

基于监督学习自动编码器图像重构

张 赛, 芮 挺*, 李华兵, 方虎生
(中国人民解放军理工大学, 江苏 南京 210007)

摘要:对于通过各种途径传播的海量信息(图像、语音及文本等)在传输途中可能会导致信息部分受损或丢失的问题,文章提出一种基于无监督学习自动编码器(Auto-Encoder, AE)用于监督学习的新方法,对受损数据进行重构。新方法模型与无监督自动编码器模型相比,提高了模型对特征的表达能力,丰富了特征“字典”的表达内容。文章以 Mnist 手写字符样本库通过添加噪声的方式模拟受损样本对新模型进行测试验证。实验表明,文章提出的新方法对受损图像重构具有良好的效果。

关键词:监督学习; 自动编码器; 图像重构; 特征表达

0 引言

一般而言,有很多因素都会引起数字图像上的局部信息缺损,如在数字图像的获取、传输、处理、压缩和解压缩过程中都有可能引起图像部分信息的损坏或丢失。为了保证图像信息的完整性,需要对这些受损图像的缺失部分进行重构。所谓图像修复就是对图像上信息缺损区域进行信息填充的过程,其目的就是对于有信息缺损的图像进行恢复^[1]。

Bertalmio 等^[2]在参考了修复者手工修复过程后,于 2000 年提出了数字图像修复这一概念。Bertalmio, Sapiro, Caselles 和 Ballester^[2]四人(即 BSCB 模型)最先使用基于偏微分方程(Partial Differential Equation, PDE)模型来解决图像修复问题。2004 年, Criminisi^[3]等提出一种基于样本块的修复方法。该修复方法同时考虑了图像的结构和纹理信息,以等照度线法线方向,待修复区域边界像素点的优先级,采用 SSD 来搜索最优匹配块的方法。2006 年, Komodakis^[4]提出的基于优先权的方向传播算法(Backpropagation Algorithm),其将图像修复问题描述成离散全局最优问题,其目标函数即求马尔科夫随机场(Markov Random Field, MRF)的能量最小化函数。

以无监督学习为代表的自动编码器(Auto-Encoder, AE)在图像分类和识别领域有着显著的优势,它可以通过自学习的方式学习提取数据自身的精简表达,利用精简的编码特征重构出数据本身的信息。由于该模型对特征“字典”的学习受到训练样本自身信息的局限性,如果对受损数据进行学习,则特征“字典”经过学习获取的特征也一定是不完整的。

综上所述,为了充分发挥经典 AE 特征提取的能力和监督学习的优势,本文基于无监督学习 AE 模型

用于监督学习的方式,提出了基于监督学习自动编码器数据重构的新方法。以 Mnist 字符库的图像中引入不同比例的冲击噪声模拟受损样本作为测试对象,将其用于对受损图像进行重构测试。通过新模型对受损图像重构后的分类结果显示,证明了新模型对受损图像重构的有效性。

1 无监督自动编码器

1986 年, Rumelhart 等^[5]提出自动编码器的概念,并将其用于高维复杂数据处理,促进了神经网络的发展。AE 是一个浅层网络, AE 参数模型的目标是最小化输入数据 X 与重构数据 Z 之间的平均重构误差:

$$J(W, b) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|Z^i - X^i\|^2 \right) \quad (1)$$

其中, m 为训练样本个数; Z^i 表示第 i 个输入样本重建的新样本; X^i 表示第 i 个输入样本。

自动编码器模型包括输入层(Input Layer)、隐含层(Hidden Layer)、输出层(Output Layer),它利用反向传播算法训练该浅层网络,学习一个近似恒等函数,并利用前向传播将原始输入 X 通过权值转化编码成隐含层的低维表达 X_e ,使得该低维编码数据能够近似重构出输入数据的原始形式。

因此 AE 的输出目标就是通过将目标输出设置得和输入一样,从而进行数据自我复原训练,能够从无标记数据中学习代表性特征,逼近输入原始数据,是一个无监督学习模型。从输入层到隐含层的压缩低维表达过程可以称作模型的编码阶段(Encoder),从隐含层的压缩特征映射还原输出层的近似原始数据的过程称为解码阶段(Decoder)。

