

การเปรียบเทียบความแม่นยำของการประมาณค่าสัญญาณข้อมูล คลื่นกำเนิดสัญญาณไฟฟ้าด้วยวิธีการสร้างแบบจำลองสมการ เชิงโครงสร้างและวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ

Accuracy Comparison of Waveform Database Generator Data Set Imputation Estimation Using Structural Equation Modeling and Multiple Regression Analysis

ณรงค์ โพธิ์* และ สมชาย ปราการเจริญ**

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อประมาณค่าสัญญาณของข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าจากฐานข้อมูลออนไลน์ของ UCI Machine Learning Repository ด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้างและวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ ซึ่งมีตัวชี้วัด 21 ตัว (จำนวน 1,000 ชุด) โดยแบ่งชุดข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่มเรียนรู้ 700 ชุด และกลุ่มทดสอบ 300 ชุด และใช้เกณฑ์เปรียบเทียบความแม่นยำ (Accuracy) จากค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (MMRE) ผลลัพธ์จากการประมาณค่าสัญญาณข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าของกลุ่มข้อมูลทดสอบแบบสุ่มข้อมูลขึ้นมาครั้งละ 1%, 5%, 10%, 15%, 30%, 50%, 70%, และ 100% ตามลำดับ ผลลัพธ์ของค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้าง คือ 3.45%, 5.86%, 8.69%, 15.89%, 8.24%, 11.07%, 11.39% และ 10.82% ตามลำดับ ส่วนวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ คือ 6.66%, 6.47%, 5.09%, 36.39%, 7.95%, 19.42%, 14.33% และ 16.02% ตามลำดับ จากผลลัพธ์ข้างต้นสรุปได้ว่า แบบจำลองสมการเชิงโครงสร้างมีความแม่นยำในการประมาณค่าสัญญาณของข้อมูลมากกว่าวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ ซึ่งมีค่าเฉลี่ยโดยรวมของค่า MMRE เท่ากับ 9.43% หรือแสดงว่ามีค่าเฉลี่ยโดยรวมของค่า

Accuracy เท่ากับ 90.57% ส่วนวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณมีค่าเฉลี่ยโดยรวมของค่า MMRE เท่ากับ 14.04% หรือแสดงว่ามีค่าเฉลี่ยโดยรวมของค่า Accuracy เท่ากับ 85.96%

คำสำคัญ: ประมาณค่าสัญญาณ แบบจำลองสมการเชิงโครงสร้าง การวิเคราะห์ความถดถอยเชิงพหุคูณ

Abstract

The objective of this research is to estimate waveform database generator data set imputation from online database of UCI Machine Learning Repository by creating Structural Equation Modeling (SEM) and Multiple Regression Analysis (MRA) with 21 indicators (1,000 data sets). Input data are divided into 2 groups including training group of 700 data sets and testing group of 300 data sets. Mean Magnitude of Relative Error (MMRE) is used for comparing the accuracy of the average relative error. The result of random testing data in the SEM from training data are 1%, 5%, 10%, 15%, 30%, 50%, 70% and 100% whereas the results MMRE of SEM are 3.45%, 5.86%, 8.69%, 15.89%, 8.24%, 11.07%, 11.39% and 10.82% re-spectively. For the method of multiple regression

* ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

** ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

analysis, the results are 6.66%, 6.47%, 5.09%, 36.39%, 7.95%, 19.42%, 14.33% and 16.02% respectively. According to the above results, it could be concluded that compared to SEM is more accurate with the average MMRE 9.43% or average accuracy 90.57%, using MRA approach with average MMRE 14.04% or average accuracy 85.96%.

Keyword: Estimation the missing data, Structural Equation Modeling, Multiple Regression Analysis.

1. บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

โดยทั่วไปงานวิจัยต้องการข้อมูลที่มีความสมบูรณ์ครบถ้วนมาใช้ในการวิเคราะห์เพื่อให้ได้ผลลัพธ์ที่แม่นยำและถูกต้องมากที่สุด แต่การเก็บข้อมูลอาจมีบางส่วนที่สูญหาย หรือไม่สมบูรณ์เกิดขึ้น ดังนั้น เพื่อให้สามารถนำข้อมูลชุดที่สมบูรณ์และชุดที่สูญหายไปมาใช้ได้ทั้งหมด จึงต้องนำชุดข้อมูลที่เกิดการสูญหายมาทดแทนค่าใหม่ด้วยวิธีประมาณค่าให้ใกล้เคียงกับค่าที่สูญหายไปให้มากที่สุดเพื่อนำกลับมาใช้ในการวิเคราะห์และพยากรณ์ข้อมูลตามที่ต้องการต่อไป หากมีบางชุดข้อมูลเกิดการสูญหายเป็นจำนวนมากๆ ผู้วิจัยจำเป็นต้องตัดข้อมูลชุดนั้นทิ้งไปเพื่อไม่ให้เกิดความเบี่ยงเบนหรือเกิดความผิดพลาดของผลลัพธ์ที่จะได้จากกระบวนการประมวลผลต่อไป

