

基于卷积神经网络的缺失数据填充方法

张网娟, 许国艳, 李敏佳, 朱 帅

(河海大学 计算机与信息学院, 江苏 南京 211100)

摘 要: 受不利因素影响, 传感器网络中的数据不可避免地存在缺失, 阻碍上层应用对数据的分析与处理. 使用能够自主提取和学习有效特征结构的卷积神经网络对缺失数据问题进行研究, 提出一种基于卷积神经网络的缺失数据填充方法. 首先, 分别针对时间序列数据的时间相关性和传感器节点间的空间相关性, 使用卷积神经网络填充模型对缺失数据进行单维度相关性的填充. 然后, 根据时间维度和空间维度的填充结果, 考虑时空相关性对缺失数据进行填充. 最后, 在真实数据集上进行实验验证, 实验结果表明考虑时空多维度相关性的填充结果优于只考虑单维度相关性的填充结果, 并与 BP 模型进行对比, 验证了卷积神经网络填充模型的有效性.

关键词: 缺失数据; 填充; 卷积神经网络; 时空相关性

Missing data imputation approach based on convolutional neural network

ZHANG Wang-juan, XU Guo-yan, LI Min-jia, ZHU Shuai

(College of Computer and Information, Hohai University, Nanjing 211100, China)

Abstract: In the sensor network, the missing data is inevitable due to many of interference factors, which hinders data analysis and processing. In this research, by using the convolution neural network(CNN) which can extract and learn effective features data independently, a missing data imputation approach based on convolution neural network was proposed. Firstly, for the temporal correlation of time series and the spatial correlation between sensor nodes, the CNN imputation model was used to estimate the missing data with single dimensional correlation. Then, based on the imputation results of temporal dimension and spatial dimension, the missing data were imputed by considering the multidimensional correlation. Finally, experimental verification is performed on the real data set. Experiment results show that the imputation result considering the temporal and spatial multidimensional correlation is better than the imputation result considering the single dimensional correlation. And compared with the BP model, the validity of the convolution neural network imputation model is verified.

Key words: missing data; imputation; convolution neural network(CNN); temporal and spatial correlation

作者简介:

张网娟 女, (1992-), 硕士研究生. 研究方向为大数据、数据管理.

E-mail: zhang_wang_juan@foxmail.com.

许国艳 女, (1971-), 博士, 副教授. 研究方向为大数据、数据起源、数据管理.

李敏佳 女, (1994-), 硕士研究生. 研究方向为社交网络、数据管理.

朱 帅 男, (1992-), 硕士研究生. 研究方向为社会网络、数据管理.