

การประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์ประยุกต์เชิงโครงข่าย โดยวิธีการเทียบเคียงโครงการที่ผ่านการวิเคราะห์การจำแนกกลุ่ม Web Based Application Development Cost Estimation Using Analogy Based with Discriminant Analysis

สมชาย ปราการเจริญ¹

Somchai Prakancharoen¹

บทคัดย่อ

ปัจจุบันการพัฒนาซอฟต์แวร์มีลักษณะเป็นซอฟต์แวร์ประยุกต์เชิงโครงข่าย เทคนิค สูตรที่ใช้ในการประมาณค่าใช้จ่าย การพัฒนาซอฟต์แวร์เดิมๆ เช่น COCOMO, SLIM ไม่สามารถประมาณค่าใช้จ่ายดังกล่าวได้ถูกต้อง การวิจัยนี้ได้ทดลองจัดสร้างวิธีการในการประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์ที่มีลักษณะเป็นซอฟต์แวร์ประยุกต์เชิงโครงข่าย (Transaction processing web based application) จากกลุ่มตัวอย่างภาคเอกชนและภาครัฐบาล โดยวิธีการเทียบเคียง กลุ่มตัวอย่าง รวบรวมจากโครงการพัฒนาซอฟต์แวร์ประยุกต์เชิงโครงข่ายของหน่วยพัฒนาซอฟต์แวร์จำนวน 247 โครงการ และได้ทำการสอบทานความถูกต้องกับโครงการอื่นๆ อีกจำนวน 30 โครงการ ผลการสอบทานพบว่า การประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์โดยการเทียบเคียงโครงการที่คล้ายคลึงกันในฐานะข้อมูลโครงการที่พัฒนาแล้วเสร็จ ที่ไม่มีการจำแนกกลุ่มโดยวิธีการยูคลิดีเนียนและใช้ทุกตัวชี้วัด ให้ความผิดพลาดในการประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์ประมาณ 180.931% ในขณะที่ค่าความผิดพลาดในการประมาณการค่าใช้จ่าย โดยวิธีการเทียบเคียงโครงการที่คล้ายคลึงกันในฐานะข้อมูลโครงการที่พัฒนาแล้วเสร็จ ที่มีการจำแนกกลุ่มโดยวิธีการยูคลิดีเนียน และ การเลือกกำหนดตัวชี้วัดที่มีนัยสำคัญในการประมาณการค่าใช้จ่ายให้ความผิดพลาดประมาณ 75.603% ซึ่งมีค่าความแม่นยำใน การพยากรณ์ที่ดีกว่าวิธีการการเทียบเคียงธรรมดาข้างต้นถึง 139.317%

คำสำคัญ : การประมาณราคา การพัฒนาซอฟต์แวร์เชิงโครงข่าย วิธีการเทียบเคียง การวิเคราะห์จำแนกกลุ่ม

ABSTRACT

Nowadays web based software has been mostly applied for software development. The traditional technique and formula for estimating the cost of software development such as COCOMO-II, SLIM could not estimate the exact cost accurately. The objective of this research was to develop a cost estimation method for web based software development. The data were collected from 247 web based software development projects of public and private software development units, and checked up with another 30 projects. It was found that the percentage of the errors of the cost estimation of the developed projects regarding the similar projects without classification of the projects by using Euclidean method with all indicators was 180.931%. Whereas, the percentage of the errors of the cost estimation of the developed

¹ ป.ร.ด. (เทคโนโลยีสารสนเทศ) ผู้ช่วยศาสตราจารย์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
รับต้นฉบับ 1 สิงหาคม 2551 รับลงพิมพ์ 28 ตุลาคม 2551



projects regarding the similar projects with classification of the projects by using Euclidean method with all indicators was 75.603%. The finding indicated that the prediction of the cost estimation of the projects with classification of the projects is more valid than without classification of the projects. The interval of error prediction was 139.317%.

Keyword : Cost Estimation, Web Based Application Software Development, Classification Analysis.

บทนำ

วิธีการประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์ที่เป็นที่ยอมรับใช้งานกันอย่างแพร่หลาย เช่น COCOMO, SLIM ได้พัฒนาสมการการประมาณการจากซอฟต์แวร์ที่มีลักษณะเชิงกระบวนการ (Functional oriented) ในปัจจุบันรูปแบบของซอฟต์แวร์มีความแตกต่างจากเดิมกล่าวคือ มีลักษณะเป็นซอฟต์แวร์ประยุกต์เชิงโครงข่าย (Web based application) ทำให้สูตรสมการข้างต้น ประมาณการผิดพลาดค่อนข้างสูง (Shen, 2005 ; Abstract) จึงมีผู้วิจัยได้เสนอแนวคิดต่างๆ ในการพัฒนาวิธีการในการประมาณการค่าใช้จ่ายให้แม่นยำยิ่งขึ้นได้แก่ (Jane Deleny, 2000 and Mendes, 2001).

