态到二进制随机状态的平滑过渡。

CRBM 采用最小化对比散度(Minimizing Contrastive Divergence, MCD)训练准则替代了仅靠 Gibbs 抽样的玻耳兹曼机^[10]的松弛搜索,大大减少了计算量。MCD 训练准则用来更新 CRBM 的权值 $\{w_{ij}\}$ 以及“噪声控制”参数 $\{a_j\}$:

$$\Delta w_{ij} = \eta_w \left(\langle s_i s_j \rangle - \langle \hat{s}_i \hat{s}_j \rangle \right) \quad (4)$$

$$\Delta \hat{a}_j = \frac{\eta_a}{a_j^2} \left(\langle s_j^2 \rangle - \langle \hat{s}_j^2 \rangle \right) \quad (5)$$

其中 \hat{s}_j 表示单元 j 的一步重构状态, $\langle \cdot \rangle$ 表示训练数据的均值, η_w 是学习率。

式(4)和式(5)表明 CRBM 的训练准则只需进行简单的加法和乘法运算,从而使得计算量不至于过大,并且可以很容易完成权值的更新过程。

3 自编码网络的实现

3.1 人工数据的降维和重构

为了表明自编码网络能够有效调整深层网络,本节利用“瑞士卷”数据集来训练一个深层自编码网络。“瑞士卷”数据集以及自编码网络结构分别如图 2 和图 3 所示。由图可见训练数据“瑞士卷”集是三维的,但实际上它们却存在于一个嵌入三维空间的二维曲面上。而三维数据集和二维曲面

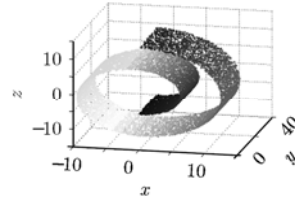


图 2 “瑞士卷”图

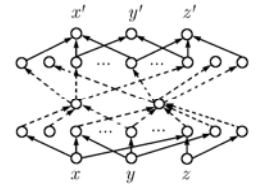
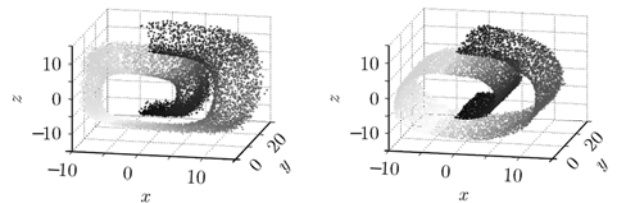


图 3 一个复杂的自编码网络

之间的关系是高度非线性的。

自编码由各层大小依次为 3-100-50-25-10-2 的编码网络和一个与之对称的解码网络构成。为了提高网络的学习能力,在输入层和中心隐层之间又加入了一些非线性中间层,描述了将三维数据降为两维以及从两维重构三维数据的高度非线性映射,如图 3 所示。自编码网络的训练数据是从“瑞士卷”上抽取的 12000 个三维点。我们首先使用 CRBM 算法初始化网络的权值以至于这些初始权值能接近最优解,然后再执行反向传播算法以获得最优解。自编码网络不仅将三维数据降到二维平面上,而且还完成了二维点到三维数据的最佳重构。图 4 分别显示了由 CRBM 对“瑞士卷”的重构图以及由自编码网络对“瑞士卷”的重构图。显然,CRBM 的重构结果与真实数据之间存在一定的误差,而自编码网络的重构效果几乎接近图 2 的真实数据。



(a)由CRBM重构的“瑞士卷”图 (b)由自编码网络重构的“瑞士卷”图

图 4

3.2 真实图像数据的降维和重构

为了表明自编码网络同样适用于真实图像数据,本节利用人体运动数据库^[11]中的图像数据集来训练一个深层自编码网络。该数据库中提供了几种类型的人体行为模板,本文选用侧面视角的步行(walk)模板进行实验。由于人体步行运动具有一定的连续性和周期性,因此只需考虑该侧面视角的人体步行姿态在周期内的运动即可,图 5 则显示了人体在一个周期内的侧面步行图像序列。将这些高维图像数据输入自编码网络进行训练,试着发现它们在低维(三维)空间的流形结构,从而可以指导人体运动的跟踪和姿态的识别。

