

การเปรียบเทียบความแม่นยำวิธีการประมาณค่าสูญหาย ระหว่างการวิเคราะห์จำแนกประเภท
กับการวิเคราะห์ความถดถอยเชิงพหุคูณ ร่วมกับการวิเคราะห์องค์ประกอบ

Imputation Methods Accuracy Comparison between Discriminant Analysis and Multiple linear
Regression Analysis with Principle Component Analysis

ณรงค์ โพธิ์ (Narong Phothi)* ดร.สมชาย ปรการเจริญ (Dr.Somchai Prakancharoen)**

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความแม่นยำวิธีการประมาณค่าสูญหาย ระหว่างการวิเคราะห์จำแนกประเภทกับการวิเคราะห์ความถดถอยเชิงพหุคูณ ร่วมกับการวิเคราะห์องค์ประกอบ ส่วนเกณฑ์เปรียบเทียบความแม่นยำใช้ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (MMRE) ข้อมูลที่ใช้ในการวิจัย คือ ข้อมูลคุณลักษณะที่พึงประสงค์ของอาจารย์ผู้สอนคอมพิวเตอร์ จำนวน 285 ชุด การประมวลผลระหว่างการวิเคราะห์จำแนกประเภทกับการวิเคราะห์ความถดถอยเชิงพหุคูณ โดยแบ่งออกเป็น 2 แบบ ได้แก่ 1) ข้อมูลไม่ผ่านการวิเคราะห์องค์ประกอบ 2) ข้อมูลผ่านการวิเคราะห์องค์ประกอบ พบว่าการนำข้อมูลไปผ่านการวิเคราะห์องค์ประกอบจะให้ความแม่นยำในการประมาณค่าสูญหายเพิ่มมากขึ้นของทุกวิธี ดังนี้ 1) วิธีการวิเคราะห์จำแนกประเภท (ก่อน MMRE=7.51% Accuracy=92.49% หลัง MMRE=7.06% Accuracy=92.94%) 2) วิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ (ก่อน MMRE=30.76% Accuracy=69.24% หลัง MMRE=22.21% Accuracy=77.79%)

ABSTRACT

The objective of this research is to compare Accuracy of imputation methods. The two methods are “Principle Component Analysis Study by Discriminant Analysis” and “Multiple Linear Regression Analysis”. Mean Magnitude of Relative Error (MMRE) is used to estimate the criteria for the efficiency comparison. The data used for the study are extracted from 285 sets of the desired characteristics and expertise of undergraduate computer instructors database.

The research processes include: 1) The data unprocessed discriminant analysis and multiple linear regression analysis. 2. The data in processed found. In summary, it was found that the best approach with principle component analysis for calculating MMRE range from 1) Discriminant analysis (before MMRE=7.51% Accuracy=92.49% after MMRE=7.06% Accuracy=92.94%), 2) Multiple linear regression analysis (before MMRE=30.76% Accuracy=69.24% after MMRE=22.21% Accuracy=77.79%) respectively.

คำสำคัญ : การประมาณค่าสูญหาย การวิเคราะห์จำแนกประเภท การวิเคราะห์ความถดถอยเชิงพหุคูณ การวิเคราะห์องค์ประกอบ

Key words : Data Imputation, Discriminant Analysis, Multiple linear Regression Analysis, Principle Component Analysis

* นักศึกษา หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ ม.เทคโนโลยีพระจอมเกล้าพระนครเหนือ

** ผู้ช่วยศาสตราจารย์ ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ ม.เทคโนโลยีพระจอมเกล้าพระนครเหนือ

1. บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

โดยทั่วไปงานวิจัยต้องการข้อมูลที่มีความครบถ้วน สมบูรณ์มาใช้ในการวิเคราะห์และพยากรณ์ เพื่อให้ได้ผลลัพธ์ที่แม่นยำและถูกต้องมากที่สุด แต่เนื่องจากการเก็บข้อมูล อาจมีบางส่วนที่ขาดหายไปหรือไม่สมบูรณ์เกิดขึ้นอยู่เสมอ ดังนั้น เพื่อให้สามารถนำข้อมูลทั้งหมดมาใช้งานได้ ผู้วิจัย จึงต้องนำชุดข้อมูลที่สูญหายมาแทนค่าด้วยวิธีประมาณค่า ที่ใกล้เคียงที่สุด เพื่อนำมาใช้ในการวิเคราะห์และพยากรณ์ ผลลัพธ์ หากข้อมูลเกิดการสูญหายมากๆ ผู้วิจัยอาจต้องตัด ข้อมูลชุดนั้นทิ้งไป

ในที่นี้ผู้วิจัยได้ใช้วิธีการวิเคราะห์จำแนกประเภทกับการวิเคราะห์ความถดถอยเชิงพหุคูณ พร้อมนำการวิเคราะห์องค์ประกอบมาทำการประมาณค่าสูญหาย ผู้วิจัย จะทำการตรวจสอบหาวิธีการที่ดีที่สุด โดยการวัดค่าเฉลี่ย ความคลาดเคลื่อนสัมพัทธ์ (Mean Magnitude of Relative Error: MMRE) หากมีค่าน้อยแสดงว่ามีความแม่นยำสูงมาก

