

# Apply Rough Set Theory to Imputation of Missing Data

Gongzhu Hu  
Department of Computer Science  
Central Michigan University  
Mount Pleasant, Michigan, USA  
hu1g@cmich.edu

## Abstract

Data science is at the forefront of technological advances that is reshaping the way we live today. It explores the scientific approaches to prepare, analyze, and understanding data. One of the primitive elements in the study of this field is **data**. Data must be analyzable, that is, data need to be clean to meet the assumption made by analysis algorithms and tools. One of the major problems with data is “missing data.” Many methods have been developed to address this problem, including imputation of the missing values. Rough set theory provides an elegant way for dealing with missing data.

In rough set theory, data is usually stored in an information table with attributes divided into condition attributes and decision attribute. Due to the uncertainty in the data, The data set is represented by formal approximations and “condition  $\rightarrow$  decision” rules can be deduced from the approximations. In this talk, a new approach to the *missing data imputation problem* is presented. In this approach, attributes are rearranged such that the attribute with missing values becomes the decision attribute so that decision rules deduced can be used to determine the missing values. For this purpose, the notion of optimal logic attribute and optimal attribute logical flow are introduced based on the roughness of rearrangements to explore the logical causal relations between attributes.

## Bibliography

Dr. Gongzhu Hu received BS in numeric analysis from Tsinghua University, China, MS in computer science from the University of Wisconsin-Madison, and PhD in computer science from Michigan State University. He joined the Computer Science Department at Central Michigan University in 1987 and is currently a professor of the department. He was the department chair from 1994 to 2007. His research interests include data mining, databases, distributed systems, and formal methods for software components. He has published over 150 papers in refereed journals and conference proceedings. Dr. Hu has served as the conference chair or program chair of many international conferences, and a member of the editorial board of the several international journals. He is a member of ACIS, ACM, a senior member of IEEE, and a senior member of ISCA. Dr. Hu served as the President of ACIS from 2010–2017, and a member of Board of Directors of ACIS and ISCA.