ในที่นี้ผู้วิจัยได้ศึกษาวิจัยการประมาณค่าสูญหายของข้อมูลด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้างและวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ โดยใช้ข้อมูลสำหรับการวิจัยจากฐานข้อมูลออนไลน์ UCI Machine Learning Repository คือ ข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้า (Waveform Database Generator Data Set) ชนิดตัวเลข จำนวน 1,000 ชุด เพื่อนำมาใช้ในการหาค่าความแม่นยำของการประมาณค่าสูญหาย โดยวัดความแม่นยำ (Accuracy) จากค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean Magnitude of Relative Error: MMRE) หาก MMRE มีค่าน้อยๆ แสดงว่ามีค่าความแม่นยำสูงมาก

1.2 วัตถุประสงค์ของการวิจัย

- 1.2.1 เพื่อประมาณค่าสูญหายข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้าง
- 1.2.2 เพื่อประมาณค่าสูญหายข้อมูลคลื่นกำเนิด

สัญญาณไฟฟ้าด้วยวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ

1.2.3 เพื่อเปรียบเทียบค่าความแม่นยำของการประมาณค่าสูญหายข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้างและวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ

1.3 ขอบเขตของการดำเนินงานวิจัย

1.3.1 ข้อมูลที่ใช้ในการดำเนินงานวิจัยคือ ข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้า (Waveform Database Generator Data Set) จากฐานข้อมูลออนไลน์ UCI Machine Learning Repository เนื่องจากเป็นข้อมูลชนิดตัวเลข จำนวน 1,000 ชุด โดยแบ่งเป็น 2 กลุ่ม คือ กลุ่มเรียนรู้ (Training) จำนวน 700 ชุด และกลุ่มทดสอบ (Testing) จำนวน 300 ชุด

1.3.2 ตัวชี้วัดของข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้า จากฐานข้อมูลออนไลน์ UCI Machine Learning Repository มีจำนวน 21 ตัว (V1, V2, ..., V21) ที่เป็นข้อมูลเกี่ยวกับคลื่นสัญญาณไฟฟ้า [10] ซึ่งงานวิจัยนี้กำหนดให้ตัวชี้วัดที่ 5 (V5) ของกลุ่มข้อมูลทดสอบเป็นค่าสูญหายเพื่อใช้เปรียบเทียบค่าความแม่นยำของการประมาณค่าสูญหาย เนื่องจากเป็นตัวชี้วัดที่สามารถเชื่อมโยงความสัมพันธ์กับตัวชี้วัดที่เหลือได้ดีที่สุด

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 วิธีแบบจำลองสมการเชิงโครงสร้าง

แบบจำลองสมการเชิงโครงสร้าง (Structural Equation Modeling: SEM) [1], [2] เป็นเทคนิคที่ใช้ในการวิเคราะห์หาความสัมพันธ์ของปัจจัยต่างๆ ที่ได้จากการสำรวจ (Exploratory) โดยวิธีการสกัดปัจจัย ผู้วิจัยจะทำการสร้างแบบจำลองของความสัมพันธ์จากปัจจัยต่างๆ ที่ควรจะเป็นตามหลักทฤษฎีหรือตามสมมติฐานของผู้วิจัยแล้วจึงทำการตรวจสอบว่าแบบจำลองความสัมพันธ์ (Default Model) ดังกล่าวสามารถอธิบายค่าความผันแปรร่วม (Covariance) ได้ดีเพียงใด แบบจำลองที่ได้จะต้องทำการตรวจสอบว่ามีความเหมาะสมหรือไม่จากค่าสถิติที่ใช้ในการตรวจสอบ ได้แก่ 1) ค่า Chi-square (χ^2) ควรจะเป็น Non Significance หรือมีค่ามากกว่า 0.05 2) ค่า Goodness of Fit Index: GFI ควรมากกว่า 0.90 3) ค่า Root Mean Square Error of Approximation: RMSEA ควรจะน้อยกว่าหรือเท่ากับ 0.06 และ 4) ค่า Hoelter's N ไม่ควรน้อยกว่า 75 ซึ่งค่า Hoelter's N นี้จะใช้ระบุขนาดของตัวอย่าง (Case) ว่ามีจำนวนเพียงพอต่อการวิเคราะห์แบบ

