



Imputation of Missing Data in Multidimensional Retail Data Sets (20110121, Dr. Arindam Banerjee)

IP Status: Issued US Patent; **Application #:** 13/204,237

Multiple Predictions of missing data

A new system provides multiple imputation of missing data elements in retail data sets used for modeling and decision support applications. The method is based on the multi-dimensional, tensor structure of the data sets and offers a fast, scalable scheme suitable for large data sets. The computer-implemented algorithm first identifies a retail data set (a measured quantity of products “p” over multiple time periods “t” and multiple retail chains and stores “s”). It then encodes dummy variables corresponding to the missing data for relevant (p,s,t) combinations and provides multiple imputations of the missing data to obtain a plurality of “complete” data sets for demand modeling.

Based on multi-dimensional, tensor structure of data sets

Current methods for imputing missing data elements in retail data sets face some limitations. First, missing data elements are typically replaced by certain point estimates for their relevant imputed values. Therefore, the resulting data set does not capture the natural variability it would have if missing data had been actually recorded instead of being imputed. This omission leads to statistical bias in subsequent analyses. Second, current procedures typically ignore data correlations along the various data set dimensions—or may only consider these correlations along a single dimension. By simultaneously considering multi-dimensional dependencies and correlations in the retail data set, much greater accuracy and statistical reliability can be obtained. This new approach applies the multi-dimensional, tensor structure of data sets to provide multiple imputation of missing data elements in retail data sets.

Phase of Development

- Prototype developed.

Benefits

- Fast, scalable scheme suitable for large data sets

Features

- Multiple imputation of missing data elements in retail data sets
- Based on the multi-dimensional, tensor representation of data sets
- Obtains a plurality of “complete” data sets for demand modeling

Applications

- Predicting missing data elements or values in retail data sets
- Multiple imputation of missing data elements

Technology ID

20110121

Category

Software & IT/Algorithms
Software & IT/Databases
Software & IT/Simulation &
Modeling

Learn more



Researchers

Arindam Banerjee, PhD

Professor, Computer Science and Engineering

[External Link](https://experts.umn.edu) (experts.umn.edu)

Publications

[*Probabilistic Tensor Factorization for Tensor Completion*](#)

Technical Report, TR 11-026

Interested in Licensing?

The University relies on industry partners to integrate software for commercial purposes. The license is available for this technology and would be for the integration, sale, manufacture or use of products claimed by the issued patent. Please contact us to share your business needs and technical interest in this technology and if you are interested in licensing the technology for further research and development.