การวิจัยนี้มุ่งพัฒนาวิธีการประมาณการค่าใช้จ่ายโดยใช้วิธีการเทียบเคียง โครงการใหม่ที่จะพัฒนาจะถูกนำมาเทียบเคียงคุณลักษณะกับโครงการเก่าๆ ที่ดำเนินการแล้วเสร็จเพื่อเลือกหาโครงการที่มีความเหมือนกันสูงสุด (Analogy) จากนั้นค่าใช้จ่ายของโครงการใหม่จะเทียบเคียงจากค่าใช้จ่ายในการพัฒนาจากโครงการที่พัฒนาแล้วเสร็จที่มีความเหมือนกันมากที่สุด ในการสืบค้นโครงการทั่วไปนิยมใช้วิธีการวัดค่าความเหมือนของทุกๆ ตัวชี้วัด-คุณลักษณะของโครงการด้วยค่า ความห่างยูคลิดีเนียน (Euclidean distance) ซึ่งเป็นวิธีการที่นิยมใช้ในกรณีข้อมูลมีค่าเป็นเชิงปริมาณและไม่ต้องคำนึงถึงระยะทาง ในงานวิจัยนี้จะทำการสืบค้นคืนโครงการโดยวิธีการเทียบเคียงโครงการที่คล้ายคลึงกันในฐานะข้อมูลโครงการที่พัฒนาแล้วเสร็จ โดยวิธีการยูคลิดีเนียนที่ไม่มีการจำแนกกลุ่มและใช้ทุกตัวชี้วัด ตามงานวิจัยของ (Shepperd, 1996) เพื่อให้การสืบค้นมีความถูกต้องยิ่งขึ้น ผู้วิจัยได้ทำการทดลองการสืบค้นโครงการที่มีการนำเทคนิค

การจำแนกกลุ่มโครงการ (Cluster analysis) การวิเคราะห์จำแนกกลุ่ม(Discriminant analysis) และการหาค่าความเหมือนโดยการเลือกระบุตัวชี้วัดเฉพาะที่มีนัยสำคัญต่อการประมาณการค่าใช้จ่าย หลังจากนั้นจะทำการประมาณการค่าความแม่นยำจากการสืบค้นของทั้งสองวิธี เพื่อทำการเปรียบเทียบสรุปหาวิธีการเทียบเคียงที่เหมาะสมต่อไป

ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

1. งานวิจัยที่เกี่ยวข้อง

Shepperd (1996) ได้กำหนดตัวชี้วัดในการสืบค้นการประมาณค่าใช้จ่ายในการพัฒนาซอฟต์แวร์โดยการเทียบเคียงโครงการซึ่งพัฒนาด้วยภาษาเดิมๆ เช่น COBOL การสืบค้นจะใช้วิธีการยูคลิดีเนียน ไรการให้น้ำหนัก และใช้ทุกตัวชี้วัดทำการเปรียบเทียบกับการประมาณค่าใช้จ่ายโดยวิธีการวิเคราะห์สมการถดถอย พบว่าการสืบค้นการประมาณค่าใช้จ่ายในการพัฒนาซอฟต์แวร์โดยการเทียบเคียงโครงการให้ค่าการประมาณการค่าใช้จ่ายได้ความแม่นยำในระดับดีกว่าวิธีการวิเคราะห์สมการถดถอย

Reifer (2002) and Briand (1999) ได้นำเสนอตัวชี้วัดเว็บ-ออบเจ็ค (Web object) เพื่อใช้ในการประมาณการขนาดของซอฟต์แวร์แทนการวัดฟังก์ชันพอยต์ (Function point) โดยจะเป็นขนาดของซอฟต์แวร์ที่มีลักษณะเชิงโครงข่ายเท่านั้น

Mendes (2001) ได้ทดลองนำตัวชี้วัดต่างๆ ที่เหมาะสมกับการประมาณการค่าใช้จ่าย กับการพัฒนาซอฟต์แวร์เชิงโครงข่ายโดยเฉพาะ มาทำการพัฒนาแบบจำลองโดยใช้เทคนิคการวิเคราะห์สมการถดถอยแบบเชิงขั้น (Stepwise regression analysis) ทำให้ทราบว่าการประมาณค่าใช้จ่ายที่แม่นยำควรใช้ตัวชี้วัดใดบ้าง

จากงานวิจัยที่นำเสนอ ผู้วิจัยจึงได้ ทำการวิจัยปรับปรุงแนวทางการวิจัยของ (Shepperd, 1996) ให้มีความทันสมัยต่อภาษาคอมพิวเตอร์(ASP and JAVA) และรูปแบบการประมวลผล(Web based application) ที่เปลี่ยนไป พร้อมนำเสนอแนวทางการสืบค้นใหม่ๆ เพื่อให้การประมวลผลที่ถูกต้องยิ่งขึ้น

ทฤษฎีที่ใช้ในการวิจัย

1. การวิเคราะห์สมการถดถอย

Regression analysis เป็นเทคนิคที่ใช้ในการพยากรณ์ตัวแปรตามจากกลุ่มของตัวแปรอิสระ(ตัวชี้วัด) หลายๆ ตัว เช่น

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \epsilon \dots\dots(1)$$

โดยที่ y = ค่าประมาณการ (ตัวแปรตาม)

α = ค่าคงที่ (Intercept แทน y)

β = ค่าสัมประสิทธิ์ถดถอย (weight)

X = ค่าของตัวชี้วัด (ตัวแปรอิสระ)

e = ค่าความคลาดเคลื่อนของการประมาณการ

2. การจำแนกกลุ่ม (Cluster analysis) (McLachlan, 2005)

เป็นเทคนิคที่นิยมใช้ในการจำแนกหรือแบ่งกลุ่มของข้อมูล(Case) โดยพิจารณาจากคุณลักษณะ-ตัวชี้วัดที่มีคุณลักษณะพิเศษที่สามารถจำแนกกลุ่มได้ การดำเนินการแยกกลุ่มต้องมีข้อมูลจำนวนหนึ่งที่ไม่จำเป็นต้องมีการแบ่งกลุ่มที่ชัดเจนแล้วมาทำการวิเคราะห์โดยวิธีการหลักๆ สองวิธีคือ "Hierarchical cluster analysis" และ "K-means cluster" วิธีการแรกเหมาะกับกลุ่มข้อมูลขนาดเล็ก หรือน้อยกว่า 200 กรณี

