



(12) 发明专利

(10) 授权公告号 CN 113113030 B

(45) 授权公告日 2022.03.22

(21) 申请号 202110302058.1

G06N 3/08 (2006.01)

(22) 申请日 2021.03.22

G06N 3/04 (2006.01)

(65) 同一申请的已公布的文献号

G06K 9/62 (2022.01)

申请公布号 CN 113113030 A

G06V 10/774 (2022.01)

G06V 10/30 (2022.01)

(43) 申请公布日 2021.07.13

(56) 对比文件

(73) 专利权人 浙江大学

CN 108601026 A, 2018.09.28

地址 310027 浙江省杭州市西湖区浙大路
38号

WO 2021007812 A1, 2021.01.21

专利权人 浙江大学舟山海洋研究中心

US 2005154584 A1, 2005.07.14

CN 111724809 A, 2020.09.29

US 2015170655 A1, 2015.06.18

(72) 发明人 陈惠芳 谢磊 忻杨璇

张健. 基于目标声源方位已知的双麦克风噪声抑制.《清华大学学报(自然科学版)》.2011,
谢磊. 基于NGPP-SVDD的非高斯过程监控及其应用研究.《自动化学报》.2009,

(74) 专利代理机构 杭州君度专利代理事务所

(特殊普通合伙) 33240

代理人 朱亚冠

审查员 龚漂

(51) Int. Cl.

G10L 19/00 (2013.01)

G10L 21/0208 (2013.01)

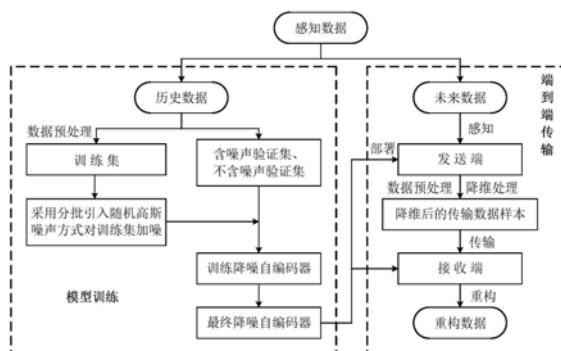
权利要求书3页 说明书7页 附图4页

(54) 发明名称

一种基于降噪自编码器的高维受损数据无线传输方法

(57) 摘要

本发明公开了一种基于降噪自编码器的高维受损数据无线传输方法。本发明方法包括模型训练和端到端传输。模型训练首先对历史感知数据集进行数据预处理,基于K-折交叉验证方法对其进行划分;然后构建降噪自编码器模型,并基于提出的新型分批引入随机高斯噪声的加噪方式训练降噪自编码器模型。端到端传输首先将训练得到的降噪自编码器拆分两部分部署于发送端和接收端,然后在发送端对未知类型噪声干扰的感知数据进行预处理和降维操作,将降维后的数据传输至接收端,最后在接收端执行重构操作,获得未受损感知数据的重构数据。本发明方法能有效进行高维受损感知数据的降维传输及降噪处理和重构,在恶劣环境下进行数据收集时滤除噪声干扰和降维传输。



1. 一种基于降噪自编码器的高维受损数据无线传输方法,包括模型训练和端到端传输,其特征在于:

所述的模型训练用于训练降噪自编码器,获得性能良好的传输架构;具体方法是:

步骤(1)对历史感知数据集进行数据预处理,对数据执行重采样、标准化以及添加噪声操作,组成具有 N_{samp} 个样本的标准化不含噪数据集 D 和含噪数据集 D_n ,样本长度均为 N_{len} ;

步骤(2)分割标准化不含噪数据集 D 和含噪数据集 D_n :将不含噪数据集 D 和含噪数据集 D_n 中 N_{samp} 个样本划分为 K 组训练-验证集 $\{D_{\text{train}}; D_{\text{valid}_n}\}_k$,满足 $N_{\text{train},k} + N_{\text{valid}_n,k} = N_{\text{samp}}$; $N_{\text{train},k}$ 为 $\{D_{\text{train}}\}_k$ 的样本数, $N_{\text{valid}_n,k}$ 为 $\{D_{\text{valid}_n}\}_k$ 的样本数,组号 $k=1, \dots, K$;

步骤(3)构建降噪自编码器模型,模型输入为含噪标准化序列 \tilde{x}'_{samp} ,模型输出为重构样本 \hat{x}'_{samp} ;

步骤(4)训练降噪自编码器模型,具体是:

(4-1)初始化:设置最大迭代轮数 N_{maxepoch} ,当前迭代轮数 $N_{\text{epoch}}=1$;设置停止准则中的阈值 thresh ;设置隐藏层的最大神经元个数 $N_{\text{hmax}} = \frac{2}{3}N_{\text{len}}$,当前隐藏层神经元个数 $N_{\text{hidd}}=1$;使用Xavier初始化方法随机初始化模型权值参数 w 和偏置参数 b ;设置当前训练集的组号为 $k=1$;

(4-2)分批:将训练集 $\{D_{\text{train}}\}_k$ 样本均匀分成 N_{batch} 个小批;

(4-3)加噪处理:随机选取一个小批,并生成一个在 $(0, 1)$ 上均匀分布的随机数 r ,将 $r/2$ 作为高斯分布的标准差,根据该标准差生成高斯噪声叠加到该小批上;

(4-4)前向传播:将加噪小批中的所有样本 \tilde{x}'_{samp} 输入降噪自编码器模型中,得到输出重构样本 \hat{x}'_{samp} ;

(4-5)后向传播:根据原始样本 x'_{samp} 和重构样本 \hat{x}'_{samp} 计算损失函数 $Loss = \frac{1}{I} \sum_{i=1}^I (\hat{x}'_i - x'_i)^2$, \hat{x}'_i 表示该小批中第 i 个含噪样本的重构输出值, x'_i 表示该小批中第 i 个含噪样本的原始值;根据梯度下降法更新模型参数 $w' = w - \eta \frac{\partial Loss}{\partial w}$, $b' = b - \eta \frac{\partial Loss}{\partial b}$, η 为学习率;