จำลองสมการโครงสร้างหรือไม่

2.2 วิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ

การวิเคราะห์ความถดถอยเชิงพหุคูณ (Multiple Regression Analysis: MRA) [3], [4], [5], [6] เป็นวิธีการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ระหว่างตัวแปรตาม (Y) หรือตัวแปรเกณฑ์ (Criterion Variable) จำนวน 1 ตัว กับตัวแปรอิสระ (X) หรือตัวแปรพยากรณ์ หรือตัวแปรทำนาย (Predictor Variable) ตั้งแต่ 2 ตัวขึ้นไป เป็นเทคนิคทางสถิติที่อาศัยความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรมาใช้ในการทำนาย โดยเมื่อทราบค่าตัวแปรหนึ่งก็สามารถทำนายอีกตัวแปรหนึ่งได้ ผลการวิเคราะห์ข้อมูลที่ได้อยู่ในรูปของสมการทำนาย สิ่งสำคัญที่ต้องการหาในการวิเคราะห์การถดถอยเชิงพหุคูณคือ สัมประสิทธิ์สหสัมพันธ์พหุคูณ สมการพยากรณ์ในรูปคะแนนดิบ หรือในรูปคะแนนมาตรฐาน หรือทั้งคู่และความคลาดเคลื่อนมาตรฐานในการพยากรณ์ ตามสูตร

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + e \quad (1)$$

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

$x_1 \dots x_n$ คือ ค่าของตัวชี้วัด (ตัวแปรอิสระ)

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

$\beta_1 \dots \beta_n$ คือ ค่าสัมประสิทธิ์ความถดถอย

e คือ ค่าความคลาดเคลื่อน

2.3 วิธีการวิเคราะห์ปัจจัย

การวิเคราะห์ปัจจัย (Factor analysis) [7] เป็นเทคนิคที่ใช้ในการสกัดปัจจัย (Factor หรือ Component) จากกลุ่มของตัวชี้วัดที่มีความสัมพันธ์ต่อกัน ปัจจัยที่ได้จะถูกใช้แทนกลุ่มของตัวชี้วัดที่มีความเป็นกลุ่มเดียวกัน เป็นเทคนิคที่ใช้ในการลดมิติหรือจำนวนตัวแปรประจักษ์ (Manifest Variable) และการพิจารณาความเหมาะสมของการสกัดปัจจัยจะตรวจสอบจากค่าสถิติ Kaiser-Meyer-Olkin: KMO (โดยค่า KMO ควรีค่ามากกว่า 0.60) ปัจจัยที่สกัดได้มีความถูกต้องเพียงใดจะพิจารณาจากค่าความ สามารถในการอธิบายค่าความผันแปร รวมของปัจจัยทั้งหมด (Total Variance Explained) กับความ ผกผันของทุกตัวแปรประจักษ์ที่ไม่มีการสกัดปัจจัย หากค่าที่ได้มีค่าเป็นร้อยละสูง (Cumulative Explained Variance) หมายถึงปัจจัยนั้นสามารถเป็นตัวแทนของตัวชี้วัดได้ดีเช่นกัน

$$F_j = w_{j1}x_1 + w_{j2}x_2 + \dots + w_{jp}x_p + e \quad (2)$$

โดยที่ F_j คือ ปัจจัยหรือองค์ประกอบหลักตัวที่ j

w_{jp} คือ สัมประสิทธิ์ของตัวแปรที่ jp

x_p คือ ตัวแปรที่ p

e คือ ค่าความคลาดเคลื่อน

2.4 วิธีการประมาณความแม่นยำ

การประมาณค่าความแม่นยำ (Accuracy Evaluation Criterion) [1],[8] จากวิธีการสร้างแบบจำลองสมการเชิงโครงสร้างและวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ ซึ่งต้องมีค่าความแม่นยำเข้ากันได้ (Model Best Fit) โดยจะถูกนำไปตรวจสอบกับชุดข้อมูลที่ทราบค่าจริง (Actual Data) กับผลลัพธ์จากการประมาณค่าของชุดข้อมูลใหม่ (Predicted Data) ด้วยการนำไปคำนวณหาความคลาดเคลื่อนสัมพัทธ์ (Magnitude of Relative Error: MRE) ตามสูตร