3. การวิเคราะห์การจำแนกกลุ่ม (Discriminant analysis) (McLachlan, 2005)

เป็นเทคนิคในการสร้างสมการที่ใช้ในการจำแนกกลุ่มของข้อมูล (Case) จากกลุ่มของข้อมูลที่มีการระบุการแยกกลุ่มไว้แล้ว สมการนี้จะใช้ในการวิเคราะห์การจำแนกกลุ่มของกรณีใหม่ว่าจะเป็นสมาชิกในกลุ่มใดต่อไป เทคนิคที่ใช้ในการพัฒนาสมการมีหลายวิธีได้แก่ Maximum likelihood

เป็นวิธีการที่จัดแยกกลุ่มโดยพิจารณาจากค่าความน่าจะเป็นสูงสุด ข้อมูลจะต้องมีลักษณะเป็น Multivariate normality วิธี Distance function เป็นวิธีที่ใช้เทคนิคการวัดระยะทางมหาลาโนบิท (Mahalanobis) และ วิธี Linear classification function ซึ่งใช้การคำนวณค่า Discriminant score ที่สูงที่สุดในการอยู่ในกลุ่มหนึ่งๆ เป็นวิธีที่นิยมมากที่สุด

4. การวัดค่าความห่างยูคลิเดียน (Euclidean distance) (Kardi, 2007)

เป็นเทคนิคที่ใช้ในการวัดความห่างของสองกรณีโดยวัดระยะทางของตัวชี้วัดหนึ่ง (x_i) กับตัวชี้วัดเดียวกันที่อาจมีค่าที่แตกต่างกัน (y_i) ของอีกกรณี กระทำจนครบทุกตัวชี้วัด ($i=1,n$) หากมีการพิจารณาหน้าหนัก (w_i) จะมีการกำหนดค่าให้ดังสมการ

$$d = \sqrt{\sum_{i=1}^n w_i (X_i - Y_i)^2} \dots\dots(2)$$

ค่า Distance ที่คำนวณได้จะมีค่าระหว่าง 0 ถึง $-\infty$ จึงต้องทำให้เป็นปกติโดยการลบค่า Distance ที่ได้โดยค่า Max distance แล้วหารด้วย Max distance $[(d - d_{\max}) / d_{\max}]$ เมื่อได้ค่า d ที่เป็นปกติ (Normalized) แล้วจะทำการแปลงเป็นค่าความเหมือน (Similarity value) โดยสมการ $S = 1 - d_{\text{normalized}}$

5. การประมาณการความแม่นยำ (Evaluation Criterion) (Shepperd, 1996)

วิธีการประมาณการที่สร้างขึ้นจะต้องมีความแม่นยำในการประมาณการ โดยนำไปทดสอบกับกลุ่มข้อมูลอีกชุดหนึ่ง ที่ทราบค่าใช้จ่ายจริง (Actual cost) ผลจากการประมาณการค่าใช้จ่ายโครงการใหม่ (Predicted cost) จะถูกนำมากำหนดเป็นค่าความคลาดเคลื่อนสัมพัทธ์ (Magnitude of relative error - MRE) ดังสมการ

$$MRE_i = \frac{[Actual\ cost_i - Predicted\ cost_i]}{Actual\ cost_i} \dots\dots(3)$$

หากมีหลายโครงการจะนำค่าความคลาดเคลื่อนสัมพัทธ์ของแต่ละโครงการมารวมกันและหาค่าเฉลี่ยดังสมการ (4)



$$MMRE = \frac{1}{n} \sum_{i=1}^n \frac{\text{Actual cost}_i - \text{Predicted cost}_i}{\text{Actual cost}_i} \times 100 \dots\dots(4)$$

หากค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean - MRE : MMRE) มีค่าสูง จะหมายถึงเปอร์เซ็นต์ของความคลาดเคลื่อนสูง หากค่า MMRE = 0 หมายถึงค่าของการพยากรณ์เท่ากับค่าจริงทุกๆ ค่า นั่นคือ MMRE ยิ่งน้อย หมายถึงมีแม่นยำสูง

วัตถุประสงค์การวิจัย

1. เพื่อพัฒนาวิธีการประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์ประยุกต์เชิงโครงข่ายโดยวิธีการเทียบเคียง
2. เพื่อประมาณการความแม่นยำในการพยากรณ์

สมมติฐานการวิจัย

การสืบค้นโครงการพัฒนาซอฟต์แวร์ที่มีความคล้ายคลึงกันจากกลุ่มโครงการที่มีการจำแนกกลุ่มและให้ความสำคัญในตัวชี้วัดที่มีนัยสำคัญต่อการจำแนกกลุ่มจะทำให้การสืบค้นโครงการได้ใกล้เคียงกับโครงการพัฒนาซอฟต์แวร์ใหม่ได้ความแม่นยำกว่าการสืบค้นโครงการที่ไม่มีการจำแนกกลุ่มและไม่มีการเลือกสรรตัวชี้วัดที่มีนัยสำคัญต่อการจำแนกโครงการ

วัสดุ อุปกรณ์และวิธีการวิจัย

ในการวิจัยเพื่อพัฒนาวิธีการประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์ประยุกต์เชิงโครงข่ายโดยวิธีการเทียบเคียงมีขั้นตอนดังนี้