1.2 ขอบเขตของการวิจัย

1.2.1 ข้อมูลที่ใช้ในการศึกษาวิจัย คือ ข้อมูล คุณลักษณะที่พึงประสงค์ของอาจารย์ผู้สอนคอมพิวเตอร์ จำนวน 285 ชุด (Case) ปี 2550 จากคณะเทคโนโลยีสารสนเทศ ม.เทคโนโลยีพระจอมเกล้าพระนครเหนือ

1.2.2 งานวิจัยนี้กำหนดให้ค่าของ “ผู้สอน มีความสามารถอธิบายพื้นฐานของระบบปฏิบัติการ (Operating System) ให้เป็นที่เข้าใจได้” (แอทริบิวต์ K411) เท่านั้นที่เป็นค่าสูญหาย เพื่อใช้ในการเปรียบเทียบ

1.2.3 ลักษณะของการสูญหายของข้อมูลจัดอยู่ในประเภทการสูญหายเชิงสุ่ม (Missing Random: MAR) (Garson G. David. 2005)

1.3 วัตถุประสงค์ของการวิจัย

1.3.1 เพื่อเปรียบเทียบความแม่นยำของวิธีการประมาณค่าทดแทนค่าสูญหายข้อมูลระหว่างการวิเคราะห์

จำแนกประเภทกับการวิเคราะห์ความถดถอยเชิงพหุคูณ ร่วมกับการวิเคราะห์องค์ประกอบวิเคราะห์

1.3.2 เพื่อประมาณความแม่นยำในการประมาณค่าทดแทนของค่าสูญหายของวิธีการวิเคราะห์จำแนกประเภทกับการวิเคราะห์ความถดถอยเชิงพหุคูณ

2. ทฤษฎีและเทคโนโลยีที่เกี่ยวข้อง

2.1 วิธีการวิเคราะห์จำแนกประเภท (Discriminant Analysis: DA) เป็นเทคนิคการวิเคราะห์ความสัมพันธ์ หรือ การหาสาเหตุที่สอดคล้องกัน (กัลยา. 2542, 2550, Geoffrey. 2005, มนต์ชัย. 2550) เพื่อเปรียบเทียบลักษณะของกลุ่มประชากร 2 กลุ่มขึ้นไป โดยค่ามากอยู่ที่กลุ่มใด จะถือว่าข้อมูลใหม่จัดอยู่ในกลุ่มนั้น ตามสูตร

$$\hat{d} = a + b_1x_1 + b_2x_2 + \dots + b_px_p \quad (1)$$

โดยที่ $\hat{d}$ = ค่าอำนาจการจำแนกประเภท

a = ค่าคงที่ (ส่วนตัดแกน Y, เมื่อ $X = 0$)

$x_1, \dots, x_n$ = ค่าตัวแปรอิสระ, ตัวแปรจำแนกกลุ่ม

$b_1, \dots, b_n$ = ค่าสัมประสิทธิ์ของสมการจำแนกกลุ่ม

2.2 วิธีการวิเคราะห์ความถดถอยเชิงพหุ (Multiple linear Regression Analysis: MRA) เป็นเทคนิคที่ใช้ในการพยากรณ์ตัวแปรตามและการประมาณค่าของข้อมูล (ชูศรี. 2541, กัลยา. 2542, 2550, มนต์ชัย. 2550) ตามสูตร

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + e \quad (2)$$

โดยที่ $\hat{y}$ = ค่าประมาณการ (ตัวแปรอิสระ)

$x_1 \dots x_n$ = ค่าของตัวชี้วัด (ตัวแปรอิสระ)

β_0 = ค่าคงที่ (ส่วนตัดแกน Y, เมื่อ $X = 0$)

$\beta_1 \dots \beta_n$ = ค่าสัมประสิทธิ์ความถดถอย

e = ค่าความคลาดเคลื่อน

2.3 วิธีการวิเคราะห์องค์ประกอบ (Principle Component Analysis: PCA) เป็นการนำความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ใช้เป็นข้อมูลกับองค์ประกอบหลัก เพื่ออธิบายความแปรปรวนของกลุ่มข้อมูลที่มี

องค์ประกอบปัจจัยหลักมีความแปรปรวนมาก-น้อยตามลำดับจนได้องค์ประกอบหลัก (หรือปัจจัย) ที่สามารถเป็นตัวแทนอธิบายการผันแปรของทุกตัวแปรได้ครบถ้วน (Miguel, and Crreira, 1997) สำหรับสมการที่ใช้ในการประมาณค่าปัจจัยที่ j

$$F_j = w_{j1}x_1 + w_{j2}x_2 + \dots + w_{jp}x_p + e \quad (3)$$

โดยที่ x_i = ตัวแปรที่ 1

w_j = สัมประสิทธิ์ของตัวแปรที่ j

F_j = ปัจจัยหรือองค์ประกอบหลักตัวที่ j

นอกจากนี้แล้วเทคนิค PCA ต้องมีการประมาณค่าน้ำหนักของปัจจัย (Factor Loading) ซึ่งจะนำมาใช้พิจารณาว่ามีตัวแปรใดบ้างที่ควรอยู่ในปัจจัยเดียวกัน โดยดูจากค่าน้ำหนักของปัจจัยที่มากที่สุดเป็นหลัก