2 监督学习自动编码器模型

对上述经典的 AE 模型可以看作是一个特殊的 3

作者简介:张赛(1991—),男,江苏徐州人,硕士研究生;研究方向:机器学习,计算机视觉。

***通信作者:**芮挺(1972—),男,江苏南京人,副教授,博士;研究方向:机器学习,计算机视觉。

层反向传播(Back Propagation, BP)神经网络。通过学习提取的特征仅仅是输入样本自身仅存的特征。如果输入样本自身是不完整的或者说是存在信息缺失的,那么通过传统 AE 模型的学习,特征“字典”获取的表征一定是不充分的。

为了弥补无监督学习对特征表达能力的局限性,合理利用监督学习的优势,本文提出了对经典无监督学习 AE 模型进行监督学习的新方法。新模型将原无监督 AE 模型的目标函数(即输入样本 X 与重构样本 Z 之间的平均重构误差最小化)改为最小化监督标签 X_{label} 与重构样本 Z 之间的平均重构误差,目标函数变为:

$$J_{\text{supervised}}(W, b) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \| Z^i - X_{\text{label}}^i \|^2 \right) \quad (2)$$

其中, m 为训练样本个数; X_{label}^i 表示第 i 个给予监督的理想样本。监督学习自动编码器模型结构如图 1 所示。

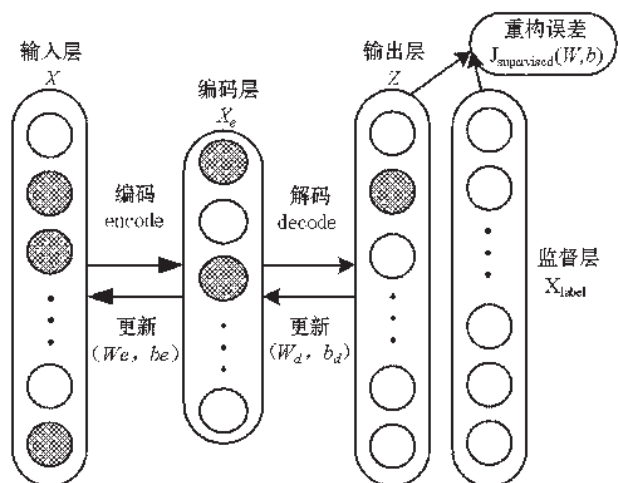


图 1 监督学习单层 AE 模型结构

经典 AE 模型可在网络学习的过程中,获取的特征表达具有局限性,通过学习提取的特征仅仅是输入样本自身固有的特征。本文采用监督学习的方式,改进经典 AE 训练策略,借助监督学习的优势来弥补经典 AE 模型对样本特征表达能力的局限性。

3 实验与分析

3.1 实验数据

手写字符识别是图像分类中最具代表性的经典问题,以这个问题作为测试对象,验证本文提出的有监督学习 AE 模型对受损样本数据重建的效果。本文采用 Mnist 手写字符数字样本库作为本实验的数据集,此数据库分为训练集和测试集,其中训练集包含 60 000 个样本数据,测试集包含 10 000 个样本数据。在输入网络训练前,对每个样本都是进行了尺度标准化及图像灰度化的预处理(每个样本尺度为 28×28),

同时对数据集分别添加了 0~50% 不等的椒盐冲击噪声模拟多种样本受损情况。图 2 为选取的部分理想样本(原始数据)和多种受损程度不同样本。



图 2 不同受损程度的样本

3.2 实验步骤

本文提出在无监督 AE 模型的框架下加以有效的监督,对受损样本进行重建的实验步骤如下:

(1)对单层的 AE 模型的输出端加以理想样本进行监督,以添加噪声的模拟受损数据集作为模型的输入,对网络进行训练,学习受损样本中的有效特征信息,利用理想样本对重建样本的监督来调整训练过程中的网络参数,即对学习的参数 W 进行完善。

(2)抽取学习完善的特征权值 W (权值包括:输入源到特征编码之间的特征权值和特征编码到输出之间的权值),建立一个单向的前馈神经网络(见图 3)对受损的测试样本进行样本的重建。