实验中每幅模板图像的大小为 61×93,通过行堆叠的方



图 5 侧面视角的一个周期人体步行序列

式可将其转化为 5673 维实向量。自编码网络由各层大小依次分别为 5673-1000-500-250-100-3 的编码网络和一个与之对称的解码网络构成。将单个人的 walk 图像序列输入自编码网络进行训练,其中包含 6 个步行周期,每个周期由 15 帧图像构成。经过自编码网络的训练,人体步行的高维图像序列映射于三维空间中有回路的周期性曲线,这一结果与人体步行的周期性相吻合,图 6 则显示了降维结果,图 6(a)和图 6(b)都是同一 walk 序列降维到三维空间的结果,只是显示的视角不同而已。图中 3 条曲线分别表示该 walk 序列中前 3 个 walk 周期的降维结果。

Walk 训练序列与重构序列的对比如图 7 所示。由此可见,自编码网络对高维图像数据也有很好的降维和重构效果。

4 结束语

本文引入了一种称为自编码网络(autoencoder)的双向



图 7

参 考 文 献

- [1] Tenenbaum J B, Silva V de, and Langford J C. A global geometric framework for nonlinear dimensionality reduction [J]. *Science*, 2000, 290: 2319-2323.
- [2] Roweis S T and Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. *Science*, 2000, 290: 2323-2326.
- [3] Rahimi A, Recht B, and Darrell T. Learning to transform time series with a few examples [J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007, 29(10): 1759-1775.
- [4] Wang Liang and Suter D. Learning and matching of dynamic shape manifolds for human action recognition [J]. *IEEE Trans. on Image Processing*, 2007, 16(6): 1646-1661.
- [5] Hinton G E and Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313: 504-507.
- [6] Zeng Xian-Hua, Luo Si-Wei, and Wang Jiao. Auto-associative neural network system for recognition [C]. Proceedings of the sixth international conference on machine learning and cybernetics, Hong Kong, 2007: 2885-2890.
- [7] Hinton G E. Training products of experts by minimizing

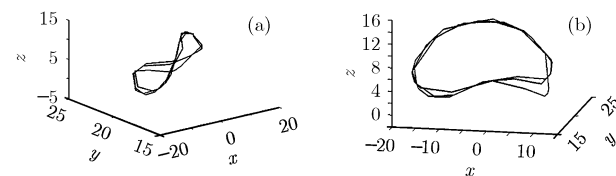


图 6 侧面视角人体步行序列的三维降维结果

非线性深层神经网络,该方法采用 CRBM 的网络结构,用以挖掘高维数据所嵌入的低维嵌套。自编码网络不仅能提供从高维数据空间到低维嵌套的映射,而且也可以给出相反的逆映射,亦即自编码网络给出了输入的高维数据与低维嵌套之间的双向映射,从而克服了大多数非线性降维方法所不具备的逆映射问题。将自编码网络用于人工数据和真实图像数据的实验表明,自编码网络对高维数据具有很好的降维和重构功能。

- contrastive divergence. *Neural Computation*[J]. 2000, 14(8): 1771-1800.
- [8] Chen H and Murray A F. A continuous restricted Boltzmann machine with hardware-amenable learning algorithm [C]. Proceedings of 12th Int. Conf. on Artificial neural networks (ICANN2002), Madrid, Spain, 2002: 358-363.
- [9] Chen H and Murray A F. Continuous restricted Boltzmann machine with an implementable training algorithm [J]. *IEE Proceedings of Vision, Image and Signal Processing*, 2003, 150(3): 153-158.
- [10] Hinton G E. Training products of experts by minimizing contrastive divergence. London: Gatsby Computational Neuroscience Unit, Technical Report: GCNU TR 2000-004, 2000.
- [11] Blank M. Human action database. <http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html>. 2007.4.

胡昭华: 女, 1981 年生, 博士生, 研究方向为视觉跟踪、模式识别、粒子滤波及机器学习等。
宋耀良: 男, 1960 年生, 教授, 博士生导师, 主要研究方向为自适应信号处理、雷达信号处理、混沌通信等。