2.4 วิธีการประมาณความแม่นยำ (Evaluation Criterion) จากการใช้วิธีต่างๆ ที่สร้างขึ้น (สมชาย, 2551) ต้องมีความแม่นยำเข้ากันได้ (Model best fit) จะถูกนำไปทดสอบกับกลุ่มข้อมูลอีกชุดหนึ่งที่ทราบค่าจริง (Actual Data) ผลจากการพยากรณ์ข้อมูลชุดใหม่ (Predicted Data) จะถูกนำมาคำนวณหาความคลาดเคลื่อนสัมพัทธ์ (Magnitude of Relative Error: MRE) โดยใช้สูตร

$$MRE_i = \frac{|actual_i - predicted_i|}{actual_i} \quad (4)$$

หากมีหลายชุดข้อมูลการคำนวณให้นำค่าความคลาดเคลื่อนสัมพัทธ์ ของแต่ละข้อมูลมารวมกันและหาค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean-MRE: MMRE) หาก MMRE มีค่ามากๆ แสดงว่ามีความคลาดเคลื่อนสูง แต่ถ้า MMRE = 0 (ศูนย์) แสดงว่าค่าของการพยากรณ์เท่ากับค่าจริงทุกค่า นั่นคือ MMRE ยิ่งมีค่าน้อยๆ แสดงว่ามีความแม่นยำสูงมากๆ ในการประมาณค่าสูญหายของข้อมูล ตามสูตร

$$MMRE = \frac{1}{n} \sum_{i=1}^n \frac{|actual_i - predicted_i|}{actual_i} \times 100 \quad (5)$$

$$Accuracy = 100 - MMRE \quad (6)$$

2.5 การสูญหายเชิงสุ่ม (Missing At Random: MAR) โดยข้อมูลที่สูญหายจะขึ้นอยู่กับค่าสังเกต (Garson G. 2005) แต่จะไม่เกิดจากการแจกแจงสุ่มภายใต้เงื่อนไขที่อ้างอิงจากตัวแปรอิสระ ซึ่งเป็นเหตุการณ์ที่ตัวแปร (Item) ที่สูญหายแต่ละตัวต่างมีความสัมพันธ์ซึ่งกันและกัน (Correlation) กับตัวแปรอื่นๆ ดังนั้น การประมาณค่าทดแทนสามารถกระทำได้โดยผ่านวิธีการต่างๆ เช่น วิธีการวิเคราะห์หารูปแบบ (Pattern Analysis) การวิเคราะห์ความถดถอยเชิงพหุคูณ (Multiple Regression) วิธีเอ็มไอ (Multiple Imputation) เป็นต้น

2.6 งานวิจัยที่เกี่ยวข้อง

Hidetomo. (2007) ได้ศึกษาวิธีการจัดกลุ่มข้อมูลจากค่าข้อมูลเริ่มต้นและค่าข้อมูลที่สูญหายด้วยวิธี Fuzzy C-Means (FCM) มีความแม่นยำมากที่สุด

ไกรรุ่ง และพยุ. (2551) ได้วิจัยการเลือกคุณลักษณะด้วยวิธีสมาชิกที่ใกล้ที่สุดสำหรับการแทนที่ค่าข้อมูลที่ขาดหายด้วยวิธีการที่ใกล้ที่สุดแบบให้น้ำหนักของข้อมูลไมโครอาร์เรย์ พบว่าใช้วิธี KNNFSW Impute มีความแม่นยำมากที่สุด

จริยา, ประสิทธิ์ และอ่ำไพ. (2551) ได้ศึกษาการเปรียบเทียบวิธีการประมาณค่าสูญหายในการวิเคราะห์ความถดถอยพหุคูณ วิธีนี้ให้ความแม่นยำสูงกว่าการแทนค่าด้วยวิธีอื่นๆ

ณรงค์ และคณะ (2551) ได้วิจัยการเปรียบเทียบผลการแยกประเภทของก้อนเนื้อจากภาพแมมโมแกรมโดยการใช้เทคนิคการทำเหมืองข้อมูลในการทำนายผลโรคมะเร็งเต้านม พบว่าการใช้วิธีโครงข่ายประสาทเทียมให้ความแม่นยำสูงสุด

ณรงค์ และคณะ (2552) ได้วิจัยการศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยวิธีการทางสถิติ พบว่าวิธีการวิเคราะห์จำแนกประเภทมีความแม่นยำมากที่สุด

ณรงค์ และคณะ (2552) ได้วิจัยการศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยวิธีการวิเคราะห์จำแนกประเภท ต้นไม้การตัดสินใจ และ ค่าเฉลี่ย พบว่าวิธีการวิเคราะห์จำแนกประเภท ให้ความแม่นยำสูงสุด

สมชาย (2551) ได้ศึกษาวิธีการประมาณการเวลาในการพัฒนาซอฟต์แวร์ประยุกต์เชิงโครงสร้างโดยวิธีแบบจำลองสมการโครงสร้าง มีความแม่นยำมากที่สุด