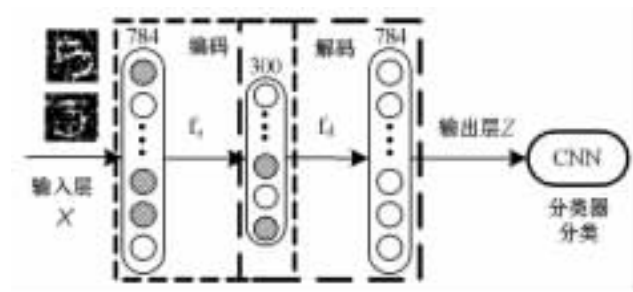
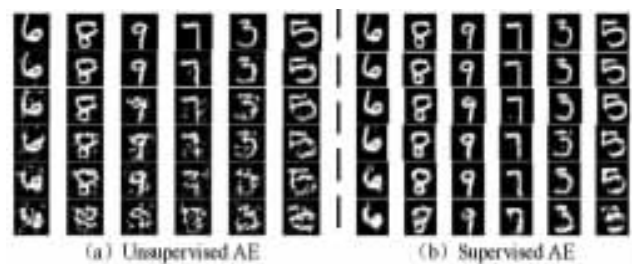


图 3 基于监督学习 AE 数据重构模型结构

(3)预训练一个卷积神经网络(Convolutional Neural Networks, CNN)作为分类器,将重建的新数据集输入进行最终的分类型测试。重建结果如图 4 所示。

3.3 实验结果分析

通过 CNN 分类的错误率比较模型的优劣,结果如表 1 所示。分类结果显示,有损样本通过无监督 AE 模型的重构,由于受损样本中受损信息的重要性是不



(a) 无监督 AE 模型依次对 5 种受损样本重构的结果 (第一行为无损样本, 从第二行开始, 依次是对含有 10%, 20%, 30%, 40%, 50% 样本受损情况的重构); (b) 监督学习 AE 模型对受损样本重构的结果。

图 4 受损样本在不同结构的 AE 模型下重构的结果对比

表 1 不同程度受损样本在不同模型重构下的 CNN 分类结果

噪声比例 (%)	噪声 10	噪声 20	噪声 30	噪声 40	噪声 50
CNN 直接分类 (%)	6.75	25.76	48.50	67.59	80.13
无监督 AE 模型 (%)	4.71	17.73	34.36	50.19	60.46
监督 AE 模型 (%)	2.06	3.11	5.83	10.49	19.56

可预测的, 一些缺失了重要信息的受损样本严重干扰了监督模型对特征的学习, 使得学习的特征表达不够充分, 随着受损程度的不断增加, 对受损样本重构的效果越差, 如图 4(a) 所示。

而通过监督学习 AE 模型对受损样本集依次进行重构分类, 根据受损程度的不同, 分类的错误率在受损样本直接进行分类的基础上分别降低了 4.69%~57.82% 不等, 其受损程度越严重, 重构的效果越显著, 其重构效果如图 4(b) 所示。通过重构的效果相比, 基于监督学习 AE 模型对受损样本重构的效果明显优于无监督学习 AE 模型重构的效果, 验证了本文提出

的监督学习 AE 模型对受损样本重构的有效性。

4 结语

在当前大数据的时代背景下, 图像作为记录数字化信息的一种主要方式, 在信息传播的过程中难免会造成部分数据的损坏或丢失, 降低了信息的使用价值。本文提出了一种基于无监督模型框架下加以监督的新方法, 有效地改善了无监督学习 AE 模型对受损样本特征的表达, 提高了数据重构能力。通过合理利用先验, 对受损数据进行重构, 弥补数据部分缺失的信息, 有效地维护了数据信息的实际利用价值, 证明了本文所提新方法的有效性。

参考文献

- [1] 张红英, 彭启琮. 数字图像修复技术综述[J]. 中国图象图形学报, 2007(1): 1-10.
- [2] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting: proceedings of international conference on computer graphics and interactive techniques [C]. 2000: 417-424.
- [3] CRIMINISI A, EREZ. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004(9): 120-121.
- [4] KOMODAKIS N. Image completion using global optimization[J]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2006(1): 442-452.
- [5] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors [J]. Nature, 1986(323): 533-536.

(责任编辑 傅金睿)

Image reconstruction based on the auto-encoder of supervised learning

Zhang Sai, Rui Ting*, Li Huabing, Fang Husheng

(The PLA University of Science and Technology, Nanjing 210007, China)

Abstract: In this paper, we propose a new method which is based on unsupervised learning of auto-encoders to supervised learning, which can be used to reconstruct mass information (images, voice and text, etc.) propagated through various channels that may lead to the damaged or loss of information. Compared with the unsupervised auto-encoder model, the new model improves the feature expression of the model, and enriches the expression content of the feature “dictionary”. In this paper, the Mnist handwritten samples database simulates damaged samples by adding noise to test the new model. Experiments show that the proposed method has a good effect on the reconstructed image.

Key words: supervised learning; auto-encoder; image reconstruction; feature expression