การประมาณค่าความพยายามของโครงการเป็นระยะโดยวิธีโครงข่ายประสาท

นายพิชัย จอดพิมาย

1313607376

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ภาควิชาคณิตศาสตร์และวิทยาการ
คอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

PHASE-WISE PROJECT EFFORT ESTIMATION USING NEURAL NETWORKS

Mr. Pichai Jodpimai

13136073 76

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Science and
Information Technology
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2013

Copyright of Chulalongkorn University

5173837523

Thesis Title	PHASE-WISE PROJECT EFFORT ESTIMATION USING NEURAL NETWORKS
By	Mr. Pichai Jodpimai
Field of Study	Computer Science and Information Technology
Thesis Advisor	Associate Professor Peraphon Sophatsathit, Ph.D.
Thesis Co-Advisor	Professor Chidchanok Lursinsap, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.....Dean of the Faculty of Science
(Professor Supot Hannongbua, Dr.rer.nat.)

THESIS COMMITTEE

.....Chairman
(Asis Unyapoth, Ph.D.)

.....Thesis Advisor
(Associate Professor Peraphon Sophatsathit, Ph.D.)

.....Thesis Co-Advisor
(Professor Chidchanok Lursinsap, Ph.D.)

.....Examiner
(Associate Professor Pornsiri Muenchaisri, Ph.D.)

.....Examiner
(Associate Professor Wiwat Vatanawood, Ph.D.)

.....External Examiner
(Suradet Jitprapaikulsarn, Ph.D.)

พิชัย จอดพิมาย : การประมาณค่าความพยายามของโครงการเป็นระยะโดยวิธีโครงข่ายประสาท. (PHASE-WISE PROJECT EFFORT ESTIMATION USING NEURAL NETWORKS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. พีระพนธ์ โสพัศสถิตย์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ศ. ดร. ชิดชนก เหลือสินทรัพย์, 74 หน้า.

การประมาณค่าความพยายามของโครงการผิด เป็นสาเหตุสำคัญของระยะเวลาของโครงการผิดพลาด กำไรต่ำ และไม่ได้รับการประมูล หลายๆแนวทางการประมาณการที่มีอยู่ดำเนินการประมาณค่าความพยายามโดยรวมของโครงการ การศึกษานี้เสนอแนวทางการประมาณค่าความพยายามของโครงการในแต่ละขั้นตอนของการพัฒนาโครงการแล้วรวมเป็นค่าความพยายามทั้งหมดของโครงการ โดยใช้การวิเคราะห์แบบละเอียดเพื่อเพิ่มการมองเห็นโครงการ แนวทางที่นำเสนอถูกดำเนินการในสี่ขั้นตอน ขั้นตอนแรกให้ความสนใจในการจัดเก็บรวบรวมข้อมูลโครงการจากองค์กรพัฒนาซอฟต์แวร์ ขั้นตอนที่สองเปลี่ยนข้อมูลเชิงคุณภาพไปเป็นข้อมูลเชิงปริมาณ ขั้นตอนที่สามค้นหาคุณลักษณะของโครงการที่เกี่ยวข้องสำหรับการประมาณค่าความพยายาม ขั้นตอนสุดท้ายดำเนินการสร้างแบบจำลองการประมาณค่าความพยายามของโครงการโดยวิธีโครงข่ายประสาท ผลลัพธ์แสดงให้เห็นว่าแนวทางที่นำเสนอใช้คุณลักษณะของโครงการที่เกี่ยวข้องในจำนวนเล็กน้อย แต่ให้การประมาณที่ถูกต้องสูงกว่าแนวทางที่มีอยู่แล้ว โดยการใช้มาตรวัดความถูกต้อง และการทดสอบทางสถิติในการประเมินประสิทธิภาพและการทวนสอบ การสนับสนุนของการประมาณการเป็นระยะที่นำเสนอสามารถเป็นประโยชน์ต่อผู้จัดการโครงการในการติดตาม และควบคุมโครงการอย่างละเอียด

1313607376

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์

สาขาวิชา วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ

ปีการศึกษา 2556

ลายมือชื่อนิสิต .....

