

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการการวิจัยแห่งชาติ



E47287

## การปรับเขตทหารสำหรับการจําแนกประเภท

นายภาสกร ตีเขนรัมย์อนันต์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2553

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

600254252

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการการวิจัยแห่งชาติ



การปรับเฉพาะสำหรับการจำแนกประเภท



นายภาสกร ตั้งชนะชัยอนันต์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2553

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย



CUSTOMIZATION FOR CLASSIFICATION

Mr. Pasakorn Tangchanachaianan

A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2010

Copyright of Chulalongkorn University

Thesis Title      CUSTOMIZATION FOR CLASSIFICATION  
By                    Mr. Pasakorn Tangchanachaianan  
Field of Study     Computer Engineering  
Thesis Advisor    Professor Boonserm Kijirikul, Ph.D.

---

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

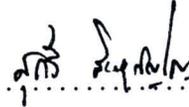
 ..... Dean of the Faculty of Engineering  
(Associate Professor Boonsom Lerdhirunwong, Dr.Ing.)

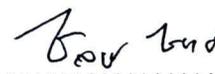
#### THESIS COMMITTEE

 ..... Chairman  
(Professor Prabhas Chongstitvattana), Ph.D.)

 ..... Thesis Advisor  
(Professor Boonserm Kijirikul, Ph.D.)

 ..... Examiner  
(Assistant Professor Chotirat Ratanamahatana, Ph.D.)

 ..... Examiner  
(Assistant Professor Sukree Sinthupinyo, Ph.D.)

 ..... Examiner  
(Assistant Professor Cholwich Nattee, Ph.D.)

ภาสกร ตั้งชนะชัยอนันต์ : การปรับเฉพาะสำหรับการจำแนกประเภท (CUSTOMIZATION FOR CLASSIFICATION) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ศ.ดร. บุญเสริม กิจศิริกุล, 76 หน้า.

**E47287**

การปรับเฉพาะเป็นกระบวนการทางการเรียนรู้ของเครื่องในการแก้ไข โมเดลที่เรียนรู้มาแล้ว เพื่อให้สามารถนำมาใช้กับข้อมูลเพิ่มเติมได้ดีขึ้น สิ่งที่แตกต่างกันจากการสร้าง โมเดลใหม่ขึ้นมาจากข้อมูลเพิ่มเติมเลยก็คือการปรับเฉพาะนั้นจะอาศัยสมมติฐานที่ว่าความรู้ โมเดลที่เรียนรู้มาแล้วนั้นจะเป็นข้อมูลที่มีประโยชน์ซึ่งจะช่วยให้การปรับปรุงผลในการเรียนรู้ของเครื่องขึ้นได้ อย่างไรก็ตาม ข้อจำกัดทั่วไปของอัลกอริทึมในการเรียนรู้เพิ่มเติมก็คืออัลกอริทึมเหล่านั้นมักจะขึ้นอยู่กับ โมเดลที่ใช้ นั่นก็ส่งผลให้จะต้องมีอัลกอริทึมใหม่สำหรับ โมเดลแต่ละประเภทและเกิดมีปัญหาในกรณีที่ไม่สามารถหาอัลกอริทึมที่เหมาะสมมาใช้ได้เนื่องจากไม่มีอัลกอริทึมสำหรับ โมเดลนั้นหรือไม่ทราบ โมเดลที่กำลังใช้อยู่

ดังนั้นเราจึงเสนอทางเลือกที่จะทำการฝึกเพิ่มเติม โดยใช้อัลกอริทึมที่ขึ้นอยู่กับประเภทของงาน เนื่องจากข้อจำกัดขั้นต่ำสุดของการที่ โมเดลนั้นจะถูกนำไปใช้ประโยชน์ได้ก็คือต้องรู้ว่า โมเดลนั้นมีไว้สำหรับงานอะไร จากงานในด้านการเรียนรู้ของเครื่องทั้งหมด เราได้นำเสนออัลกอริทึมสำหรับการจำแนกประเภท และการลดมิติ และได้ทำการทดสอบอัลกอริทึมเหล่านั้น โดยใช้ชุดข้อมูลที่สร้างจากข้อมูลที่มีอยู่จริง ผลการทดสอบแสดงให้เห็นถึงประสิทธิผลของวิธีการที่นำเสนอ

ภาควิชา ... วิศวกรรมคอมพิวเตอร์... ลายมือชื่อนิสิต .....  
 สาขาวิชา ... วิศวกรรมคอมพิวเตอร์... ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก...  
 ปีการศึกษา ..... 2553 .....