สรุปงานวิจัยที่ได้ศึกษานี้เป็นการใช้เทคนิคการประมาณค่าสูญหายและจำแนกประเภทข้อมูล ซึ่งมีความแตกต่างกันไปขึ้นอยู่กับประเภทและลักษณะของข้อมูลที่นำมาใช้ สำหรับงานวิจัยนี้ผู้วิจัยจึงได้ทดลองการประมาณค่าสูญหายกับฐานข้อมูลคุณลักษณะที่พึงประสงค์ของอาจารย์ผู้สอนคอมพิวเตอร์

3. วิธีการดำเนินงานวิจัย

วิธีการเปรียบเทียบความแม่นยำวิธีการประมาณค่าสูญหาย ระหว่างการวิเคราะห์จำแนกประเภทกับการวิเคราะห์ความถดถอยเชิงพหุคูณ ร่วมกับการวิเคราะห์องค์ประกอบ โดยผู้วิจัยมีขั้นตอนและวิธีการดังต่อไปนี้

3.1 แบ่งฐานข้อมูลคุณลักษณะที่พึงประสงค์ของอาจารย์ผู้สอนคอมพิวเตอร์ จำนวน 285 ชุด ออกเป็น 2 กลุ่ม คือ กลุ่มข้อมูลเรียนรู้ (Training Data) 200 ชุด และกลุ่มข้อมูลทดสอบ (Testing Data) 85 ชุด

ตารางที่ 1 แอทริบิวต์ฐานข้อมูลที่ใช้ในการวิจัย

Name	Description	Rank
M11	ผู้สอนให้คำแนะนำเมื่อนักศึกษาทำงานผิดพลาด	1-5
M12	ผู้สอนมีพฤติกรรมที่แสดงออกด้วยวาจาหรือการกระทำที่ให้การช่วยเหลือต่อนักศึกษาโดยไม่หวังผลตอบแทน	1-5
M13	ผู้สอนให้ออกาสให้นักศึกษาแก้ตัวเมื่อมีการทำไม่ถูกต้อง	1-5
M14	ผู้สอนตอบแทนบุญคุณของผู้มีอุปการะคุณ หน่วยงาน สถานที่ ตามโอกาสอันสมควร	1-5
M15	ผู้สอนมีความมั่นคงทางอารมณ์ สุขุม รอบคอบ ใช้เหตุผลตัดสินใจ	1-5
M16	ผู้สอนมีความร่วมมือและประสานงานกับนักศึกษา ผู้มีส่วนได้ส่วนเสีย บุคคลต่าง ๆ และชุมชน	1-5
M17	ผู้สอนมีความเที่ยงตรง เปิดเผย ใจกว้าง กับทุกๆ คน	1-5
M18	ผู้สอนปฏิบัติตนตามกฎระเบียบของหน่วยงานและ	1-5

Name	Description	Rank
	ประเทศชาติ	
M19	ผู้สอนให้ความเท่าเทียมกันกับนักศึกษาทุกคนเหมือนกัน	1-5
M110	ผู้สอนมีความเป็นกัลยาณมิตร และเอื้ออาทรต่อนักศึกษา เช่น รักและเมตตา ช่วยเหลือ เอาใจใส่ ให้กำลังใจโดยเสมอภาค	1-5

ตารางที่ 1 แอทริบิวต์ฐานข้อมูลที่ใช้ในการวิจัย (ต่อ)