1. ศึกษาวรรณกรรมอ้างอิง งานวิจัยที่เกี่ยวข้อง

โดยพิจารณา รวบรวม ค้นหา ตัวชี้วัดที่ได้มีการเผยแพร่ว่าเป็นตัวชี้วัดในการประมาณการค่าใช้จ่ายที่เหมาะสมและจำเป็นต่อการพัฒนาวิธีการประมาณการค่าใช้จ่ายซอฟต์แวร์ประยุกต์เชิงโครงข่ายจากแหล่งสำคัญๆ ได้แก่ (Shepperd, 1996 ; Reifer, 2002 ; Mendes, 2001 ; Briand, 1999 ; Boehm, 2004 and Sommerville, 2005)

2. ทำการจัดหมวดหมู่ตัวชี้วัดต่างๆ

พร้อมการวิเคราะห์ วิเคราะห์ แก้ไข ปรับปรุง โดยผู้เชี่ยวชาญจากภาคเอกชน และภาคราชการ จำนวนรวม 30 คน โดยใช้วิธีเดลฟายเทคนิคครวมสองครั้ง ผลการ

ปรับปรุงแก้ไขเพิ่มเติมทำให้ได้ตัวชี้วัดจำนวน 8 หมวด 38 ตัวชี้วัด ดังนี้

หมวดที่ 1 ตัวชี้วัดเกี่ยวข้องกับซอฟต์แวร์ที่จะพัฒนา (Product)

- 1.1 ขนาดของซอฟต์แวร์ (Size web object)
- 1.2 ความน่าเชื่อถือ (Reliability)
- 1.3 ความซับซ้อน (Complexity)
- 1.4 ประสิทธิภาพ - ความเร็ว (Performance)
- 1.5 ขนาดของฐานข้อมูล (Database size)
- 1.6 รูปแบบ โครงสร้างของซอฟต์แวร์และความเป็นอิสระ (Modularity)
- 1.7 การป้องกันและตรวจสอบข้อมูล (Data verification and validation)
- 1.8 การรักษาความปลอดภัยซอฟต์แวร์ (Web application security)
- 1.9 ความสะดวกพึงพอใจต่อการใช้งาน (Web-usability)
- 1.10 ความสะดวกในการติดตั้งการใช้งาน (Configurability)

หมวดที่ 2 ตัวชี้วัดเกี่ยวข้องกับบุคลากร (Personnel)

- 2.1 ความสามารถของผู้พัฒนาโปรแกรม (Programmer capability)
- 2.2 ประสบการณ์ของผู้พัฒนาโปรแกรม (Programmer experience)
- 2.3 ความสามารถของผู้วิเคราะห์ระบบ (Analyst capability)
- 2.4 ประสบการณ์ของผู้วิเคราะห์ระบบ (Analyst experience)
- 2.5 ความร่วมแรงร่วมใจในทีมงาน (Team cohesion)
- 2.6 ความมั่นคงในทีมงานต่อการพัฒนาโปรแกรม (การคงอยู่ - การลาออกงาน) (Team stability)
- 2.7 ประสบการณ์ของทีมต่อระบบงานที่จะพัฒนา (Team application experience)
- 2.8 ความสามารถของผู้วางโครงสร้างสถาปัตยกรรม (Web architecture capability)

2.9 ความสามารถของผู้ออกแบบ (Web designer capability)

หมวดที่ 3 กรรรมวิธีการพัฒนาซอฟต์แวร์(S/W development method)

3.1 รูปแบบกระบวนการพัฒนาซอฟต์แวร์ (Water fall, evolutionary)

3.2 ระดับการพัฒนาความสามารถของกระบวนการพัฒนาซอฟต์แวร์ (Capability maturity model)

3.3 การวิเคราะห์ความต้องการของระบบ (Requirement engineering)

หมวดที่ 4 ฐานระบบงาน (Platform)

4.1 ระดับความยากของฐานระบบงาน (Platform difficulty เช่น Client/server, three tier, distributed object)

4.2 ความแตกต่างของระบบซอฟต์แวร์ที่เป็นปัญหาในการแลกเปลี่ยนข้อมูลระหว่างระบบที่พัฒนากับระบบงานอื่นๆ (Interoperability /heterogeneity)

4.3 ความยากของภาษาที่ใช้ในการพัฒนา (Language difficulty)

4.4 การปรับเปลี่ยนฐานระบบงาน (Platform volatility)

หมวดที่ 5 การบริหารจัดการโครงการ (Management/project)

5.1 ประสบการณ์ของผู้บริหารโครงการ (Project manager experience)

5.2 จำนวนโครงการที่ประสบความสำเร็จ (% of successproject - complete)

5.3 สิ่งอำนวยความสะดวกในการทำงาน (Infrastructure /resource availability and dedicated to project)

หมวดที่ 6 โปรแกรมสนับสนุนการพัฒนาซอฟต์แวร์ (CASE tool)

6.1 ประสบการณ์ในการใช้โปรแกรมสนับสนุน (CASE tool experience)

6.2 ฟังก์ชันของ CASE tools (CASE tool functions)

6.3 ประเภทของ CASE tool (CASE tool category)

6.4 โปรแกรมสนับสนุนการออกแบบเว็บ (Web design /develop tool)

หมวดที่ 7 ข้อจำกัดโครงการ (Constraint)

7.1 ความเร่งรัดของกำหนดการ/ระยะเวลาโครงการในการพัฒนาซอฟต์แวร์ (Schedule)