(4-6)验证:若所有小批均已选取过,则将含噪验证集 $\{D_{\text{valid}_n}\}_k$ 输入模型中,计算不含噪的验证集样本 x'_{samp} 和重构样本 \hat{x}'_{samp} 的均方误差 MSE_{valid_n} ,并计算其与保存的验证集最小均方误差 $\{MSE_{\text{valid}_n}\}_{\min}$ 的绝对差值 $\varepsilon = |MSE_{\text{valid}_n} - \{MSE_{\text{valid}_n}\}_{\min}|$;否则返回步骤(4-3);

(4-7)迭代:若迭代轮数 $N_{\text{epoch}} < N_{\text{maxepoch}}$ 且 $\varepsilon > \text{thresh}$,则更新并保存 $\{MSE_{\text{valid}_n}\}_{\min}$,返回步骤(4-2)进行下一轮迭代, $\{MSE_{\text{valid}_n}\}_{\min}$ 为本轮保存的最小 MSE_{valid_n} ;否则结束训练,保存第 k 组的最佳模型参数 $\{w_{\text{best}}, b_{\text{best}}\}_{k, N_{\text{hidd}}}$;

若隐藏层神经元个数 $N_{\text{hidd}} < N_{\text{hmax}}$,则返回步骤(4-1)进行下一个不同隐藏层神经元个数

的模型训练;

所述的端到端传输用于将训练得到的降噪自编码器模型部署至两端,实现数据的降维传输及降噪重构;方法是:

步骤(a) 部署训练得到的降噪编码器:从K个降噪自编码器模型选择出最优的模型参数;将降噪自编码器模型的编码阶段的超参数 W^e 和 b^e 部署于发送端;将编码阶段的非线性激活函数 f 和解码阶段的超参数 W^d 和 b^d ,以及非线性激活函数 g 部署于接收端;

步骤(b) 在发送端处理噪声类型未知的感知数据;

步骤(c) 将降维后的传输数据样本 Y_{samp} 传输至接收端;

步骤(d) 在接收端重构未受损感知数据:对接收数据样本 Y_{samp} 执行解码、反标准化以及重组操作,得到未受损感知数据的重构数据 \hat{x}_t 。

2. 如权利要求1所述的基于降噪自编码器的高维受损数据无线传输方法,其特征在于:步骤(1)具体是:

重采样操作:对历史感知数据使用符号间隔 T_s 进行采样,获取长度为 N_{len} 的离散序列

$x = [x_{mN_{\text{len}}+1}, x_{mN_{\text{len}}+2}, \dots, x_{(m+1)N_{\text{len}}}]^T$, x_t 表示离散序列中第 t 个采样点的离散值, $t = (mN_{\text{len}}+1), (mN_{\text{len}}+2), \dots, (m+1)N_{\text{len}}$, T 表示转置操作, m 表示重采样轮次;

标准化操作:对离散序列 x 进行最大最小化处理,获取标准化序列 $x' = [x'_{mN_{\text{len}}+1}, x'_{mN_{\text{len}}+2}, \dots, x'_{(m+1)N_{\text{len}}}]^T$,

每个样本点取值范围是 $(-1, 1)$,第 t 个采样点的标准化值 $x'_t = 2 \times \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} - 1$, x_{\max} 为离散序列中最大离散值, x_{\min} 为离散序列中最小离散值;

添加噪声操作:在标准化序列上叠加随机生成的与多种实际场景相符的噪声 n ,得到含噪的标准化序列 $\tilde{x}' = [\tilde{x}'_{mN_{\text{len}}+1}, \tilde{x}'_{mN_{\text{len}}+2}, \dots, \tilde{x}'_{(m+1)N_{\text{len}}}]^T$,第 t 个采样点的含噪标准化值 $\tilde{x}'_t = x'_t + n$ 。

3. 如权利要求2所述的基于降噪自编码器的高维受损数据无线传输方法,其特征在于:步骤(2)具体是:将 \tilde{x}'_{samp} 馈入有一层输入层,一层隐藏层和一层输出层的降噪自编码器模型;输入层的神经元个数为 N_{in} ,隐藏层的神经元个数为 N_{hidd} ,输出层的神经元个数为 N_{out} ,满足 $N_{\text{in}} = N_{\text{out}} = N_{\text{len}}$;降噪自编码器模型的激活函数为 \tanh ,优化算法为随机梯度下降法(SGD);隐藏层的输入即为传输的低维信号 y_{samp} ,输出层的输出为模型重构的信号 \hat{x}'_{samp} 。

4. 如权利要求3所述的基于降噪自编码器的高维受损数据无线传输方法,其特征在于:步骤(b)具体是:

一系列噪声类型未知的感知数据的通式为 $\tilde{x}_{t'} = x_{t'} + n_{t'}$, $t' = 1, 2, \dots, N_T$, N_T 表示该数据的总序列长度, t' 时刻未受损感知数据值为 $x_{t'}$, t' 时刻噪声值为 $n_{t'}$;

对 $\tilde{x}_{t'}$ 执行重采样、标准化操作,得到样本大小为 $N_{\text{samp}} \times N_{\text{len}}$ 的待传输数据样本

$\tilde{X}'_{\text{samp}} = [\tilde{x}'_1, \tilde{x}'_2, \dots, \tilde{x}'_{N_{\text{samp}}}]$; \tilde{x}'_m 表示第 m 个标准化数据样本, $m = 1, 2, \dots, N_{\text{samp}}$, 维度为 N_{len} ,操作方法与步骤(1)相同;

对 \tilde{X}'_{samp} 执行降维操作,得到样本大小为 $N_{\text{samp}} \times N_{\text{hidd}}$ 的传输数据样本 $Y_{\text{samp}} = [y_1, y_2, \dots, y_{N_{\text{samp}}}]$;