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม .....

5173837523 : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
KEYWORDS: PHASE-WISE ESTIMATION / A NEURAL NETWORK / PROJECT SIZE / SOFTWARE PROJECT EFFORT

PICHAI JODPIMAI: PHASE-WISE PROJECT EFFORT ESTIMATION USING NEURAL NETWORKS. ADVISOR: ASSOC. PROF. PERAPHON SOPHATSATHIT, Ph.D., CO-ADVISOR: PROF. CHIDCHANOK LURSINSAP, Ph.D., 74 pp.

Inaccurate project effort estimation is a significant cause of missed project deadline, low profit, and lost bids. Many existing estimation approaches tend to predict an overall effort of a project. This study proposes a phase-wise effort estimation approach to predict an effort of each phase of a project using fine-grained analysis to increase project visibility. The proposed approach is carried out in four steps. The first step concerns collection of project data from software development organizations. The second step transforms qualitative data to quantitative data. The third step finds only relevant features of a project for an effort estimation. The last step establishes a phase-wise effort estimation model using neural networks. The results show that the proposed approach utilizes only few necessary and relevant features to yield more accurate estimation than existing approaches by means of accuracy metrics and statistical hypothesis tests in performance evaluation and cross validation. Contributions of the proposed phase-wise estimation could benefit project managers in fine-grained project monitoring and control.

1313607376

Department:	Mathematics and Computer Science	Student's Signature	.....
Field of Study:	Computer Science and Information Technology	Advisor's Signature	.....
Academic Year:	2013	Co-Advisor's Signature	.....

ACKNOWLEDGEMENTS

I would like to acknowledge my thesis advisors, Associate Professor Dr. Peraphon Sophatsathit and Professor Dr. Chidchanok Lursinsap for helpful guidance and encouragement. They have suggested the solutions to many experimental problems and helped finish this thesis in time.

Finally, I wish to thank Electronic Government Agency (Public Organization) and VP Advance Company for providing software project data and administrative support, without whom the work will never be made possible.

1313607376

CONTENTS

Page

THAI ABSTRACT iv

ENGLISH ABSTRACT v

ACKNOWLEDGEMENTS vi

CONTENTS vii

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER I INTRODUCTION 1

1.1 Problem and Motivation 1

1.2 Objectives 1

1.3 Scope and Limitations 2

1.4 Contributions 2

CHAPTER II BACKGROUND AND RELATED WORK 4

2.1 Software Size Measurement 4

2.1.1 Lines of Code Measurement 4

2.1.2 Functional Size Measurement 4

2.2 Estimation Techniques 6

2.2.1 Artificial Neural Networks 6

2.2.2.1 Multilayer Perceptron 7

2.2.2.2 Fuzzy Neural Network 8

2.2.2.3 Radial Basis Function Network 10

2.2.3 Fuzzy Inference System 11

2.2.4 Support Vector Regression 11

2.2.5 Regression Analysis 13

2.2.6 Classification and Regression Tree 13

2.2.7 Analogy 14

2.3 Phase Effort Distribution 15

2.4 Estimation Techniques 16

1313607376

Page

CHAPTER III PROPOSED METHODOLOGY 19
3.1 Data Collection 19
3.2 Data Transformation 24
3.3 Feature Selection 27
3.4 Building Estimation Model and Estimating Software Effort. 33
CHAPTER IV RESULTS AND DISCUSSION 35
4.1 Performance Evaluation 35
4.1.1 Error Metrics 35
4.1.2 Statistical Hypothesis Tests 36
4.1.3 Cross-Validation 37
4.2 Parameters Set-up of Neural Networks and Other Estimation Techniques. 38
4.3 Results of Features Selection Process 39
4.4 Results of Effort Estimation 41
4.4.1 Comparison of Estimation with and without Feature Selection 45
4.4.2 Comparison of Phase-Wise and Overall Effort Estimations 49
4.4.3 Comparison of Phase-Wise and Adaptive Phase-Wise Estimations 51
4.4.4 Comparison of Phase-Wise and COCOMO II Models 52
4.5 Discussion 55
4.5.1 Likely Used and Coverage Estimation Ranges of Estimation Models. 55
4.5.2 Limitations of Estimation Models. 56
4.5.3 Effect of Programming Language 57
CHAPTER V CONCLUSION 58
REFERENCES 59
APPENDIX 63
APPENDIX A 64
APPENDIX B 67
APPENDIX C 69