7.2 ความสะดวกในการเปลี่ยนแปลงความต้องการได้ (Requirement volatility)

7.3 การจัดหาทรัพยากรทางด้านวิชาการและการสนับสนุนด้านเทคนิค (Resource/technical support)

หมวดที่ 8 ตัวชี้วัดอื่นๆ

8.1 ความสามารถในการพัฒนาจากโครงสร้างเดิม (Reusable project reusability)

8.2 แผนงานการพัฒนาซอฟต์แวร์ (Web blue print)

3. จัดสร้างแบบสอบถาม

เพื่อใช้ในการรวบรวมข้อมูลตามตัวชี้วัดข้างต้น ทำการทดสอบความถูกต้องความเข้าใจความต้องการตรงตามวัตถุประสงค์ของแบบสอบถามต่อวัตถุประสงค์ของงานวิจัย โดยการทำการสอบถามจากกลุ่มตัวอย่างทดลองตอบแบบสอบถามจำนวน 10 คนเพื่อให้ข้อเสนอแนะในการปรับปรุง แก้ไข แบบสอบถาม แบบสอบถามที่ได้รับการปรับปรุงแล้วได้ถูกนำไปใช้ในการรวบรวมข้อมูลจากกลุ่มตัวอย่างผู้พัฒนาซอฟต์แวร์ประยุกต์เชิงโครงข่าย ซึ่งรวบรวมจากแหล่งซอฟต์แวร์เฮาส์ และผู้พัฒนาซอฟต์แวร์อิสระรวมจำนวนเท่ากับ 285 โครงการ

4. ตรวจสอบความถูกต้องของแบบสอบถาม

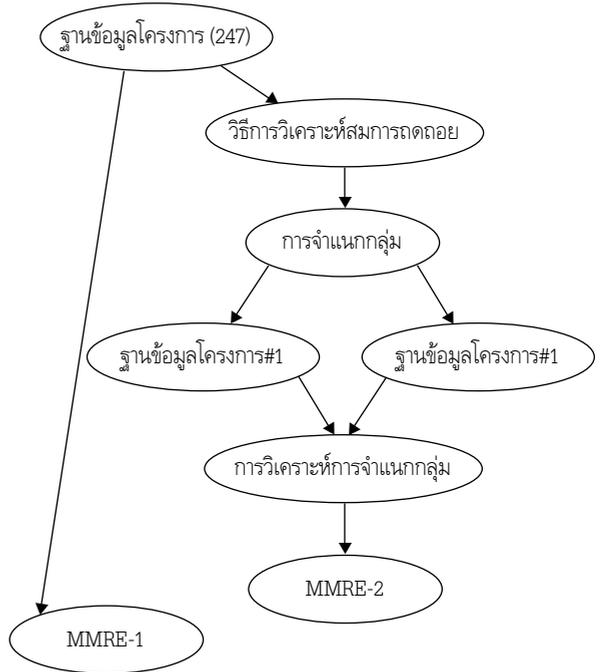
หลังจากรวบรวมแบบสอบถามแล้วได้มีการตรวจสอบความถูกต้องของแบบสอบถาม และคัดออกแบบสอบถามที่มีลักษณะการตอบที่ผิดปกติ หรือมีความแตกต่างจากกลุ่มตัวอย่างทุกๆ ไป (Outlier) ทำให้คงเหลือแบบสอบถามที่เหมาะสมเข้าสู่การจัดทำแบบจำลองจำนวน 277 โดยที่โครงการทั้งหมดเป็นโครงการขนาดเล็กถึงกลาง (โดยพิจารณาจากค่าใช้จ่ายระหว่าง 10,000 - 1,000,000 บาท)



รวม 277 โครงการนี้จะถูกนำไปพัฒนา วิธีการประมาณการ ค่าใช้จ่าย และนอกจากนี้ผู้วิจัยได้คัดเลือก 30 โครงการ สำหรับการสอบทานค่าความแม่นยำ (MMRE)