Na me	Description	Ra nk
T211	ผู้สอนอธิบายตรงประเด็นและยกตัวอย่างประกอบให้เข้าใจชัดเจน	1-5
T212	ผู้สอนใช้อุปกรณ์/เครื่องมือ/สถานที่ที่มีความเหมาะสมกับจำนวนนักศึกษาและเนื้อหาวิชา	1-5
T213	ผู้สอนให้คำแนะนำที่ดี สามารถนำมาปฏิบัติได้	1-5
T214	ผู้สอนแนะนำนักศึกษาให้ใช้เครื่องมือต่าง ๆ หรือข้อมูลใหม่ ๆ ที่สามารถค้นคว้าเพิ่มเติมได้	1-5
T215	ผู้สอนใช้ภาษาที่ง่าย ชัดเจน ในการถามคำถามเพื่อให้ผู้ฟังเข้าใจง่าย ไม่สับสนกับคำถาม	1-5
T221	ผู้สอนมีเอกสารแผนการสอน พร้อมทั้งอธิบายจุดประสงค์ ขอบเขต เนื้อหาของวิชา แนะนำวิธีการเรียนการสอนและการประเมินผลวิชาเรียนให้นักเรียนทราบ ก่อนการเรียนการสอนจะเริ่มต้น	1-5
T222	ผู้สอนมีการกำหนดเวลาในกิจกรรมการเรียนการสอนได้อย่างเหมาะสม	1-5
T223	ผู้สอนยอมรับข้อมูลใหม่ที่มีเหตุผลถูกต้องหรือดีกว่าเดิม	1-5
T224	ผู้สอนสามารถกระตุ้นผู้เรียนให้คิดหาความเชื่อมโยง ระหว่างเนื้อหาวิชาอื่น ๆ	1-5
T225	ผู้สอนมีช่องทางในการตอบปัญหาหรือข้อสงสัยของนักศึกษาหลายช่องทาง เช่น MSN, E-mail, Webboard	1-5
T231	ผู้สอนสอบถามความรู้และความเข้าใจของนักศึกษาก่อนเรียนทุกครั้ง	1-5
T232	ผู้สอนสอบถามความรู้และความเข้าใจของนักศึกษาหลังเรียนทุกครั้ง	1-5
T233	ผู้สอนมีการสอบถามตรงตามเนื้อหา วัตถุประสงค์ที่กำหนดไว้ในแผนการสอน และควรมีการเฉลยข้อคำถาม	1-5
B311	ผู้สอนแต่งกาย สะอาด เรียบร้อย เหมาะสมกับกาลเทศะ	1-5
B312	ผู้สอนแสดงออกทางวาจาและกิริยาด้วยความเป็นมิตรและให้เกียรตินักศึกษา	1-5
B313	ผู้สอนมีท่าสัณยา และท่าสัณยาในการสอนอย่างเต็มเปี่ยม	1-5
B312	ผู้สอนสามารถรับรู้พฤติกรรมของนักศึกษาทุกคนได้	1-5
B322	ผู้สอนมีการติดตามความเคลื่อนไหวและการเปลี่ยนแปลงของความรู้ใหม่ๆ	1-5
B323	ผู้สอนสามารถควบคุมอารมณ์ตนเอง โดยใช้เหตุผลในการแก้ไขปัญหา	1-5
B324	ผู้สอนแลกเปลี่ยนและยอมรับฟังความคิดเห็นของผู้อื่นเกี่ยวกับแนวทางการทำงานหรือแก้ปัญหาด้านการสอน	1-5
B325	ผู้สอนปฏิบัติตนเป็นตัวอย่างที่ดีกับนักศึกษาและบุคลากรรอบข้าง	1-5
B326	ผู้สอนสามารถนำทักษะความรู้มาปรับใช้ให้เกิดประโยชน์สูงสุด	1-5
K411	ผู้สอนมีความสามารถอธิบายพื้นฐานของระบบปฏิบัติการ	1-5

Na me	Description	Ra nk
	(Operating System) ให้เป็นที่เข้าใจได้	
K412	ผู้สอนมีความสามารถอธิบายพื้นฐานของการประมวลผลข้อมูล (Data Processing) ให้เป็นที่เข้าใจได้	1-5
K413	ผู้สอนมีความสามารถอธิบายพื้นฐานของการเขียนโปรแกรม (Computer Programming) ให้เป็นที่เข้าใจได้	1-5

ตารางที่ 1 แอทริบิวต์ฐานข้อมูลที่ใช้ในการวิจัย (ต่อ)

Name	Description	Ra nk
K414	ผู้สอนมีความสามารถอธิบายพื้นฐานของโครงสร้างข้อมูลและการวิเคราะห์อัลกอริทึม (Data Structure and Algorithm Analysis) ให้เป็นที่เข้าใจได้	1-5
K415	ผู้สอนมีความสามารถอธิบายพื้นฐานของระบบสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ (LAN, WAN) ให้เป็นที่เข้าใจได้	1-5
K416	ผู้สอนมีความสามารถอธิบายพื้นฐานของการวิเคราะห์และออกแบบระบบ (System Analysis and Design) ให้เป็นที่เข้าใจได้	1-5
K417	ผู้สอนมีความสามารถอธิบายพื้นฐานของระบบฐานข้อมูล (Database) ให้เป็นที่เข้าใจได้	1-5
K421	ผู้สอนมีการศึกษาค้นคว้าเพื่อพัฒนาตนเองในด้านต่าง ๆ อย่างต่อเนื่อง	1-5
K422	ผู้สอนนำผลงานการวิจัยมาใช้อย่างเหมาะสมในการเรียนการสอน	1-5
S511	ผู้สอนชมเชยและให้กำลังใจเมื่อนักศึกษาให้ความสนใจ และปฏิบัติตามในการรักษามรดกของชุมชน อาทิ การจัดทำเว็บไซต์ เพื่อถ่ายทอดมรดกของชุมชน	1-5
S512	ผู้สอนมีการประยุกต์ใช้เทคโนโลยีกับศิลปะและวัฒนธรรมไทย อาทิ การสร้างตัวอักษร หรือการส่งเสริมการออกแบบด้วยวัฒนธรรมไทย	1-5
S521	ผู้สอนมีส่วนร่วมในกิจกรรมต่าง ๆ ที่ทางชุมชนจัดขึ้นมา	1-5
S522	ผู้สอนนำอุปกรณ์ด้าน ICT มาช่วยเสริมความรู้ ทักษะ ให้กับชุมชน เช่น การใช้อินเทอร์เน็ต และโปรแกรมสำเร็จรูป เป็นต้น	1-5
S523	ผู้สอนเป็นผู้ประสานความร่วมมือระหว่างบ้านและชุมชน	1-5