$$MRE_i = \frac{|Actual_i - Predicted_i|}{Actual_i} \quad (3)$$

หากชุดข้อมูลมีปริมาณมากต้องนำมาหาค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean-MRE: MMRE) หากผลลัพธ์ที่ได้พบว่า MMRE มีค่ามากแสดงว่าค่าที่ได้จากการประมาณค่าสูญหายมีความแม่นยำน้อย แต่ถ้า MMRE มีค่าน้อยหรือเข้าใกล้ศูนย์ (0) แสดงว่าค่าที่ได้มีความแม่นยำ (Accuracy) สูงหรือใกล้เคียงกับค่าจริงมาก ตามสูตร

$$MRE_i = \frac{1}{n} \sum_{i=1}^n \frac{|Actual_i - Predicted_i|}{Actual_i} \times 100 \quad (4)$$

$$Accuracy = 100 - MMRE \quad (5)$$

2.5 งานวิจัยที่เกี่ยวข้องกับการประมาณค่าสูญหาย

จากการศึกษาวิจัยของนักวิจัยที่ได้ศึกษาเกี่ยวกับวิธีการประมาณค่าสูญหายของข้อมูล อาทิเช่น สมชาย [1] ได้ใช้วิธีแบบจำลองสมการโครงสร้างในการประมาณการเวลาในการพัฒนาซอฟต์แวร์ประยุกต์เชิงโครงข่าย Hidetomo [9] ได้ศึกษาวิธีการจัดกลุ่มข้อมูลจากค่าข้อมูลเริ่มต้นและค่าข้อมูลที่สูญหายด้วยวิธี Fuzzy C-Means (FCM) ซึ่งมีความแม่นยำในการใช้งาน ส่วนงานวิจัยของณรงค์ และสมชาย [10] ได้วิจัยการศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยวิธีแบบจำลองสมการโครงสร้าง ระหว่างแบบผ่านและไม่ผ่านวิธีการวิเคราะห์จำแนกประเภท พบว่า การนำข้อมูลไปผ่านกระบวนการวิเคราะห์จำแนกประเภทก่อนนำมาแทนค่าในสมการแบบจำลองสมการโครงสร้างให้ความแม่นยำสูงสุด เพราะเป็นการกรองข้อมูลที่อยู่ในกลุ่มเดียวกันมาไว้ด้วยกัน ก่อนนำมาใช้งาน Huazhen Fang และคณะ [11] ได้ทำ



การประมาณค่าพารามิเตอร์ที่ไม่มีข้อมูลนำเข้า-ออกด้วยวิธี Recursive Minimum Component Analysis (RMCA) ซึ่งได้ผลลัพธ์ที่ดีเช่นกัน Rufus [12] ใช้สมการ SEM ในการแก้ปัญหาข้อมูลสูญหายทำให้ได้ข้อมูลใหม่ที่มีความสมบูรณ์ใกล้เคียงกับข้อมูลจริงมากยิ่งขึ้น Angela [13] ได้ใช้ MRA เข้ามาช่วยในการปรับค่าสัญญาณของข้อมูลที่ไม่สมบูรณ์ให้มีความสมบูรณ์ยิ่งขึ้นทำให้ผลลัพธ์จากการใช้ข้อมูลมีความแม่นยำขึ้น

สรุปงานวิจัยที่ข้างต้นเป็นการใช้เทคนิคการประมาณค่าสัญญาณของข้อมูลที่มีความแตกต่างกัน ซึ่งขึ้นอยู่กับประเภทและลักษณะของข้อมูลที่นำมาใช้งาน สำหรับงานวิจัยนี้ผู้วิจัยจึงได้ศึกษาเปรียบเทียบความแม่นยำของการประมาณค่าสัญญาณของข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้างและวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ จำนวน 1,000 ชุด จากฐานข้อมูลออนไลน์ของ UCI Machine Learning Repository

3. วิธีการดำเนินงานวิจัย

การเปรียบเทียบความแม่นยำของการประมาณค่าสัญญาณข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้างและวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ มีขั้นตอนการดำเนินงานดังต่อไปนี้

3.1 การแบ่งกลุ่มข้อมูลเพื่อใช้ในการดำเนินงานวิจัย

โดยการแบ่งฐานข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้า จำนวน 1,000 ชุด เป็น 2 กลุ่ม (70:30) คือ กลุ่มเรียนรู้ จำนวน 700 ชุด และกลุ่มทดสอบ จำนวน 300 ชุด