5. ขั้นตอนการดำเนินการวิจัย

โครงการจำนวน 247 โครงการจะถูกนำไปวิเคราะห์หา ตัวชี้วัดที่มีนัยสำคัญต่อการประมาณการค่าใช้จ่ายในการ พัฒนาซอฟต์แวร์โดยวิธีการวิเคราะห์สมการถดถอย ตัวชี้วัด นี้จะถูกนำไปใช้ในการจำแนกกลุ่มของฐานข้อมูล โครงการ(Cluster analysis)โดยจะจำแนกออกเป็นสองกลุ่ม ในแต่ละกลุ่มข้อมูลจะถูกนำมาสร้างสมการในการวิเคราะห์ จำแนกกลุ่ม(Discriminant analysis) เมื่อต้องการประมาณ การค่าใช้จ่ายโครงการอื่นๆ เพื่อสอบทานความถูกต้องจะใช้ 30 โครงการที่คัดเลือกไว้มาทำการสืบค้นจากฐานข้อมูล โครงการ (247 โครงการ) การดำเนินการจะเริ่มด้วยการนำ ข้อมูลของโครงการใหม่ป้อนเข้าสู่สมการการจำแนกกลุ่มเพื่อ ทากกลุ่ม (#1 หรือ #2) แล้วจึงไปทำการสืบค้นในฐานข้อมูลโดย พิจารณาโครงการที่อยู่ในกลุ่มเดียวกัน การสืบค้นจะใช้วิธีการ วัดค่าความห่าง(Euclidean distance) จากนั้นทำการปรับให้ เป็นค่าปกติ และแปลงเป็นค่าร้อยละของความเหมือนต่อไป โครงการที่มีค่าร้อยละของความเหมือนทุกโครงการตั้งแต่ 80% ขึ้นไปจะถูกนำค่าใช้จ่ายของโครงการเหล่านั้น มาหาค่า เฉลี่ย เพื่อใช้เป็นค่าประมาณการค่าใช้จ่ายของโครงการที่นำ ไปเทียบเคียงต่อไป ค่าความแตกต่างของค่าใช้จ่ายที่ได้จาก การเทียบเคียง กับค่าใช้จ่ายจริงของโครงการที่นำมาสอบทาน จะถูกคำนวณค่าไว้เป็นค่าความคาดเคลื่อนของโครงการนั้นๆ (Magnitude of relative error: MRE) ดำเนินการให้ครบทั้ง 30 โครงการจากนั้นทำการคำนวณหาค่าเฉลี่ยค่าความ คาดเคลื่อน (Mean magnitude of relative error: "MMRE - 2") ต่อไป ส่วนการประมาณการค่าใช้จ่ายในการพัฒนา ซอฟต์แวร์ โดยการเทียบเคียงโครงการที่คล้ายคลึงกันในฐานะ ข้อมูลโครงการที่พัฒนาแล้วเสร็จทั้งหมด (247 โครงการ) โดย วิธีการยูคลิดีเนียนที่ไม่มีการจำแนกกลุ่มและใช้ทุกตัวชี้วัด ให้ ความผิดพลาดในการประมาณการค่าใช้จ่ายในการพัฒนา ซอฟต์แวร์ประมาณเป็น "MMRE-1" ผังการดำเนินการวิจัย มีดังนี้



ภาพประกอบที่ 1 ขั้นตอนการดำเนินการวิจัย

ผลการวิจัย

ผลการวิจัยมีดังนี้
ผู้วิจัยได้ทดลองประมวลผลโดยโปรแกรมสำเร็จรูป ทางสถิติโดยใช้เทคนิคต่างๆ ได้ผลดังนี้

สถิติพื้นฐานของกลุ่มตัวอย่าง

จากกลุ่มตัวอย่าง 277 โครงการ ค่าสถิติพื้นฐานของ ตัวแปรค่าใช้จ่าย ในการพัฒนาซอฟต์แวร์ มีดังนี้

ตารางที่ 1 ค่าสถิติพื้นฐานของตัวแปรค่าใช้จ่าย

N Valid	277
N Missing	0
Mean	216155.38
Median	100000.00
Mode	30000
Std. Deviation	237664.9
Minimum	10000
Maximum	1000000

การวิเคราะห์สมการถดถอย (Regression analysis-standardized)

สมการพยากรณ์ค่าใช้จ่าย ดังนี้

$$\text{Cost} = 0.885 * \text{ZV1_1_log} - 0.086 * \text{ZV1_4} + 0.089 * \text{Zv3_2} - 0.069 * \text{zv4_2} - 0.120 * \text{ZV5_1} \dots\dots\dots(5)$$

$$[R^2 = 0.891 \text{ F} = 42.782 \text{ Sig} = 0.000 \text{ Tolerance mean} = 0.435 \text{ Durbin_watson} = 1.44]$$

นั่นคือ ตัวชี้วัดหมวดที่ 1 zv1_1_log (ตัวที่ 1 ที่ทำการแปลงค่าโดยสมการ Log เพื่อให้มีการแจกแจงแบบปกติพร้อมการทำให้เป็นค่ามาตรฐาน(Standardized-Z) แล้ว) ตัวชี้วัด zv1_4 ตัวชี้วัด zv3_2 ตัวชี้วัด zv4_2 และตัวชี้วัด zv5_1 ทั้งห้าตัวชี้วัดเป็นตัวชี้วัดที่มีนัยสำคัญต่อการพยากรณ์ค่าใช้จ่าย ในงานวิจัยนี้จะใช้ตัวชี้วัดนี้ในการจำแนกกลุ่มข้อมูลต่อไป

การจำแนกกลุ่ม(Cluster analysis)

จากตัวชี้วัดทั้งห้าจากสมการที่ 5 ได้ถูกนำมาเป็นตัวแปรในการจัดแบ่งกลุ่มของข้อมูล 247 โครงการ โดยวิธี Hierarchical analysis จำนวน 2 กลุ่มให้มีค่าความห่างไม่เกิน 20 หน่วย พบว่าในกลุ่มที่ 1 มีจำนวนรวม 223 โครงการ และ 15 โครงการในกลุ่มที่ 2 และ 9 โครงการที่ไม่สามารถระบุกลุ่มได้เมื่อทำการตรวจสอบกับประเภทของโครงการที่เก็บรวบรวมมา ประเภทที่ 1 เป็นโครงการในภาคเอกชนโดยส่วนมากและเช่นกันภาคราชการในประเภท กลุ่มที่ 2

การวิเคราะห์การจำแนกกลุ่ม (Discriminant analysis)

การระบุกลุ่มจากสมการที่ 5 ได้ถูกนำมาวิเคราะห์หาสมการที่ใช้ในการพยากรณ์กลุ่มต่อไป

จากการตรวจสอบความเป็นปรกติของการแจกแจงปกติของหลายตัวแปร (Multivariate normality) โดยการตรวจสอบค่า Box's F ให้ค่าที่มีนัยสำคัญโดยมีค่า F=1.069 นั่นคือ ข้อมูลมีการแจกแจงแบบการแจกแจงปกติของหลายตัวแปรจริง