第 m 个传输数据样本为 $y_m = \tilde{x}_m W^e + b^e$ 。

5. 如权利要求4所述的基于降噪自编码器的高维受损数据无线传输方法,其特征在于:步骤(d)具体是:

解码操作:对接收到的数据样本 Y_{samp} 执行矩阵向量乘法、加法以及非线性激活操作,得到原始标准化感知数据重构样本 $\hat{X}'_{\text{samp}} = [\hat{x}'_1, \hat{x}'_2, \dots, \hat{x}'_{N_{\text{samp}}}]$;第 m 个解码后的原始标准化感知数据重构样本为 $\hat{x}'_m = g(f(y_m)) \cdot W^d + b^d$;

反标准化操作:对解码后的原始标准化感知数据重构样本执行最大最小反标准化操作,获取原始值域内的原始感知数据重构样本 $\hat{X}_{\text{samp}} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{N_{\text{samp}}}]$;第 m 个原始感知数据重构样本值 $\hat{x}_m = \frac{\hat{x}'_m + 1}{2} \times (x_{\text{max}} - x_{\text{min}}) + x_{\text{min}}$;

重组操作:对反标准化后的原始感知数据重构样本按行拼接,获取序列长度为 N_T 的未受损感知数据的重构数据 \hat{x}_t 。

一种基于降噪自编码器的高维受损数据无线传输方法

技术领域

[0001] 本发明涉及信号处理领域,具体涉及一种无线传感器网络中基于降噪自编码器的高维受损数据降维及降噪一体化可靠传输方法。

背景技术

[0002] 随着通信及信息技术的高速发展,信息需求和待处理数据量也随之增加,如何高效地处理大量网络数据并最大限度地降低传输、存储、计算开销就成为了我们亟待解决的一大问题。无线传感器网络已广泛应用于多个领域,诸如环境监测、军事侦察、精准农业等。但由于传感器的体积决定了它们的计算、存储、通信资源都十分有限,无法负荷巨量数据的计算和传输。此外,传感器易受恶劣环境的影响,在感知数据过程中存在噪声与干扰,致使干净数据的获取难度急剧增加。

[0003] 目前,常用的无线传感器网络高维数据可靠数据传输方法主要由数据采集、数据处理和数据重构三部分组成。为避免“维数灾难”问题,目前常用的数据传输方法主要基于压缩感知理论,旨在通过对待传输数据执行降维操作,以降低实际传输量。压缩感知包括字典、观测矩阵和重构算法三个方面。压缩感知主要通过字典和重构算法的设计实现降噪,通过观测矩阵的设计实现降维,并且其应用的前提条件是信号可压缩,主要用于解决线性问题,因此对于愈加复杂的通信环境和现实待解决问题的适应性较低。

[0004] 申请号201910604340.8的中国发明专利申请公开了时序贝叶斯压缩采样及信号解压缩重构方法及数据丢失恢复方法,其局限性在于所述方法涉及测量矩阵、基向量矩阵的设计,以及稀疏性建模、贝叶斯概率求解、超参数的快速优化估计等,设计成本、时间成本及计算成本较高,不适用于资源受限的设备。申请号202010259663.0和202010910559.3的中国发明专利分别申请公开了一种面向压缩感知的稀疏多带信号盲重构方法和一种基于字典双学习的压缩感知信号重构方法及系统,其局限性在于预设了待处理信号是可压缩的,且将信号重构问题转化为线性规划问题,降噪功能只面向特定类型的噪声干扰,不适用于复杂噪声干扰下的非线性数据的降维及降噪重构。申请号202011254453.9的中国发明专利申请公开了一种基于随机解调结构的心电信号压缩采样装置及方法,其局限性在于将心电信号线性映射到低维空间,可能损失关键特征,且没有考虑信号存在噪声干扰的问题。

[0005] 随着机器学习的蓬勃发展,为复杂环境下无线传感器网络实现数据降维和降噪一体化传输提供了新可能。自编码器最初应用于计算机视觉、图像处理及语音处理领域,它能够从受损图像中学习到更为鲁棒的特征,从而重构原始图像,也能够部署于无线传感器网络,减少待传输数据量,进而降低通信、计算和存储成本,并且自编码器的非线性也决定了它能够解决更为复杂的非线性问题。因此,自编码器在网络信号处理方法具备更大的优势,为无线传感器网络中高维受损数据传输实现降维和降噪一体化提供了新思路。

发明内容

[0006] 本发明的目的是针对无线传感器网络部署环境恶劣、干起节点资源有限、待传输

数据量大且受损的情况下,目前网络信号处理方法计算和设计成本较高、受场景限制、精度及压缩率较低的问题,提出一种基于降噪自编码器的高维受损数据无线传输方法。

[0007] 本发明方法包括模型训练和端到端传输。

[0008] 所述的模型训练用于训练降噪自编码器,获得性能良好的传输架构;具体方法是:

[0009] 步骤(1)对历史感知数据集进行数据预处理,具体如下:

[0010] 对数据执行重采样、标准化以及添加噪声操作,组成具有 N_{samp} 个样本的标准化不含噪数据集 D 和含噪数据集 D_n ,样本长度均为 N_{len} ;其中,

[0011] 重采样操作:对历史感知数据使用符号间隔 T_s 进行采样,获取长度为 N_{len} 的离散序列

$x = [x_{mN_{\text{len}}+1}, x_{mN_{\text{len}}+2}, \dots, x_{(m+1)N_{\text{len}}}]^T$, x_t 表示离散序列中第 t 个采样点的离散值, $t = (mN_{\text{len}}+1),$