3.2 การวิเคราะห์ปัจจัยของกลุ่มข้อมูลเรียนรู้

โดยนำข้อมูลกลุ่มเรียนรู้มาวิเคราะห์ปัจจัยใหม่ด้วยวิธีการสกัดปัจจัยแบบ Principle Component Analysis: PCA เพื่อหาความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ใช้เป็นข้อมูลองค์ประกอบหลักของตัวแปร คือ การผสมเชิงเส้นตรง (Linear Combination) ของตัวแปรที่อธิบายการผันแปรของข้อมูลได้มากที่สุด และใช้วิธีการหมุนแกนแบบตั้งฉาก (Varimax) เพื่อที่จะลดจำนวนตัวแปรที่มีน้ำหนักปัจจัยมากบนแต่ละปัจจัยให้เหลือน้อยที่สุดจะทำให้ได้เฉพาะตัวแปรที่มีค่าสัมประสิทธิ์ในการรวมตัวแบบเชิงเส้นสูง ซึ่งการหมุนแกนจะเป็นวิธีการที่จะทำให้สมาชิกของแต่ละตัวแปรในปัจจัยหนึ่งๆ ชัดเจนมากยิ่งขึ้น พบว่ามีค่า KMO เท่ากับ 0.957 ดังตารางที่ 1 และได้

ปัจจัยหลักจากการสกัดปัจจัยใหม่จำนวน 3 ปัจจัยหลัก คือ F1, F2 และ F3 ดังตารางที่ 2

ตารางที่ 1 ผลลัพธ์ของค่า KMO and Bartlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.957
Bartlett's Test of Sphericity	Approx. Chi-Square	7.694E3
	df	210
	Sig.	0.000

ตารางที่ 2 ผลลัพธ์ของปัจจัยและตัวชี้วัดจาก PCA

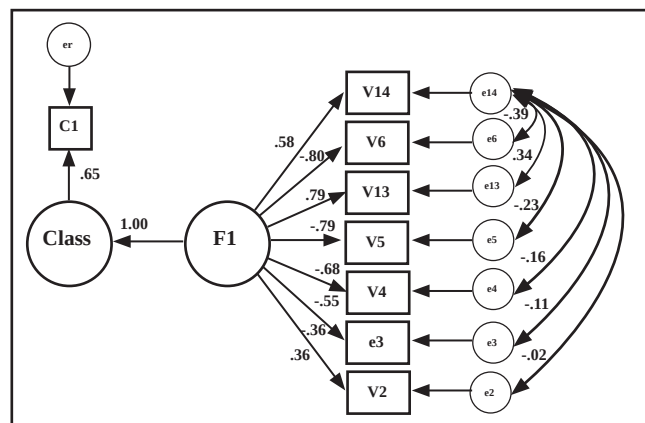
ชื่อปัจจัย	ตัวชี้วัด
F1	V15, V7, V14, V8, V16, V6, V13, V17, V5, V9, V18, V4, V19, V3, V2, V20
F2	V11, V10, V12
F3	V21, V1

3.3 การสร้างแบบจำลองสมการเชิงโครงสร้าง

โดยการนำข้อมูลปัจจัยหลักทั้ง 3 ที่ได้จากการวิเคราะห์ปัจจัยมาสร้างเป็นแบบจำลองสมการเชิงโครงสร้าง ส่วนแบบจำลองที่เหมาะสมที่สุดตรวจสอบจากค่าสถิติของความเข้ากันได้ของแบบจำลองต่อข้อมูลตัวอย่าง (Goodness of Fit) ได้แก่ ค่า χ^2 , RMSEA, GFI และ Hoelter's N ซึ่งเป็นค่าความพอเพียงของขนาดตัวอย่าง (Sample Case) ซึ่งได้ผลลัพธ์ดังตารางที่ 3 และแบบจำลองสมการเชิงโครงสร้างดังภาพที่ 1 พร้อมสมการที่เหมาะสมในการประมาณค่าสัญญาณของแบบจำลองเชิงโครงสร้างตามสมการที่ 6-8

ตารางที่ 2 ผลลัพธ์ของปัจจัยและตัวชี้วัดจาก PCA

Model	χ^2	P	GFI	RMSEA	Hoelter's N
Default	23.400	0.054	0.992	0.031	707



ภาพที่ 1 แบบจำลองสมการ SEM ประเภท Standardized



$$\text{Class} = (0.65 \times C1) + 0.392 \quad (6)$$

$$F1 = \text{Class} / 1.00 \quad (7)$$

$$V5 = (F1 - ((0.58 \times V14) - (0.80 \times V6) + (0.79 \times V13) - (0.68 \times V4) - (0.55 \times V3) - (0.36 \times V2))) / 0.79 \quad (8)$$