3.2 นำข้อมูลเรียนรู้มาหาสมการที่ดีที่สุดในแต่ละวิธี โดยแบ่งการวิจัยออกเป็น 2 แบบ ได้แก่ 1) ข้อมูลไม่ผ่านการวิเคราะห์องค์ประกอบ 2) ข้อมูลผ่านการวิเคราะห์องค์ประกอบ (Classes, 2005)

3.3 นำข้อมูลทดสอบมาแทนค่าในสมการที่ได้ในข้อ 3.2 เพื่อใช้ประมาณการหาค่าทดแทนค่าสูญหายของข้อมูล

3.4 นำค่าที่ได้จากข้อ 3.3 แทนค่าลงในชุดข้อมูลที่ย่อยแล้วนำข้อมูลทั้งหมดไปทำซ้ำตามวิธีในข้อ 3.2 และ 3.3 อีกครั้งเพื่อให้ได้ค่าทดแทนของค่าสูญหายที่ดีที่สุด

ที่สุด ผลจากการทดลองโดยใช้โปรแกรมสำเร็จรูปทางสถิติที่ระดับความเชื่อมั่น 95% (Yamane, 1967) มีดังต่อไปนี้

3.4.1 การประมาณค่าสูญหายจากสมการวิเคราะห์ความถดถอยเชิงพหุ แบบข้อมูลไม่ผ่านการวิเคราะห์องค์ประกอบได้สมการประมาณค่า ดังสมการที่ 7 และตารางที่ 2

$$\hat{y} = -0.067 + (0.170 * B325) + (0.421 * K412) + 0.474 \quad (7)$$

ตารางที่ 2 ค่าสัมประสิทธิ์ของสมการ MRA

Coefficients					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-0.067	0.332		-0.201	0.841
M11	0.040	0.064	0.041	0.624	0.533
:					
B324	-0.046	0.070	-0.049	-0.657	0.512
B325	0.170	0.074	0.168	2.288	0.023
B326	0.035	0.076	0.036	0.466	0.642
K412	0.421	0.070	0.449	5.977	0.000
K413	0.090	0.071	0.090	1.277	0.203
:					
S523	0.045	0.057	0.057	0.784	0.435

a. Dependent Variable: K411

3.4.2 การประมาณค่าสูญหายจากสมการวิเคราะห์จำแนกประเภท แบบข้อมูลไม่ผ่านการวิเคราะห์องค์ประกอบได้สมการประมาณค่า 4 สมการ ดังสมการ (บางส่วน) ที่ 8-11 และตารางที่ 3

$$\hat{d}_1 = -58.808 + (1.345 * M11) + (0.540 * M12) + \dots + (0.967 * S523) \quad (8)$$

$$\hat{d}_2 = -75.152 + (0.347 * M11) + (0.499 * M12) + \dots + (0.301 * S523) \quad (9)$$

$$\hat{d}_3 = -96.873 + (0.817*M11) + (0.252*M12) + \dots + (0.447*S523) \quad (10)$$

$$\hat{d}_4 = -120.367 + (0.965*M11) - (0.767*M12) + \dots + (0.982*S523) \quad (11)$$

ตารางที่ 3 ค่าสัมประสิทธิ์ของสมการ DA

Classification Function Coefficients				
	K411			
	น้อย	ปานกลาง	มาก	มากที่สุด
M11	1.345	0.347	0.817	0.965
M12	0.540	0.499	0.252	-0.767
M13	0.377	0.742	1.106	1.525
:				
S523	0.967	0.301	0.447	0.982
(Constant)	-58.808	-75.152	-96.873	-120.367
Fisher's linear discriminant functions				

3.4.3 การประมาณค่าสูญหายจากสมการวิเคราะห์ความถดถอยเชิงพหุ แบบข้อมูลผ่านการวิเคราะห์องค์ประกอบ ได้สมการประมาณค่า ดังสมการที่ 12 และตารางที่ 4

$$\hat{y} = 0.496 + (0.421*K412) + (0.117*K417) + 0.476 \quad (12)$$

ตารางที่ 4 ค่าสัมประสิทธิ์ของสมการ MRA

Coefficients					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0.496	0.217		2.278	0.024
K412	0.421	0.057	0.449	7.409	0.000
K413	0.120	0.064	0.119	1.886	0.061
K414	0.048	0.060	0.052	0.801	0.424
K415	0.070	0.056	0.081	1.239	0.217
K416	0.100	0.062	0.110	1.617	0.108

K417	0.117	0.049	0.140	2.385	0.018
------	-------	-------	-------	-------	-------

a. Dependent Variable: K411

3.4.4 การประมาณค่าสูญหายจากสมการวิเคราะห์จำแนกประเภท แบบข้อมูลผ่านการวิเคราะห์องค์ประกอบ ได้สมการประมาณค่า 4 สมการ ดังสมการ (บางส่วน) ที่ 13-16 และตารางที่ 5

$$\hat{d}_1 = -19.712 + (6.101*K412) + (3.306*K413) + \dots + (1.573*K417) \quad (13)$$

$$\hat{d}_2 = -36.330 + (7.567*K412) + (5.214*K413) + \dots + (2.740*K417) \quad (14)$$

$$\hat{d}_3 = -49.269 + (9.125*K412) + (5.389*K413) + \dots + (3.334*K417) \quad (15)$$

$$\hat{d}_4 = -66.983 + (11.479*K412) + (6.119*K413) + \dots + (3.666*K417) \quad (16)$$