$(mN_{\text{len}}+2), \dots, (m+1)N_{\text{len}}$, T 表示转置操作, m 表示重采样轮次。

[0012] 标准化操作:对离散序列 x 进行最大最小化处理,获取标准化序列

$x' = [x'_{mN_{\text{len}}+1}, x'_{mN_{\text{len}}+2}, \dots, x'_{(m+1)N_{\text{len}}}]^T$, 每个样本点取值范围是 $(-1, 1)$, 第 t 个采样点的标准化值

$x'_t = 2 \times \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} - 1$, x_{\max} 为离散序列中最大离散值, x_{\min} 为离散序列中最小离散值。

[0013] 添加噪声操作:在标准化序列上叠加随机生成的与多种实际场景相符的噪声 n ,得到含噪的标准化序列 $\tilde{x}' = [\tilde{x}'_{mN_{\text{len}}+1}, \tilde{x}'_{mN_{\text{len}}+2}, \dots, \tilde{x}'_{(m+1)N_{\text{len}}}]^T$, 第 t 个采样点的含噪标准化值 $\tilde{x}'_t = x'_t + n$ 。

[0014] 步骤(2)分割标准化不含噪数据集 D 和含噪数据集 D_n :将不含噪数据集 D 和含噪数据集 D_n 中 N_{samp} 个样本划分为 K 组训练-验证集 $\{D_{\text{train}}, D_{\text{valid}_k}\}_k$, 满足 $N_{\text{train},k} + N_{\text{valid}_k,k} = N_{\text{samp}}$; $N_{\text{train},k}$

为 $\{D_{\text{train}}\}_k$ 的样本数, $N_{\text{valid}_k,k}$ 为 $\{D_{\text{valid}_k}\}_k$ 的样本数, 组号 $k=1, \dots, K$ 。

[0015] 步骤(3)构建降噪自编码器模型,模型输入为 \tilde{x}'_{samp} ,隐层输出为 h_{samp} ,模型输出为 \hat{x}'_{samp} ,已知原始样本为 x'_{samp} 。构建方法具体如下:

[0016] 将 \tilde{x}'_{samp} 馈入有一层输入层,一层隐藏层和一层输出层的降噪自编码器模型。输入层的神经元个数为 N_{in} ,隐藏层的神经元个数为 N_{hidd} ,输出层的神经元个数为 N_{out} ,满足 $N_{\text{in}} = N_{\text{out}} = N_{\text{len}}$ 。降噪自编码器模型的激活函数为 \tanh ,优化算法为随机梯度下降法(SGD)。隐藏层的输入即为传输的低维信号 y_{samp} ,输出层的输出为模型重构的信号 \hat{x}'_{samp} 。

[0017] 步骤(4)训练降噪自编码器模型,具体是:

[0018] (4-1)初始化:设置最大迭代轮数 $N_{\text{max_epoch}}$,当前迭代轮数 $N_{\text{epoch}} = 1$;设置停止准则

中的阈值 thresh ;设置隐藏层的最大神经元个数 $N_{\text{hmax}} = \frac{2}{3}N_{\text{len}}$,当前隐藏层神经元个数 N_{hidd}

$= 1$;使用Xavier初始化方法随机初始化模型权值参数 w 和偏置参数 b ;设置当前训练集的组号为 $k=1$;

[0019] (4-2)分批:将训练集 $\{D_{\text{train}}\}_k$ 样本均匀分成 N_{batch} 个小批;

[0020] (4-3)加噪处理:随机选取一个小批,并生成一个在 $(0, 1)$ 上均匀分布的随机数 r ,将 $r/2$ 作为高斯分布的标准差,根据该标准差生成高斯噪声叠加到该小批上;

[0021] (4-4) 前向传播:将加噪小批中的所有样本 \tilde{x}'_{samp} 输入降噪自编码器模型中,得到输出重构值 \hat{x}'_{samp} ;

[0022] (4-5) 后向传播:根据原始样本 x'_{samp} 和重构样本 \hat{x}'_{samp} 计算损失函数

$$Loss = \frac{1}{I} \sum_{i=1}^I (\hat{x}'_i - x'_i)^2, \hat{x}'_i \text{ 表示该小批中第 } i \text{ 个含噪样本的重构输出值, } x'_i \text{ 表示该小批中第 } i$$

个含噪样本的原始值;根据梯度下降法更新模型参数 $w' = w - \eta \frac{\partial Loss}{\partial w}$, $b' = b - \eta \frac{\partial Loss}{\partial b}$, η 为学习率;

[0023] (4-6) 验证:若所有小批均已选取过,则将含噪验证集 $\{D_{valid_n}\}_k$ 输入模型中,计算不含噪的验证集样本 x'_{samp} 和重构样本 \tilde{x}'_{samp} 的均方误差 MSE_{valid_n} , 并计算其与保存的验证集最小均方误差 $\{MSE_{valid_n}\}_{min}$ 的绝对差值 $\varepsilon = |MSE_{valid_n} - \{MSE_{valid_n}\}_{min}|$; 否则返回步骤 (4-3);

[0024] (4-7) 迭代:若迭代轮数 $N_{epoch} < N_{max_epoch}$ 且 $\varepsilon > \text{thresh}$, 则更新并保存 $\{MSE_{valid_n}\}_{min}$, 返回步骤 (4-2) 进行下一轮迭代, $\{MSE_{valid_n}\}_{min}$ 为本轮保存的最小 MSE_{valid_n} ; 否则结束训练, 保存第 k 组的最佳模型参数 $\{w_{best}, b_{best}\}_{k, N_{hidd}}$;

[0025] 若隐藏层神经元个数 $N_{hidd} < N_{hmax}$, 则返回步骤 (4-1) 进行下一个不同隐藏层神经元个数的模型训练。

[0026] 所述的端到端传输用于将训练得到的降噪自编码器模型部署至两端, 实现数据的降维传输及降噪重构; 具体方法是:

[0027] 步骤 (a) 部署训练得到的降噪编码器:

[0028] 从 K 个降噪自编码器模型选择出最优的模型参数;