3.4 การวิเคราะห์ความถดถอยเชิงพหุคูณ

โดยการนำข้อมูลปัจจัยหลักเฉพาะ F1 ของกลุ่มข้อมูลเรียนรู้ ซึ่งประกอบด้วยตัวชี้วัดของคลื่นสัญญาณไฟฟ้าทั้งหมดจำนวน 16 ตัว (งานวิจัยนี้กำหนดให้ตัวชี้วัดที่ 5 (V5) ของกลุ่มข้อมูลทดสอบเป็นค่าสัญญาณ) มาใช้งานเฉพาะตัวชี้วัดที่ 2, 3, 4, 5, 6, 13 และ 14 (V2, V3, V4, V5, V6, V13, V14) ซึ่งเป็นตัวชี้วัดที่ตรงกับผลลัพธ์สุดท้ายของแบบจำลองสมการเชิงโครงสร้าง เพื่อไม่ให้เกิดความแตกต่างกันของการประมาณค่าสัญญาณของข้อมูลด้วยวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ โดยได้ผลลัพธ์ของค่าความคาดเคลื่อน ($e = 0.649$) ดังตารางที่ 4 และ 5 พร้อมสมการที่เหมาะสมในการประมาณค่าสัญญาณของวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณตามสมการที่ 9

$$V5 = (C1 - (1.213 - (0.024 \times V2) - (0.040 \times V3) - (0.084 \times V4) - (0.038 \times V6) + (0.113 \times V13) - (0.050 \times V14) + 0.649)) / 0.150 \quad (9)$$

ตารางที่ 4 ผลลัพธ์ของค่าความคาดเคลื่อนของสมการ MRA

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.616 ^a	0.379	0.373	0.649
a. Predictors: (Constant), V14, V2, V3, V4, V5, V13, V6				
b. Dependent Variable: C1				

3.5 การเปรียบเทียบผลการประมาณค่าสัญญาณของข้อมูล

โดยการนำข้อมูลกลุ่มทดสอบจำนวน 300 ชุด ซึ่งกำหนดให้ค่าตัวชี้วัดที่ 5 (V5) เป็นค่าสัญญาณสำหรับทดสอบการประมาณค่าหาตำแหน่งค่าสัญญาณของข้อมูลจากสมการแบบจำลองสมการเชิงโครงสร้างและสมการของการวิเคราะห์ความถดถอยเชิงพหุคูณ ด้วยการสุ่มข้อมูลกลุ่มทดสอบขึ้นมาครั้งละ 1%, 5%, 10%, 15%, 30%, 50%, 70% และ 100% แล้วนำข้อมูลที่ได้จากการสุ่มนี้ไปประมวลผลเพื่อหาค่าตำแหน่งค่าสัญญาณของข้อมูล ดังนี้

ตารางที่ 5 ผลลัพธ์ของการตรวจสอบความเหมาะสมของสมการ MRA

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.213	0.105		11.499	0.000
	V2	-0.024	0.025	-0.031	-0.980	0.327
	V3	-0.040	0.024	-0.058	-1.651	0.099
	V4	-0.084	0.022	-0.145	-3.749	0.000
	V5	-0.150	0.021	-0.310	-7.210	0.000
	V6	-0.038	0.021	-0.084	-1.802	0.072
	V13	0.113	0.022	0.226	5.033	0.000
	V14	-0.050	0.020	-0.109	-2.526	0.012
a. Dependent Variable: C1						

3.5.1 การประมาณค่าสัญญาณข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้างโดยการนำข้อมูลที่ได้จากการสุ่มข้อมูลกลุ่มทดสอบไปคำนวณตามสมการที่ 6-8

3.5.2 การประมาณค่าสัญญาณข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าด้วยวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณโดยการนำข้อมูลที่ได้จากการสุ่มข้อมูลกลุ่มทดสอบไปคำนวณตามสมการที่ 9

3.5.3 ผลลัพธ์ของการเปรียบเทียบผลการประมาณค่าสัญญาณของข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าจากค่า MMRE และค่า Accuracy ได้ดังตารางที่ 6