ตารางที่ 5 ค่าสัมประสิทธิ์ของสมการ DA

Classification Function Coefficients				
	K411			
	น้อย	ปานกลาง	มาก	มากที่สุด
K412	6.101	7.567	9.125	11.479
K413	3.306	5.214	5.389	6.119
K414	2.197	1.849	2.310	2.335
K415	0.525	1.272	1.481	1.843
K416	1.149	2.011	2.553	2.827
K417	1.573	2.740	3.334	3.666
(Constant)	-19.712	-36.330	-49.269	-66.983
Fisher's linear discriminant functions				

3.5 ตรวจสอบความแม่นยำของผลลัพธ์ที่ได้จากสมการของแต่ละวิธีด้วยการหาค่าความคลาดเคลื่อนสัมพัทธ์ (MRE) ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (MMRE) และค่าความแม่นยำ (Accuracy) ได้ดังตารางที่ 6

ตารางที่ 6 ค่าเฉลี่ยความคลาดเคลื่อนของผลการวิจัย

	MRA with out PCA	DA with out PCA	MRA with PCA	DA with PCA
MMRE	30.76 %	7.51 %	22.21 %	7.06 %
Accuracy	69.24 %	92.49 %	77.79 %	92.94 %

4. สรุปผลการวิจัย

สรุปผลการเปรียบเทียบความแม่นยำวิธีการประมาณค่าสูญหายของข้อมูลคุณลักษณะที่พึงประสงค์ของอาจารย์ผู้สอนคอมพิวเตอร์ ระหว่างการวิเคราะห์จำแนกประเภทกับการวิเคราะห์ความถดถอยเชิงพหุคูณ ร่วมกับการวิเคราะห์องค์ประกอบ โดยแบ่งเป็น 2 แบบ ได้แก่ 1) ข้อมูลไม่ผ่านการวิเคราะห์องค์ประกอบ 2) ข้อมูลผ่านการวิเคราะห์องค์ประกอบ ซึ่งได้ผลการดีที่สุดเรียงลำดับจากมากไปหาน้อย ได้ดังนี้

ลำดับที่ 1 การประมาณค่าสูญหายจากสมการวิเคราะห์จำแนกประเภท แบบข้อมูลไม่ผ่านการวิเคราะห์องค์ประกอบ เป็นวิธีที่ดีที่สุด เนื่องจากมีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ 7.06% หรือ แสดงว่ามีความแม่นยำในการประมาณค่าสูญหายถูกต้อง 92.94%

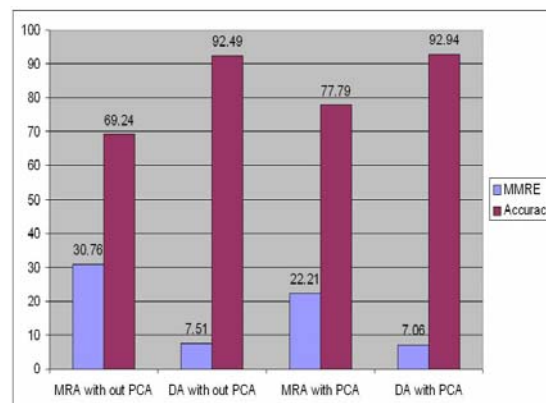
ลำดับที่ 2 การประมาณค่าสูญหายจากสมการวิเคราะห์จำแนกประเภท แบบข้อมูลไม่ผ่านการวิเคราะห์องค์ประกอบ เป็นวิธีที่ดีลำดับที่ 2 เนื่องจากมีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ 7.51% หรือ แสดงว่ามีความแม่นยำในการประมาณค่าสูญหายถูกต้อง 92.49%

ลำดับที่ 3 การประมาณค่าสูญหายจากสมการวิเคราะห์ความถดถอยเชิงพหุ แบบข้อมูลผ่านการวิเคราะห์องค์ประกอบ เป็นวิธีที่ดีลำดับที่ 3 เนื่องจากมีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ 22.21% หรือ แสดงว่ามีความแม่นยำในการประมาณค่าสูญหายถูกต้อง 77.79%

ลำดับที่ 4 การประมาณค่าสูญหายจากสมการวิเคราะห์ความถดถอยเชิงพหุ แบบข้อมูลไม่ผ่านการวิเคราะห์องค์ประกอบ เป็นวิธีที่ดีลำดับที่ 4 เนื่องจากมี

ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ 30.76% หรือ แสดงว่ามีความแม่นยำในการประมาณค่าสูญหายถูกต้อง 69.24%

สรุปผลการวิจัยเป็นกราฟเปรียบเทียบค่า MMRE และ Accuracy ของแต่ละวิธีได้ดังภาพที่ 1