ตารางที่ 2 ผลการทดสอบการแจกแจงปกติของหลายตัวแปร

Box's M	18.525
F	1.069
df ₁	15
df ₂	2244.348
Sig.	.380

ในการวิเคราะห์การจำแนกกลุ่มโดยการกำหนดใช้วิธีการคำนวณความน่าจะเป็นจากขนาดของกลุ่มตัวอย่างเนื่องจากกลุ่มที่ 1 มีจำนวนมากกว่ากลุ่มที่ 2 ค่อนข้างสูง การจำแนกกลุ่มใช้เทคนิค Within-group เนื่องจากมีการกระจาย ($\sigma_1 = \sigma_2$) ที่ใกล้เคียงกัน การคัดเลือกตัวแปรอิสระเข้าสู่สมการวิเคราะห์การจำแนกกลุ่มใช้วิธีการแบบ Stepwise ผลการสร้างสมการ ผลการวิเคราะห์ได้สมการที่สามารถจำแนกกลุ่มได้ดี โดยพิจารณาจากค่า Wilks' Lambda มีค่านัยสำคัญทางสถิติที่นัยสำคัญทางสถิติ .01

ตารางที่ 3 ค่า Wilks' Lambda ของการวิเคราะห์จำแนกกลุ่ม

Test of Function (s)	Wilks' Lambda	Chi-square	df	Sig
	.651	100.312	5	.000

สมการวิเคราะห์การจำแนกกลุ่ม มีค่าสัมประสิทธิ์การจำแนก (ไม่เป็นค่ามาตรฐาน) ดังสมการที่ 6 สำหรับกลุ่มที่ 1 และสมการที่ 7 สำหรับกลุ่มที่ 2

$$D1 = 0.711 + 0.072 * \text{zv1_1_log} + 0.042 * \text{zv5_1} + 0.103 * \text{zv4_2} + 0.102 * \text{zv3_2} + 0.033 * \text{zv1_4} \dots\dots\dots(6)$$

$$D2 = -4.649 - 1.067 * \text{zv1_1_log} - 0.62 * \text{zv5_1} - 1.537 * \text{zv4_2} - 1.518 * \text{zv3_2} - 0.484 * \text{zv1_4} \dots\dots\dots(7)$$

หากพิจารณาค่าสัมประสิทธิ์การจำแนกที่เป็นค่ามาตรฐานจะพบว่าปัจจัย Zv4_2 มีความสำคัญสูงสุดในการจำแนกกลุ่มที่ 0.492 ดังตารางที่ 4



ตารางที่ 4 ค่าสัมประสิทธิ์การจำแนกที่เป็นค่ามาตรฐาน

	Function
Zv1_1	1.000
Zv5_1	.365
Zv4_2	.208
Zv3_2	.492
Zv1_4	.477
	.164

ในการพัฒนาสมการวิเคราะห์การจำแนกกลุ่มและสอบทานสมการ ให้ค่าความแม่นยำที่ระดับเดียวกันคือ 95.50% โดยมีกรณีที่ไม่สามารถจำแนกกลุ่มได้จำนวน 9 กรณี

ตารางที่ 5 ผลการสร้างสมการวิเคราะห์การจำแนกและการสอบทานสมการ

		Group	Predicted Group		Total Membership
			Membership		
			1	2	
Original	Count	1	213	10	223
		2	0	15	15
	%	1	95.5	4.5	100.0
		2	0	100.0	100.0
Cross-validated	Count	1	213	10	223
		2	0	15	15
	%	1	95.5	4.5	100.0
		2	0	100.0	100.0

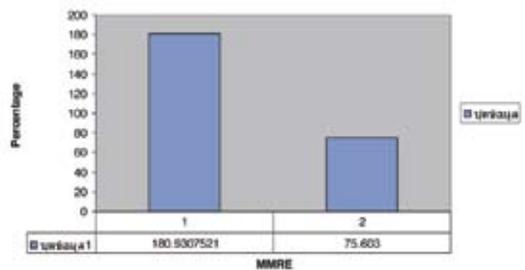
A : cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

B : 95.8% of original grouped cases correctly classified.

C : 95.8% of cross-validated grouped cases correctly classified.

ผลการดำเนินการสืบค้นโครงการ และหาค่าความแม่นยำในการพยากรณ์ (MMRE)

ผลการสืบค้น(วิธีที่1)โดยวิธีการยูคลิเดียนจากฐานข้อมูลโครงการที่ไม่มีการจำแนกกลุ่มและใช้ทุกตัวชี้วัดปราศจากการให้น้ำหนักให้ความผิดพลาดในการประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์ประมาณเป็น “MMRE-1” เท่ากับ “180.930%” และค่าความผิดพลาดในการประมาณค่าใช้จ่าย “MMRE-2” โดยวิธีการเทียบเคียงโครงการที่คล้ายคลึงกัน (วิธีที่2) ในฐานข้อมูลจากโครงการที่พัฒนาแล้วเสร็จที่มีการจำแนกกลุ่ม โดยวิธีการยูคลิเดียนที่มีการเลือกกำหนดตัวชี้วัดที่มีนัยสำคัญในการประมาณการค่าใช้จ่ายและมีการให้น้ำหนัก ให้ค่าประมาณการเท่ากับ “75.603%” ดังการดำเนินการในภาพ ที่ 1 มีผลดังนี้



ภาพที่ 1 ค่าร้อยละของความผิดพลาดในการประมาณการค่าใช้จ่ายจากการสืบค้นโครงการวิธีที่ 1 และ วิธีที่ 2