ตารางที่ 6 ผลลัพธ์ MMRE และ Accuracy ของสมการ SEM และ MRA

Model Summary ^b				
Random Data	SEM		MRA	
	MMRE	Accuracy	MMRE	Accuracy
1%	3.45	96.55	6.66	93.34
5%	5.86	94.14	6.47	93.53
10%	8.69	91.31	5.09	94.91
15%	15.89	84.11	36.39	63.61
30%	8.24	91.76	7.95	92.05
50%	11.07	88.93	19.42	80.58



Model Summaryb				
Random	SEM		MRA	
Data	MMRE	Accuracy	MMRE	Accuracy
70%	11.39	88.61	14.33	85.67
100%	10.82	89.18	16.02	83.98
Average	9.43	90.57	14.04	85.96

4. ผลการดำเนินงานวิจัย

สรุปผลการประมาณค่าสูญหายข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้า ซึ่งมีข้อมูลจำนวน 1,000 ชุด และแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มเรียนรู้ จำนวน 700 ชุด เพื่อใช้ในการสร้างสมการ และแบบจำลองในการประมาณค่าสูญหาย และกลุ่มทดสอบ จำนวน 300 ชุด เพื่อใช้ตรวจสอบความแม่นยำในการประมาณค่าสูญหายของข้อมูลโดยการสุ่มข้อมูลขึ้นมาครั้งละ 1%, 5%, 10%, 15%, 30%, 50%, 70%, และ 100% ดังนี้

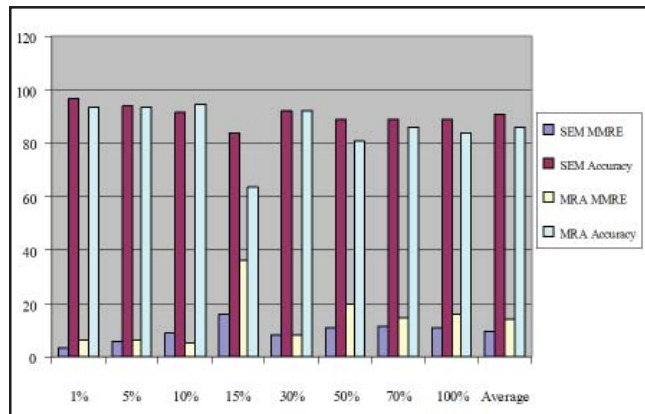
4.1 ผลลัพธ์จากการประมาณค่าสูญหายข้อมูลด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้าง ซึ่งได้ค่า MMRE คือ 3.45%, 5.86%, 8.69%, 15.89%, 8.24%, 11.07%, 11.39% และ 10.82% ตามลำดับ

4.2 ผลลัพธ์จากการประมาณค่าสูญหายข้อมูลด้วยวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ ซึ่งได้ค่า MMRE คือ 6.66%, 6.47%, 5.09%, 36.39%, 7.95%, 19.42%, 14.33% และ 16.02% ตามลำดับ

4.3 สรุปผลการเปรียบเทียบความแม่นยำของการประมาณค่าสูญหายข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าด้วยวิธีการสร้างแบบจำลองสมการเชิงโครงสร้างและวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ พบว่า วิธีการสร้างแบบจำลองสมการเชิงโครงสร้างมีความแม่นยำในการประมาณค่าสูญหายของข้อมูลมากกว่าวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ ซึ่งมีค่าเฉลี่ยโดยรวมของค่า MMRE เท่ากับ 9.43% หรือแสดงว่ามีค่าเฉลี่ยโดยรวมของค่า Accuracy เท่ากับ 90.57% ส่วนวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณมีค่าเฉลี่ยโดยรวมของค่า MMRE เท่ากับ 14.04% หรือแสดงว่ามีค่าเฉลี่ยโดยรวมของค่า Accuracy เท่ากับ 85.96% ดังภาพที่ 2

5. วิเคราะห์ผลและข้อเสนอแนะ

จากผลลัพธ์ของการดำเนินงานวิจัยของการประมาณค่าสูญหายข้อมูลคลื่นกำเนิดสัญญาณไฟฟ้าด้วยวิธีการสร้าง