ภาพที่ 1: กราฟเปรียบเทียบค่า MMRE และ Accuracy

5. วิจัยนผลและข้อเสนอแนะในการทำวิจัยต่อไป

1) การจัดกลุ่มข้อมูล (Grouping) แบบหลัก (Column) และแบบแถว (Row) ทั้งสองอย่างนี้ทำให้ได้ข้อมูลที่เป็นกลุ่มเดียวกัน การประมาณค่าทดแทนจึงได้ค่าความแม่นยำสูงสุด

2) การใช้เทคนิคจากผลการวิจัยนี้เพื่อเปรียบเทียบความแม่นยำของวิธีการประมาณค่าสูญหายสำหรับตัวเลขนั้น ในกรณีที่ข้อมูลเชิงคุณภาพ ควรใช้เทคนิคการจับกลุ่มข้อมูลแบบคร่าวๆ (Rough Set) เพื่อใช้ในการประมาณค่าสูญหายประเภทข้อมูลเชิงคุณภาพ

6. เอกสารอ้างอิง

- กัลยา วานิชย์บัญชา. (2542). การวิเคราะห์ข้อมูลด้วย SPSS for Windows. กรุงเทพฯ: ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย.
- กัลยา วานิชย์บัญชา. (2550). การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 2. กรุงเทพฯ: ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย.

- ไกรรุ่ง เสงพะพรหม และพวง มีสัจ. (2551). การเลือก
คุณลักษณะด้วยวิธีสมาชิกที่ใกล้ที่สุดสำหรับการ
แทนที่ค่าข้อมูลที่ขาดหายด้วยวิธีการที่ใกล้
ที่สุดแบบให้น้ำหนักของข้อมูล ไมโครอาร์เรย์.
การประชุมทางวิชาการระดับชาติด้าน
คอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่
4. กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอม
เกล้าพระนครเหนือ.
- จริยา แสงสุวรรณ, ประสิทธิ์ พัทธพงษ์ และอำไพ
ทองธีรภาพ. (2551). การศึกษาเปรียบเทียบ
วิธีการประมาณค่าสูญหายในการวิเคราะห์การ
ถดถอยพหุคูณ. วารสารวิทยาศาสตร์ประยุกต์. ปี
ที่ 7 (ฉบับที่ 1) มิ.ย. 51, หน้า 1-7. กรุงเทพฯ:
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนคร
เหนือ.
- ชูศรี วงศ์รัตน. (2541). เทคนิคการใช้สถิติเพื่อการวิจัย.
พิมพ์ครั้งที่ 7. กรุงเทพฯ: เทพเนรมิตการพิมพ์.
- ณรงค์ โพธิ, วาทีนี น้อยเพียร, ภัทราวุฒิ แสงศิริ, ภรณ์ยา
อำมฤรัตน์, เดช ธรรมศิริ, พวง มีสัจ และ
สมชาย
ปราการเจริญ. (2551). การเปรียบเทียบผลการแยก
ประเภทของก้อนเนื้อจากภาพแมมโมแกรมโดย
การใช้เทคนิคการทำเหมืองข้อมูลในการทำนาย
ผลของโรคมะเร็งเต้านม. การประชุมทาง
วิชาการ CIT2008 ครั้งที่ 2. กรุงเทพฯ:
มหาวิทยาลัยรังสิต.
- ณรงค์ โพธิ, สมชาย ปราการเจริญ, เดช ธรรมศิริ, ภรณ์ยา
อำมฤรัตน์ และวาทีนี น้อยเพียร. (2552).
การศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหาย
โดยวิธีการทางสถิติ. การประชุมทางวิชาการ
ระดับชาติด้านคอมพิวเตอร์และเทคโนโลยี
สารสนเทศ ครั้งที่ 5. กรุงเทพฯ: มหาวิทยาลัย
เทคโนโลยีพระจอมเกล้า พระนครเหนือ.
- มนต์ชัย เทียนทอง. (2550). สถิติและวิธีการวิจัยทาง
เทคโนโลยีสารสนเทศ. พิมพ์ครั้งที่ 1. กรุงเทพฯ:
สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- สมชาย ปราการเจริญ. (2551). การประมาณการเวลาใน
การพัฒนาซอฟต์แวร์ประยุกต์เชิงโครงข่าย
โดยวิธี แบบจำลองสมการโครงสร้าง. วารสาร
เทคโนโลยีสารสนเทศ. ปีที่ 4 (ฉบับที่ 7) ม.ค.-
มิ.ย. 2551, กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยี
พระจอมเกล้าพระนครเหนือ.
- Classes M. (2005). Factor analysis tutorial. Sam
Houston University: USA
- Garson G. David. (2005). Data Imputation for Missing
Values. North Carolina University: USA.
- Geoffrey J. McLachlan.. (2005). Discriminant Analysis
and Statistical Pattern Recognition. USA.
- Hidetomo Ichihashi. (2007). Fuzzy c-Means Classifier
with Deterministic Initialization and Missing
Value Imputation. Proceedings of the 2007
IEEE Symposium on Foundations of
Computational Intelligence (FOCI 2007)
- Miguel A. and Crreira-Perpinan. (1997). A Review of
Dimension Reduction Techniques. Technical
Report CS-96-09, Dept of Computet Science.
University of Sheffield.
- Yamane, T. (1967). Elementary sampling theory. New
Jersey: Prentice-Hall.