วิจารณ์ผลการวิจัย

ผลการวิจัยมีประเด็นสำคัญนำมาอภิปรายผล ดังนี้

การประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์ที่จะทำการพัฒนาโดยการเทียบเคียงโครงการที่จะพัฒนามากับโครงการพัฒนาซอฟต์แวร์จำนวนมากที่พัฒนาแล้วเสร็จจัดเก็บอยู่ในฐานข้อมูลโดยการเทียบเคียงโครงการใหม่กับโครงการเก่าโดยการให้ความสำคัญต่อตัวชี้วัดที่มีนัยสำคัญต่อการจำแนกกลุ่มโครงการเก่าๆ การค้นหาจะทำการตรวจ

สอบก่อนว่าโครงการใหม่อยู่ในกลุ่มโครงการเก่ากลุ่มใดๆ การสืบค้นจะสืบค้นจากเฉพาะโครงการเก่าๆ ที่อยู่ในกลุ่มที่โครงการใหม่ที่จะพัฒนาถูกระบุโดยสมการวิเคราะห์การจำแนกกลุ่ม ผลการวิจัยพบว่าค่าความผิดพลาดในการประมาณการค่าใช้จ่ายในการพัฒนาซอฟต์แวร์โดยการเทียบเคียงโครงการ ให้ค่าความผิดพลาดต่ำกว่าการเทียบเคียงโครงการโดยไม่มีกรจำแนกกลุ่มและทำการสืบค้นโดยการพิจารณาทุกตัวชี้วัด โดยวิธีการที่นำเสนอให้ค่าความผิดพลาดในการประมาณการค่าใช้จ่ายโดยการเทียบเคียงโครงการเพียง 75.603% ในขณะที่วิธีการสืบค้นโดยไม่มีกรจำแนกกลุ่มและไม่เลือกสรรตัวชี้วัดให้ความผิดพลาดในการประมาณการถึง 180.930% ผลการวิจัยตรงตามสมมติฐานของการวิจัยและช่วยเสริมสร้างความแม่นยำในการสืบค้นโครงการของงานวิจัย (Shepperd,1996) ที่ไม่มีกรดำเนินการวิเคราะห์การจำแนกกลุ่มและไม่ให้ความสำคัญต่อตัวชี้วัดเฉพาะที่มีนัยสำคัญแต่อย่างใด

สรุปผลการวิจัย

ผลการวิจัยสรุปได้ดังนี้

จากผลการวิจัย ผู้วิจัยได้รวบรวมตัวชี้วัด จากงานวิจัยต่างๆ และผ่านการพิจารณาจากผู้เชี่ยวชาญ ทำให้ได้ตัวชี้วัดจำนวนทั้งสิ้น 38 ตัว ใน 8 หมวด เพื่อใช้ในการจำแนกกลุ่มวิเคราะห์จำแนกกลุ่มแล้วจึงนำไปใช้ในการพิจารณาโครงการใหม่ว่าจะอยู่กลุ่มใด ให้ไปสืบค้นเทียบเคียงที่กลุ่มนั้นๆ โดยการให้ความสนใจเฉพาะตัวชี้วัดที่มีนัยสำคัญต่อการประมาณการค่าใช้จ่ายและให้นำหนักแก่ตัวชี้วัดแตกต่างกันไป ผลการดำเนินการพบว่าวิธีการที่นำเสนอให้ค่าความแม่นยำในการประมาณการที่ใกล้เคียงกว่า วิธีการที่ใช้การสืบค้นโครงการธรรมดาถึง "139.317%" ค่าประมาณการค่าใช้จ่ายที่ได้ถึงแม้จะยังคงมีความผิดพลาดถึง "75.603%" อาจมาจากโครงการเดิมที่พัฒนาแล้วเสร็จกับโครงการใหม่ที่จะสืบค้นมีความแตกต่างกัน จึงควรมีการรวบรวมโครงการพัฒนาซอฟต์แวร์ที่แล้วเสร็จเพิ่มเติมเข้าไปในฐานข้อมูลให้มากยิ่งขึ้น และทำการศึกษาตัวชี้วัดใหม่ๆ ต่อไป

เอกสารอ้างอิง

- Barry, B. (2004). **COCOMO Tutorial** University of southern califonia.
- Donald, R. (2001). **Web Object Counting Conventions**. Reifer consultants, Inc., CA, USA.
- Donald, R. (2002). **System Level Metrics for Software Evelopment Estimation**. Univ. of southern califonia, USA.
- Darlington, R. B. (1990). **Regression and Linear Models** New York, McGraw-Hill.
- Emillia, M. (2001). **Web Metrics Estimating Design and Authoring Effort**. Univ. of Auckland, Newzealand.
- Geoffrey J. L. (2005). **Discriminant Analysis and Statistical Pattern Recognition**. Wiley Series, USA,
- Ian, S. (2005). **Software Engineering**. St Andrews Univ.UK.
- Lionel, C. B. (1999). **An Assessment and Comparison of Common Software cost Estimation Modeling Techniques**. Kaiserslaudern, Germany.
- Martin, (1996). **Effort Estimating Using Analogy**. Bournemouth Univ. USA.
- Sarah, Jane D. (2000). **The Application of CASE Base Reasoning to Early Software Project Cost Estimation**. Trinity college, UK.
- Teknomo, K. (2007) **Similarity Measurement** available at <http://people.revoked.com>, on 20 October.
- Zhihao, C. (2005). **Finding the Right Data for Software Cost Modeling**. University of southern califonia, USA.