ภาพที่ 2 เปรียบเทียบ MMRE และ Accuracy ของสมการ SEM และ MRA

แบบจำลองสมการเชิงโครงสร้างและวิธีการวิเคราะห์ความถดถอยเชิงพหุคูณ จำนวน 1,000 ชุด โดยแบ่งเป็น 2 กลุ่ม คือ กลุ่มเรียนรู้ 70% และกลุ่มทดสอบ 30% พบว่า การสุ่มทดสอบแบบ 1%-10% จะให้ค่า MMRE ที่ค่อนข้างดีเนื่องจากมีปริมาณข้อมูลน้อย จึงไม่เหมาะที่จะนำมาใช้งานเนื่องจากอาจเกิดความไม่น่าเชื่อถือของค่าใหม่ได้ ส่วนค่าการสุ่มแบบตั้งแต่ 15%-49% จะให้ค่า MMRE ที่แตกต่างกันมาก จึงไม่เหมาะที่จะนำมาใช้งานเช่นกัน สุดท้ายการสุ่มแบบตั้งแต่ 50% ขึ้นไปจะให้ค่า MMRE ที่มีความใกล้เคียงกันมาก ดังนั้น การสุ่มข้อมูลที่มีความเหมาะสมและน่าเชื่อถือของข้อมูลเพื่อนำมาใช้งานควรใช้ค่าตั้งแต่ 50% ขึ้นไป

ขอเสนอแนะในการศึกษาวิจัยครั้งต่อไปควรเพิ่มจำนวนของฐานข้อมูลเพื่อนำมาใช้งาน เพิ่มจำนวนตัวชี้วัดที่เกิดการสูญหาย และเพิ่มเกณฑ์ในการวัดความแม่นยำแบบอื่นๆ ที่เพื่อนำมาใช้ในการเปรียบเทียบผลการประมาณค่าสูญหายของข้อมูล เพื่อเป็นการยืนยันวิธีที่เหมาะสมที่สุดต่อไป

6. เอกสารอ้างอิง

- [1] สมชาย ปรากฏเจริญ. "การประมาณการเวลาในการพัฒนาซอฟต์แวร์ประยุกต์เชิงโครงข่าย โดยวิธีแบบจำลองสมการโครงสร้าง," วารสารเทคโนโลยีสารสนเทศ. ปีที่ 4 (ฉบับที่ 7) ม.ค.-มิ.ย. 2551, กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2551.
- [2] G. Garson David. *Data Imputation for Missing Values*. North Carolina State University, USA, 2005.
- [3] กัลยา วาณิชย์บัญชา. *การวิเคราะห์ข้อมูลหลายตัวแปร*. พิมพ์ครั้งที่ 2. กรุงเทพฯ: ศูนย์หนังสือแห่งจุฬาลงกรณ์



- มหาวิทยาลัย, 2550.
- [4] ชุติรี วงศ์รัตนะ. เทคนิคการใช้สถิติเพื่อการวิจัย. พิมพ์ครั้งที่ 7. กรุงเทพฯ: เทพเนรมิตรการพิมพ์, 2541.
- [5] มนต์ชัย เทียนทอง. สถิติและวิธีการวิจัยทางเทคโนโลยีสารสนเทศ. พิมพ์ครั้งที่ 1. กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2550.
- [6] บุญชม ศรีสะอาด. วิธีการทางสถิติสำหรับการวิจัยเล่ม 1. พิมพ์ครั้งที่ 4. กรุงเทพฯ : สุวีริยาสาส์น, 2547.
- [7] A. Miguel Crreira-Perpinan. *A Review of Dimension Reduction Techniques*. Technical, Report Dept of Computet Science. University of Sheffield, 1997.
- [8] Martin Sheppard. *Effort estimating using analogy*. Bournemouth University, USA, 1996.
- [9] Hidetomo Ichihashi., “Fuzzy c-Means Classifier with Deterministic Initialization and Missing Value Imputation,” *Proceedings of the 2007 IEEE Symposium on Foundations of Computational Intelligence*, 2007.
- [10] ณรงค์ โปธิ และ สมชาย ปรากฏเจริญ., “การเปรียบเทียบความแม่นยำวิธีการประมาณค่าสูญหาย โดยวิธีแบบจำลองสมการโครงสร้าง ระหว่างแบบผ่านและไม่ผ่านวิธีการวิเคราะห์จำแนกประเภท,” *การประชุมวิชาการวิทยาศาสตร์และเทคโนโลยี ครั้งที่ 8*. ปทุมธานี : มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต, 2553.
- [11] Huazhen Fang, Yang Shi, and Jian Wu. “Parameter Estimation with Missing Input/Output Data,” 2009 *American Control Conference*. Hyatt Regency Riverfront, St. Louis, MO, USA, June 10-12, 2009.
- [12] Rufus Lynn Carter. *Solutions for Missing Data in Structural Equation Modeling*. Research & Practice in Assessment Volume 1, Issue 1 March, 2006.
- [13] L. Angela Cool. *A Review of Methods for Dealing with Missing Data*. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, 28 January, 2000.