Chemometrics and Intelligent Laboratory Systems 177 (2018) 1–7

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Noisy matrix completion on a novel neural network framework

Samuel Mercier [a,b,*], Ismail Uysal [a]

[a] University of South Florida, Department of Electrical Engineering, 4202 East Fowler Avenue, Tampa, FL, 33620, USA
[b] Université de Sherbrooke, Department of Chemical and Biotechnological Engineering, 2500 De L'Université Boulevard, Sherbrooke, Quebec, J1K 2R1, Canada

ARTICLE INFO

Keywords:
Matrix completion
Noisy matrices
Neural network
Principal component analysis (PCA)
Trimmed scores regression (TSR)

ABSTRACT

A novel matrix completion algorithm based on the iterative application of neural networks is presented. It is shown that Bayesian regularization provides proper protection against overfitting, more so than early-stopping or a combination of both. The flexibility to increase the size of the hidden layer provides a better description of increasingly nonlinear relationships between the known and missing values in the data with a limited loss in generalization ability. The proposed neural network algorithm provides a more accurate estimation of missing values than current matrix completion algorithms based on iterative regression approaches or PCA applications for many datasets with fractions of missing values from 5 to 40%. The neural network algorithm performs particularly well on datasets where the number of observations significantly exceeds the number of features.

1. Introduction

Matrix completion is the general problem of estimating the missing values in a dataset from the known values. The presence of missing values in a dataset is common and can be caused by several factors, such as hardware failure, human error and data corruption, to name just a few. When the number of missing values is minimal and the missing values are distributed randomly, the incomplete observations can be removed to recreate a complete (but slightly smaller) dataset. However, if missing values are frequent or concentrated in certain regions of the dataset, the information loss resulting from the removal of the incomplete observations may become too significant for subsequent statistical inference, and estimation of these missing values is required. As such, matrix completion has a wide spectrum of applications, including new product development and product characterization [1], meteorological data [2], wastewater treatment [3], gene expression profiles [4,5], recommender systems [6,7], seismic data [8], traffic flow [9], image recovery [10] and video editing [11].

Matrix completion is generally presented as a rank-minimization problem. More specifically, a dataset $M \in \mathbb{R}^{n \times p}$ of rank $r < \min(p, n)$ can be approximated by a matrix $Y \in \mathbb{R}^{n \times p}$ minimizing the following optimization problem [12]:

$$\min rank(Y) \tag{1a}$$

$$\text{Subject to } Y_{ij} = M_{ij} \text{ for all } ij \in \Omega \tag{1b}$$

where Ω is the set of known elements ij of M. For noisy datasets, which is common in applications such as new product development and product characterization, each known value contains a noise component of a magnitude dependent on the precision of the measurement method:

$$M_{ij} = A_{ij} + E_{ij} \text{ for all } ij \in \Omega \tag{2}$$

where $A \in \mathbb{R}^{n \times p}$ is the matrix of "true" values and $E \in \mathbb{R}^{n \times p}$ is the matrix of noise. The completion of a noisy matrix is a more challenging problem, given that the underlying signal (A) and noise (E) components are unknown. To mitigate the impact of noise on the estimation of the missing values, the equality constraint in the rank-minimization problem (Eq. (1b)) is generally replaced by an inequality constraint allowing some level of deviation from the noisy matrix ($|Y_{ij} - M_{ij}| < \delta$ for all $ij \in \Omega$, where δ is the permissible deviation) [13], or the rank-minimization problem is replaced by a principal component analysis (PCA) application [6] or a regression-based approach [14,15].

Recently, Folch-Fortuny et al. [14] developed an iterative regression-based approach to perform matrix completion of noisy matrices (Fig. 1). In this approach, observations are removed one-by-one from the dataset. The observation is separated into two vectors, one containing the p_1 known values of the observation ($o_I \in \mathbb{R}^{p_1}$) and the other containing the missing values ($o_D \in \mathbb{R}^{p-p_1}$). The rest of the dataset

* Corresponding author. Université de Sherbrooke, Department of Chemical and Biotechnological Engineering, 2500 De L'Université Boulevard, Sherbrooke, Quebec, J1K 2R1, Canada.
E-mail addresses: samuel.mercier@USherbrooke.ca, samuelm@mail.usf.edu (S. Mercier).

https://doi.org/10.1016/j.chemolab.2018.04.001
Received 27 July 2017; Received in revised form 24 March 2018; Accepted 1 April 2018
Available online 6 April 2018