[0029] 将降噪自编码器模型的编码阶段的超参数 w^e 和 b^e 部署于发送端;

[0030] 将编码阶段的非线性激活函数 f 和解码阶段的超参数 w^d 和 b^d , 以及非线性激活函数 g 部署于接收端。

[0031] 步骤 (b) 在发送端处理噪声类型未知的感知数据:

[0032] 一系列噪声类型未知的感知数据的通式为 $\tilde{x}_t = x_t + n_t$, $t' = 1, 2, \dots, N_T$, N_T 表示该数据的总序列长度, t' 时刻未受损感知数据值为 $x_{t'}$, t' 时刻噪声值为 $n_{t'}$;

[0033] 对 \tilde{x}_t 执行重采样、标准化操作, 得到样本大小为 $N_{smp} \times N_{len}$ 的待传输数据样本

$\tilde{X}'_{smp} = [\tilde{x}'_1, \tilde{x}'_2, \dots, \tilde{x}'_{N_{smp}}]$; \tilde{x}'_m 表示第 m 个标准化数据样本, $m = 1, 2, \dots, N_{smp}$, 维度为 N_{len} , 操作方法与步骤 (1) 相同;

[0034] 对 \tilde{X}'_{smp} 执行降维操作, 得到样本大小为 $N_{smp} \times N_{hidd}$ 的传输数据样本

$$Y_{smp} = [y_1, y_2, \dots, y_{N_{smp}}]; \text{ 第 } m \text{ 个传输数据样本为 } y_m = \tilde{x}'_m W^e + b^e.$$

[0035] 步骤 (c) 将降维后的传输数据样本 Y_{smp} 传输至接收端。

[0036] 步骤(d)在接收端重构未受损感知数据:

[0037] 对接收数据样本 Y_{samp} 执行解码、反标准化以及重组操作,得到未受损感知数据的重构数据 \hat{x}_r ;其中,

[0038] 解码操作:对接收到的数据样本 Y_{samp} 执行矩阵向量乘法、加法以及非线性激活操作,得到原始标准化感知数据重构样本 $\hat{X}'_{\text{samp}} = [\hat{x}'_1, \hat{x}'_2, \dots, \hat{x}'_{N_{\text{samp}}}]$;第m个解码后的原始标准化感知数据重构样本为 $\hat{x}'_m = g(f(y_m)) \cdot W^d + b^d$;

[0039] 反标准化操作:对解码后的原始标准化感知数据重构样本执行最大最小反标准化操作,获取原始值域内的原始感知数据重构样本 $\hat{X}_{\text{samp}} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{N_{\text{samp}}}]$;第m个原始感知数据重构样本值 $\hat{x}_m = \frac{\hat{x}'_m + 1}{2} \times (x_{\text{max}} - x_{\text{min}}) + x_{\text{min}}$;

[0040] 重组操作:对反标准化后的原始感知数据重构样本按行拼接,获取序列长度为 N_T 的未受损感知数据的重构数据 \hat{x}_r 。

[0041] 本发明以传统降噪自编码器模型为传输机制的基础架构,引入分批添加随机高斯噪声的训练方式,改善了模型对多种噪声的泛化能力和鲁棒性,方便在资源受限的无线传感器节点中部署,辅助无线传感器网络进行可靠且高效的数据传输。与现有的无线数据传输方法相比。

[0042] 本发明具有如下有益效果:

[0043] (1) 现有的基于压缩感知理论的传输方法需要对信号做可压缩性假设,降噪性能取决于噪声的先验信息情况,降维性能受算法设计的限制。而本发明能规避这些假设条件,无需噪声的先验信息就能对抵抗多种扰动,且能最大限度地降低传输数据的维度。

[0044] (2) 现有的基于机器学习的传输方法仅单方面考虑降维或降噪功能,且其降噪性能的优越性取决于模型的深度和复杂性,对实际部署的设备具有较高的要求。而本发明使用了最简单的三层结构的降噪自编码器,易于在资源受限的设备上进行部署,可以实现降维及降噪一体化传输。

[0045] (3) 现有的传输方法并没有考虑多种类型噪声存在下的传输问题,其降噪功能有一定局限性。而本发明提出了一种新型加噪训练模式,提高了降噪自编码器的鲁棒性和对噪声的泛化能力,能有效地传输多种噪声干扰下的感知数据并从中降噪重构。

附图说明

[0046] 图1为本发明方法整体框架示意图;

[0047] 图2为本发明方法中的模型训练流程图;

[0048] 图3为本发明方法中的降噪自编码器模型的示意图;

[0049] 图4为本发明方法中降噪自编码器模型的训练流程图;

[0050] 图5为本发明方法中的端到端传输流程图。

具体实施方式

[0051] 以下结合附图并举实施例对本发明作进一步详细说明。

[0052] 本发明针对端到端传输情景中,待传输感知数据在复杂环境下受多种噪声干扰且数据量较大的问题。提出了一种基于降噪自编码器的高维受损数据无线传输方法,用于感知数据的降维传输和降噪重构。并设计降噪自编码器模型训练阶段的新型噪声引入机制,使得降噪自编码器可以在模型训练过程中充分学习数据的本质特征和噪声的特征,对多种噪声具有鲁棒性,最大限度地降低传输数据量,并从含噪数据中重构出原始数据。

[0053] 本方法的高维受损数据无线传输架构如图1所示,其中降噪自编码器模型主要由编码器和解码器两个模块构成。编码器的线性模块用于发送端处高维数据的低维映射,该部分的降维处理模块主要由编码阶段的权值参数 W^e 和偏置参数 b^e 组成。模型编码器的非线性模块以及解码器模块用于接收端数据的降噪重构,主要由隐藏层的激活函数 f 、解码阶段的参数 W^d 和 b^d 以及输出层的激活函数 g 组成。