1313607376

Page

APPENDIX D..... 73

VITA..... 74

131360737G

LIST OF TABLES

Table 1: An overview of effort estimation research work. 16
Table 2: Number of projects in each programming language. 20
Table 3: Number of projects in each application type. 20
Table 4: Number of projects in each application domain. 20
Table 5: Number of projects in each development type. 20
Table 6: Number of projects in each software process model. 21
Table 7: Number of projects in each intended market. 21
Table 8: Descriptive statistics of software efforts through phases. 21
Table 9: Values of rating scales in COCOMO II model. 24
Table 10: Values of rating scales in IFPUG FSM. 25
Table 11: Values of rating scales in Mk II FPA. 25
Table 12: Values of rating scales and weights in UCP. 26
Table 13: A number of features required for estimation techniques. 40
Table 14: Results of effort estimation in MdMRE. 42
Table 15: Results of effort estimation in PRED (0.25) 42
Table 16: Results of effort estimation in MdMER. 43
Table 17: Results of effort estimation in MdBRE. 44
Table 18: Results of effort estimation in MdlBRE. 44
Table 19: Estimation results of phase-wise and COCOMO II models for localization and globalization estimations in MdMRE. 53
Table 20: Required features of proposed phase-wise model. 54
Table 21: Estimation results for a project written by java android. 57

1313607376

LIST OF FIGURES

Figure 1: An example of multilayer perceptron architecture. 7
Figure 2: An example of fuzzy neural network diagram. 9
Figure 3: Architecture of a radial basis function network. 11
Figure 4: An example of linearly separable patterns. 12
Figure 5: Transformation of nonlinearly to linearly separable patterns. 12
Figure 6: An example of a decision tree. 14
Figure 7: Comparison of averaged percentage of effort distribution between PNNE, ISBSG, CSBSG, and COCOMO II data sets. 22
Figure 8: Comparison of averaged percentage of effort distribution between PNNE-1 and PNNE-2 data sets. 23
Figure 9: Scatter plot of KSLOC and software effort in different phases. 23
Figure 10: Examples of software project data transformed to quantitative data. 26
Figure 11: Examples of phase-wise and overall data sets. 27
Figure 12: Flows of proposed phase-wise and overall effort estimation. 33
Figure 13: Flows of proposed adaptive phase-wise effort estimation. 34
Figure 14: An example of 6-fold cross-validation. 37
Figure 15: Average number of features required for phase-wise and overall effort estimations. 40
Figure 16: Average number of features (prior phase) required for adaptive phase-wise effort estimation. 41
Figure 17: Comparison of estimation with and without feature selection in MdMRE. 46
Figure 18: Comparison of estimation with and without feature selection in PRED (0.25). 46
Figure 19: Comparison of estimation with and without feature selection in in MdMRE. 47
Figure 20: Comparison of estimation with and without feature selection in MdBRE. 48
Figure 21: Comparison of estimation with and without feature selection in MdIBRE. 48
Figure 22: Comparison of phase-wise and overall effort estimations. 50

1313607376

Figure 23: Comparison of phase-wise, sum of phase-wise, and overall estimations in MdMRE. 51
Figure 24: Comparison of phase-wise and adaptive phase-wise estimations. 52
Figure 25: Proposed and COCOMO II phase-wise estimation. 53
Figure 26: Likely used and coverage estimation ranges of all the proposed models. 56