Nomenclature

A	$n \times p$ matrix of "true" values
c	cost function of neural network training
E	$n \times p$ matrix of noise
k	size of the neural network hidden layer
M	$n \times p$ sparse matrix to be completed
M_f	$n \times p$ matrix of missing value estimates
n	number of observations in M
o_D	$1 \times (p - p_1)$ vector of missing values in an observation
o_I	$1 \times p_1$ vector of known features in an observation
p	number of features in M
p_1	number of known features in an observation
PC	principal component
PCA	principal component analysis
r	rank of M
MSE	mean squared error
T	$n \times p$ matrix of missing values estimated during neural network training
TSR	trimmed scores regression without early stopping
TSRE	trimmed scores regression with early stopping
VBPCA	variational Bayesian principal component analysis
w_1	$1 \times (p_1 + 1)k$ vector of weights connecting the neural network input and hidden layers
w_2	$1 \times (k + 1)(p - p_1)$ vector of weights connecting the neural network hidden and output layers
Z_D	$n \times (p - p_1)$ matrix of dependent features
Z_I	$n \times p_1$ matrix of independent features

Greek symbols

α	hyper parameter in the neural network training cost function
β	hyper parameter in the neural network training cost function
Ω	set of known values ij in the matrix

Subscripts

ij	element ij of the matrix
test	test

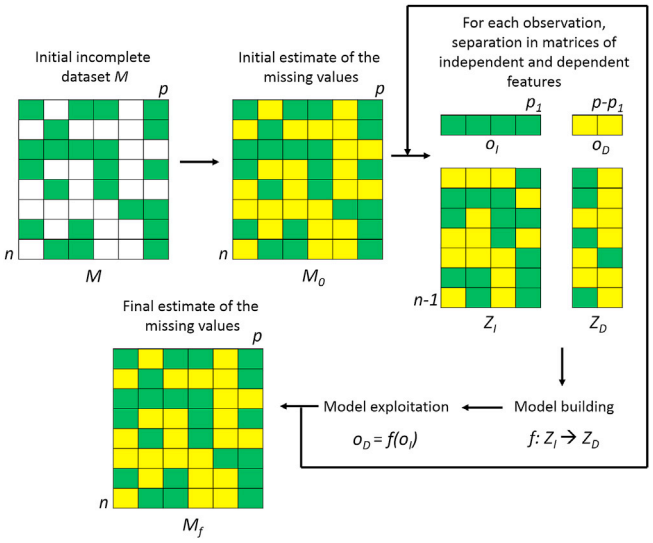

Fig. 1. Illustration of the iterative regression-based approach for matrix completion developed by Folch-Fortuny et al. [14].

is separated into matrices of independent ($Z_I \in \mathbb{R}^{(n-1) \times p_1}$) and dependent ($Z_D \in \mathbb{R}^{(n-1) \times (p- p_1)}$) features, according to the known and missing values in the removed observation, respectively. The matrices of independent and dependent features are used to train a model f, which is then used to predict the missing values of the removed observation. The accuracy of the iterative regression-based approach is strongly dependent on the selection of the model f. Linear mapping with the model coefficients estimated from the matrices of dependent and independent features preconditioned with different PCA-based approaches has been investigated. The trimmed scores regression (TSR) approach, where the independent variables of the linear map are derived from the scores of the augmented matrix $Z = [Z_I\ Z_D]$ and Z is calculated only considering the weight attributed in each principal component loading to the p_1 known features, was shown to provide accurate results for a number of datasets [14].

However, the performance of matrix completion by the current iterative regression-based approach is bounded by the strong hypothesis of a linear map, which can significantly distort the relationship between the dependent and the independent features. The relationship between the dependent and independent features depends on the observation removed from the dataset, and is generally unknown before the application of matrix completion. A desirable regression-based matrix completion algorithm would thus relax the necessity to rely on specific assumptions about the underlying relationships between the dependant and independent features and provide enough flexibility to describe a different, often nonlinear relationship for each observation removed from the dataset. The significant overfitting problems that easily arises when using limited parametric nonlinear maps should also be avoided. As such, we believe that neural networks are well-positioned to improve the problem of matrix completion with their well-demonstrated ability to reproduce a variety of linear and nonlinear relationships from the same neural network architecture (number of hidden layers and neurons per layer). Furthermore, protection against overfitting can be included in neural network training, typically in the form of regularization or early-stopping.

In this work, we propose to combine the iterative regression-based matrix completion approach of Folch-Fortuny et al. [14] (Fig. 1) with a mapping process based on neural networks. The proposed neural network algorithm is detailed in section 2.1. Section 2 also describes the four datasets considered in this work, as well as current matrix completion algorithms against which the neural network approach is compared. In section 3, the contribution of early-stopping and Bayesian regularization is investigated, the accuracy of the neural network approach is compared with state-of-the-art matrix completion algorithms, and the impact of the number of observations on the accuracy of the missing value estimates is discussed. Finally, section 4 summarises the key findings of this work.

2. Methods

2.1. Proposed approach

The matrix completion approach proposed in this work consists of introducing a neural network in the model building step of the iterative regression-based approach developed by Folch-Fortuny et al. [14] (Fig. 1). The initial estimate of the missing values in the incomplete dataset was the feature-wise average of the known values. Before the algorithm was performed, each column (feature) of the dataset was centered and scaled. During processing, the observations were removed one-by-one from the dataset and, according to the features known and