[0054] 本实施例以Intel Berkeley Lab提供的监测数据集中2004年3月1日至3月12日的来自Mote1的温度数据为例,时间序列总长度为21720,对基于降噪自编码器的高维受损数据无线传输方法进行说明。发明方法包括模型训练和端到端传输。

[0055] 如图2,模型训练具体是:

[0056] 步骤(1)对历史的温度数据集进行数据预处理,具体如下:

[0057] 对温度数据执行重采样、标准化以及添加噪声操作,组成具有 N_{samp} 个(本实施例中 $N_{\text{samp}}=362$)样本的标准化不含噪温度数据集 D 和含噪温度数据集 D_n ,样本长度均为 N_{len} (本实施例中 $N_{\text{len}}=60$)。

[0058] 步骤(2)分割标准化不含噪温度数据集 D 和含噪温度数据集 D_n ,具体是:

[0059] 基于 K -折(本实施例中 $K=5$)交叉验证的思想将不含噪温度数据集 D 和含噪温度数据集 D_n 中 N_{samp} 个样本都进行 K 等分,有 $\{D_1, D_2, \dots, D_K\}$ 和 $\{D_{n,1}, D_{n,2}, \dots, D_{n,K}\}$ 。不含噪训练集 $\{D_{\text{train}}\}_k$ 用于模型的训练,含噪测试集 $\{D_{\text{test}_n}\}_k$ 和不含噪测试集 $\{D_{\text{test}}\}_k$ 用于检验传输方法的有效性;第 k 组训练-测试集为 $\{D_{\text{train}}; D_{\text{test}_n}\}_k = \{D_1, \dots, D_{k-1}, D_{k+1}, \dots, D_K; D_{n,k}\}$,组号 $k=1, 2, 3, 4, 5$;前三组中 $\{D_{\text{train}}\}_k$ 的样本个数为 $N_{\text{train},k} = 290$, $\{D_{\text{test}_n}\}_k$ 的样本个数为 $N_{\text{test}_n,k} = 72$, $k=1, 2, 3$,后两组中 $\{D_{\text{train}}\}_k$ 的样本个数为 $N_{\text{train},k} = 289$, $\{D_{\text{test}_n}\}_k$ 的样本个数为 $N_{\text{test}_n,k} = 73$, $k=4, 5$ 。

[0060] 步骤(3)构建降噪自编码器模型,模型输入为 \tilde{x}'_{samp} ,隐层输出为 h_{samp} ,模型输出为 \hat{x}'_{samp} ,已知原始样本为 x'_{samp} 。构建方法具体如下:

[0061] 将 \tilde{x}'_{samp} 馈入如图3所示的有一层输入层,一层隐藏层和一层输出层的降噪自编码器模型。输入层的神经元个数为 N_{in} (本实施例中 $N_{\text{in}}=60$),隐藏层的神经元个数为 N_{hidd} (本实施例中 $N_{\text{hidd}}=20$),输出层的神经元个数为 N_{out} (本实施例中 $N_{\text{out}}=60$),满足 $N_{\text{in}}=N_{\text{out}}=N_{\text{len}}$ 。降噪自编码器模型的激活函数为 \tanh ,优化算法为随机梯度下降法(SGD)。隐藏层的输入即为传输的低维信号 y_{samp} ,输出层的输出为模型重构的信号 \hat{x}'_{samp} 。

[0062] 步骤(4)如图4所示,训练降噪自编码器模型,具体是:

[0063] (4-1)初始化:设置最大迭代轮数 $N_{\text{max_epoch}}=1500$,当前迭代轮数 $N_{\text{epoch}}=1$;设置停止准则中的阈值 $\text{thresh}=1e-6$;使用Xavier初始化方法随机初始化模型权值参数 w 和偏置参数 b ;设置当前训练集的组号为 $k=1$,总组数为 $K=5$;

[0064] (4-2) 分批: 将训练集 $\{D_{\text{train}}\}_k$ 样本均匀分成 N_{batch} 个 (本实施例中 $N_{\text{batch}} = \frac{290}{34} = \frac{289}{34} = 9$) 小批, 随机选取部分小批 (本实例中选取1个小批) 作为验证集 $\{D_{\text{valid}}\}_k$, 并叠加与含噪数据集同一分布的噪声, 生成对应的含噪验证集 $\{D_{\text{valid}_n}\}_k$;

[0065] (4-3) 加噪处理: 随机选取一个小批, 并生成一个在 $(0, 1)$ 上均匀分布的随机数 r , 将 $r/2$ 作为高斯分布的标准差, 根据该标准差生成高斯噪声叠加到该小批上;

[0066] (4-4) 前向传播: 将加噪小批中的所有样本 \tilde{x}'_{samp} 输入降噪自编码器模型中, 得到输出重构值 \hat{x}'_{samp} ;

[0067] (4-5) 后向传播: 根据原始样本 x'_{samp} 和重构样本 \hat{x}'_{samp} 计算损失函数 $Loss = \frac{1}{I} \sum_{i=1}^I (\hat{x}'_i - x'_i)^2$, \hat{x}'_i 表示该小批中第 i 个含噪样本的重构输出值, x'_i 表示该小批中第 i 个含噪样本的原始值; 根据梯度下降法更新模型参数 $w' = w - \eta \frac{\partial Loss}{\partial w}$, $b' = b - \eta \frac{\partial Loss}{\partial b}$, η 为学习率;

[0068] (4-6) 验证: 若所有小批均已选取过, 则将含噪验证集 $\{D_{\text{valid}_n}\}_k$ 输入模型中, 计算不含噪的验证集样本 x'_{samp} 和重构样本 \hat{x}'_{samp} 的均方误差 MSE_{valid_n} , 并计算其与保存的验证集最小均方误差 $\{MSE_{\text{valid}_n}\}_{\min}$ 的绝对差值 $\varepsilon = |MSE_{\text{valid}_n} - \{MSE_{\text{valid}_n}\}_{\min}|$; 否则返回步骤 (4-3);

