

พัชรารวรรณ ชินโรสง : การออกแบบและพัฒนาเทคนิคไฮบริดสำหรับการเติมค่าข้อมูลที่
สูญหาย (THE DESIGN AND DEVELOPMENT OF A HYBRID TECHNIQUE FOR
MISSING VALUE IMPUTATION) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.นิตยา
เกิดประสพ, 83 หน้า.

ปัจจุบันการทำเหมืองข้อมูลมีความสำคัญและเป็นประโยชน์ในหลาย ๆ ด้านเช่น การสร้าง
โมเดลเพื่อทำนายด้านการแพทย์ ด้านการศึกษา ด้านธุรกิจ ด้านวิศวกรรม และด้านอื่น ๆ แต่ถ้าชุด
ข้อมูลที่มีข้อมูลที่สูญหายจะมีผลกระทบต่อโมเดลที่ถูกสร้างขึ้นมาเพื่อการทำนาย ถ้านำโมเดลนั้น
ไปใช้ในด้านการแพทย์ เช่น การสร้างโมเดลเพื่อทำนายโรคให้กับผู้ป่วยจะทำให้โมเดลมีค่าความ
ถูกต้องต่ำและการทำนายผลอาจผิดพลาดจะทำให้ผู้ป่วยเสียชีวิตได้ ถ้าในด้านธุรกิจอาจทำให้เสีย
ทรัพย์สินถึงขั้นล้มละลายได้ งานวิจัยนี้จึงออกแบบเทคนิคมาเพื่อช่วยเติมค่าให้กับข้อมูลที่สูญ
หายซึ่งจะทำให้โมเดลที่ได้มีความถูกต้องและแม่นยำขึ้น การพัฒนาและทดสอบอัลกอริทึมจะใช้
การเขียนโปรแกรมด้วยภาษาอาร์ซึ่งเป็นภาษาเชิงฟังก์ชันที่นิยมใช้ในด้านสถิติ การออกแบบ
โปรแกรมเติมค่าให้กับข้อมูลสูญหาย จะให้โปรแกรมทำการวิเคราะห์หาเทคนิคที่เหมาะสมสำหรับ
การเติมค่าให้กับข้อมูลสูญหายที่จะทำให้ประสิทธิภาพของโมเดลดีขึ้นโดยการตรวจสอบด้วย
วิธีการสร้างโมเดลของตนไม่ตัดสินใจด้วยข้อมูลที่เตรียมไว้สำหรับฝึกสอน และจะตรวจสอบค่า
ความถูกต้องด้วยข้อมูลที่เตรียมไว้สำหรับทดสอบ ซึ่งงานวิจัยนี้จะทำให้มีเทคนิคในการทำนายค่า
สูญหายขึ้นมาใหม่ และยังสามารถใช้เทคนิคนี้กับวิธีการทำเหมืองข้อมูลแบบอื่น ๆ ได้อีกด้วย

PHATCHARAWAN CHINTHAISONG : THE DESIGN AND
DEVELOPMENT OF A HYBRID TECHNIQUE FOR MISSING VALUE
IMPUTATION. THESIS ADVISOR : ASSOC. PROF. NITTAYA
KERDPRASOP, Ph.D., 83 PP.

MISSING VALUE/IMPUTATION TECHNIQUE/R PROGRAMMING

At present data mining is important and can be useful in many domains such as medical, education, business, engineering, and others. But if the data are incomplete, they can affect the induced predictive model. If such model has been used to diagnose patients, it can have deadly impact. For business applications, the poor model can cause bankruptcy. This research thus proposes the design of missing-value imputation techniques to help improving the predictive performance of a classification model. An implementation of the proposed techniques has been done with the R language, which is the functional language used in statistics. The program has been designed to automatically select an appropriate imputation techniques for each specific kind of data types. The performance of our imputation methods has been tested through the assessment of the induced classification tree model. The imputation techniques proposed in this research can also support other kinds of data mining tasks.

School of Computer Engineering

Academic Year 2012

Student's Signature _____

Advisor's Signature _____