

数据库、信号与信息处理

基于EM和贝叶斯网络的丢失数据填充算法

李宏, 阿玛尼, 李平, 吴敏

中南大学 信息科学与工程学院, 长沙 410083

收稿日期 2008-8-19 修回日期 2008-11-3 网络版发布日期 2010-2-8 接受日期

摘要 实际应用中存在大量的丢失数据的数据集, 对丢失数据的处理已成为目前分类领域的研究热点。分析和比较了几种通用的丢失数据填充算法, 并提出一种新的基于EM和贝叶斯网络的丢失数据填充算法。算法利用朴素贝叶斯估计出EM算法初值, 然后将EM和贝叶斯网络结合进行迭代确定最终更新器, 同时得到填充后的完整数据集。实验结果表明, 与经典填充算法相比, 新算法具有更高的分类准确率, 且节省了大量开销。

关键词 [丢失数据填充](#) [参数更新器](#) [最大期望值算法 \(EM\)](#) [贝叶斯网络](#)

分类号 [TP301](#)

Imputation algorithm of missing values based on EM and Bayesian network

LI Hong, EMMANUEL Amani, LI Ping, WU Min

School of Information Science and Engineering, Central South University, Changsha 410083, China

Abstract

Dataset with missing values is quite common in real applications, and handling missing values has become a research hot issue in the classification field. This paper analyzes and compares several popular missing values imputation algorithms, and has proposed a novel imputation algorithm for missing values based on EM (Expectation Maximization) and Bayesian network. In this algorithm, the Naïve Bayesian is employed to estimate the initial values of EM algorithm, and the EM inspired approach for filling up missing values is incorporated to Bayesian network learning with the objective of ensuring the ultimate updater. As a result, the complete dataset is got after imputation. Experiment results demonstrate that the proposed algorithm enables much higher classification accuracy and lower cost when compared with other classical imputation algorithm.

Key words [missing values imputation](#) [parameter updater](#) [Expectation-Maximization \(EM\)](#) [Bayesian network](#)

DOI: 10.3778/j.issn.1002-8331.2010.05.037

通讯作者 李宏

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(516KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [参考文献](#)

服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [复制索引](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

相关信息

- ▶ [本刊中 包含“丢失数据填充”的相关文章](#)
- ▶ [本文作者相关文章](#)

- [李宏](#)
- [阿玛尼](#)
- [李平](#)
- [吴敏](#)