[0069] (4-7) 迭代: 若迭代轮数 $N_{\text{epoch}} < 1500$ 且 $\varepsilon > 1e-6$, 则更新并保存 $\{MSE_{\text{valid}_n}\}_{\min}$, 更新 $N_{\text{epoch}} = N_{\text{epoch}} + 1$, 返回步骤 (4-2) 进行下一轮迭代, $\{MSE_{\text{valid}_n}\}_{\min}$ 为本轮保存的最小 MSE_{valid_n} ; 否则结束训练, 保存第 k 组的最佳模型参数 $\{w_{\text{best}}, b_{\text{best}}\}_{k, N_{\text{hid}}}$ 。

[0070] 如图5, 所述的端到端传输具体是:

[0071] 步骤 (a) 部署训练得到的降噪编码器:

[0072] 从5个降噪自编码器模型选择出最优的模型参数;

[0073] 将降噪自编码器模型的编码阶段的超参数 w^e 和 b^e 部署于发送端;

[0074] 将编码阶段的非线性激活函数 f 和解码阶段的超参数 w^d 和 b^d , 以及非线性激活函数 g 部署于接收端。

[0075] 步骤 (b) 在发送端处理噪声类型未知的感知数据:

[0076] 一系列噪声类型未知的感知数据的通式为 $\tilde{x}_t = x_t + n_t$, $t' = 1, 2, \dots, N_T$ (本实例中 $N_T = 8430$, 噪声 n_t 服从多种类型的噪声分布);

[0077] 对 \tilde{x}_t 执行重采样、标准化操作, 得到样本大小为 $N_{\text{samp}} \times N_{\text{len}}$ (本实例中 $N_{\text{samp}} = 73$, $N_{\text{len}} = 60$) 的待传输数据样本 $\tilde{X}'_{\text{samp}} = [\tilde{x}'_1, \tilde{x}'_2, \dots, \tilde{x}'_{N_{\text{samp}}}]$; 操作方法与步骤 (1) 相同;

[0078] 对 \tilde{X}'_{samp} 执行降维操作, 得到样本大小为 $N_{\text{samp}} \times N_{\text{hid}}$ (本实例中 $N_{\text{hid}} = 20$) 的传输数据样本 $Y_{\text{samp}} = [y_1, y_2, \dots, y_{N_{\text{samp}}}]$; 操作方法与步骤 (1) 相同。

[0079] 步骤 (c) 将降维后的传输数据样本 Y_{samp} 传输至接收端。

[0080] 步骤(d)在接收端重构未受损感知数据:

[0081] 对接收数据样本 Y_{samp} 执行解码、反标准化以及重组操作,得到未受损感知数据的重构数据 \hat{x}_r ,操作方法与步骤(d)相同。

[0082] 上述实例所述的内容仅仅是对本发明实现形式的列举,本发明的保护范围不应限于实施例所陈述的具体形式,本发明的保护范围也应包含在本发明基础上所构思的同类发明方法。

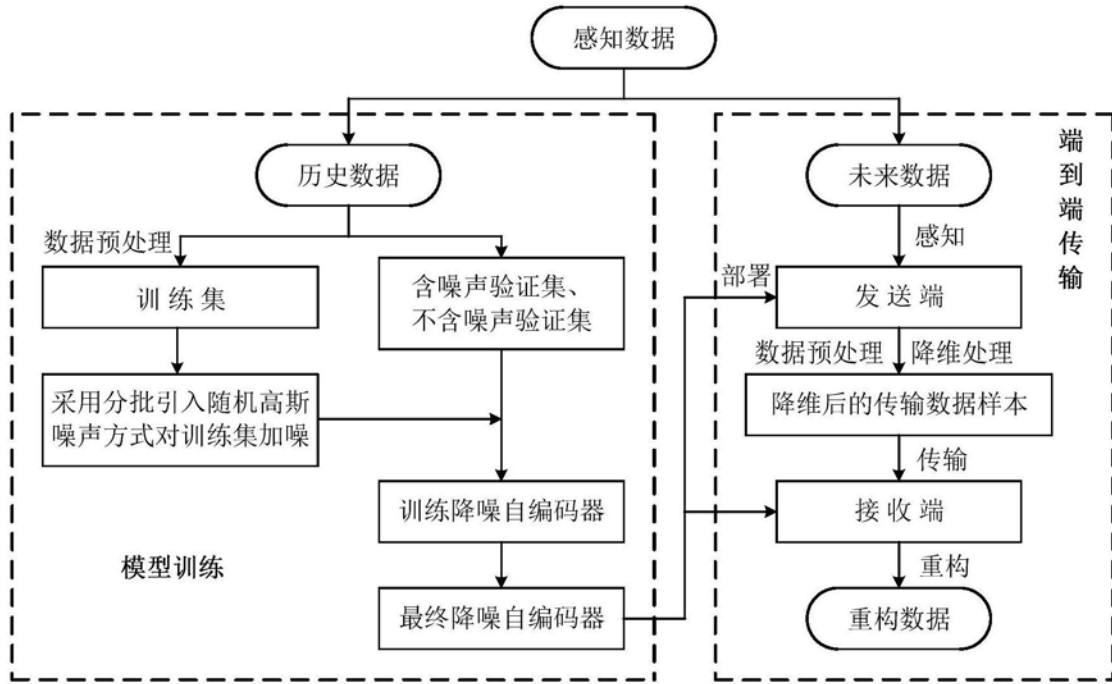


图1

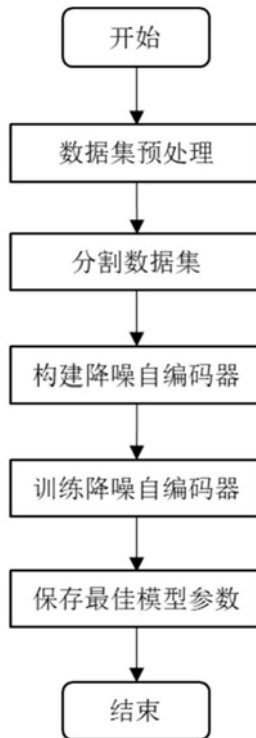


图2

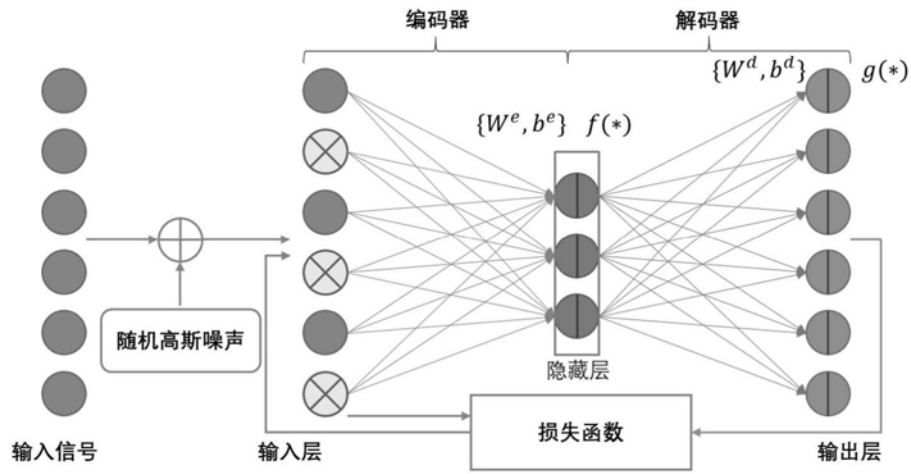


图3

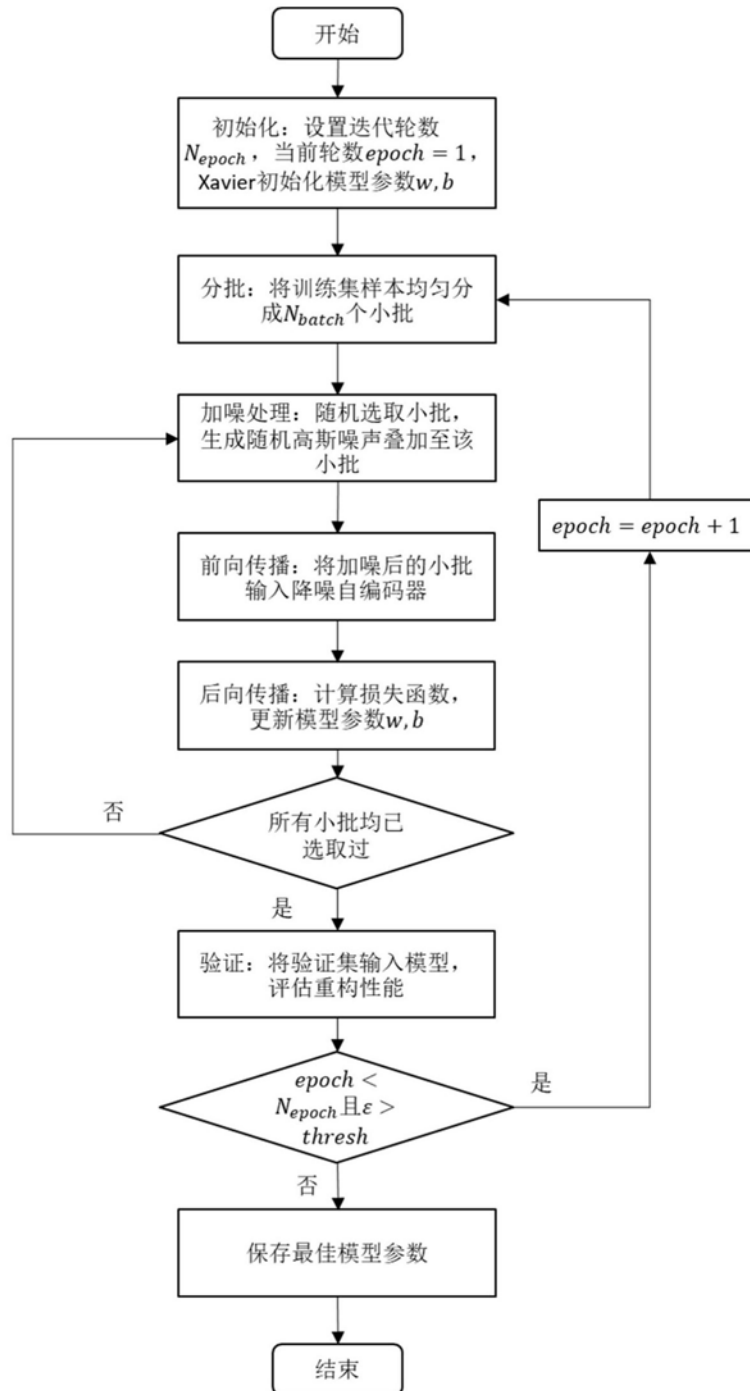


图4

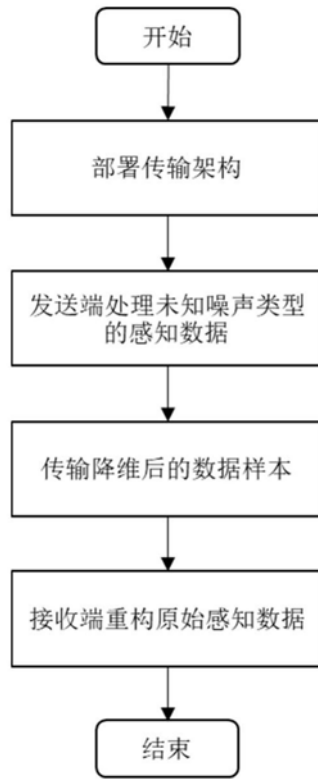


图5