##4971822121 : MAJOR COMPUTER ENGINEERING

KEYWORDS : CUSTOMIZATION / CLASSIFICATION / DIMENSIONALITY  
REDUCTION

PASAKORN TANGCHANACHAIANAN : CUSTOMIZATION FOR  
CLASSIFICATION. ADVISOR : PROF. BOONSERM KIJSIRIKUL, Ph.D.,  
76 pp.

**E47287**

Customization is a machine learning approach in modifying a learned model to be better adapted with additional data. In contrast to training a new model from the additional data, customization relies on the hypothesis that knowing the learned model will serve as useful information which could improve the result of machine learning. However, a common limitation for customization algorithms is that they are specific on the type of model. This results in requiring new algorithms for each type of models and problem could arise if the user is unable to determine appropriate algorithms due to the lack of algorithms for that model or the difficulty in determining the type of models.

Therefore, we propose an alternative of performing customization with the algorithm that is based on the task since the minimum requirement for a model to be useful is to know the task it is for. From among all possible tasks, we present customization frameworks and algorithms for classification and dimensionality reduction and perform experiments on them using real-world dataset. This experimental results show the effectiveness of our proposed methods.

Department : . Computer Engineering . Student's Signature  .....

Field of Study : . Computer Engineering . Advisor's Signature  .....

Academic Year : .....2010.....

## Acknowledgments

First of all, I would like to express my gratitude to my advisor, for being patient and understanding and having his faith on a student like me, and also trying to support me in every way he can. I am grateful to Prof. Prabhas Chongstitvattana for his concern due to my abnormal situation and as well as other of my thesis committee members who have to give away their precious time and provide me with their comments.

I would like to thank Rattachat Chatpatanasiri, who provides a lot of advice concerning my works, Teesid Korsrilabutr, the owner of L<sup>A</sup>T<sub>E</sub>X template which I modify into my dissertation, Tanasanee Phienthrakul, for keeping me company even when there is almost no other appearing member in our lab, and also for other MIND lab members for keeping me company and giving me precious experiences. Also, I would like to extend my thanks to other graduate students in our departments whom either keep me company and even provide me with supports that come in all forms.

And most importantly, I wish to thank the Thailand Research Fund for giving me scholarship, which greatly assists me in financial issue.

# Contents

	Page
<b>Abstract (Thai)</b> . . . . .	iv
<b>Abstract (English)</b> . . . . .	v
<b>Acknowledgments</b> . . . . .	vi
<b>Contents</b> . . . . .	vii
<b>List of Tables</b> . . . . .	x
<b>List of Figures</b> . . . . .	xii
<b>Chapter</b>	
<b>I Introduction</b> . . . . .	<b>1</b>
1.1 Objectives . . . . .	3
1.2 Scope . . . . .	3
1.3 Procedure . . . . .	4
1.4 Contributions . . . . .	4
1.5 Organization of the Thesis . . . . .	4
<b>II Background</b> . . . . .	<b>5</b>
2.1 Conventions . . . . .	5
2.2 Customization . . . . .	5
2.3 Classification . . . . .	6
2.3.1 Weighted Nearest Neighbor . . . . .	6
2.3.2 Max Wins . . . . .	6
2.3.3 Adaptive Directed Acyclic Graph . . . . .	6
2.3.4 Optical Character Recognition . . . . .	7
2.3.5 Handwriting Recognition . . . . .	8
2.4 Dimensionality Reduction . . . . .	8
2.4.1 Linear Dimensionality Reduction . . . . .	10
2.4.1.1 Principal Component Analysis . . . . .	10
2.4.1.2 Linear Discriminant Analysis . . . . .	11
2.4.2 KPCA Methods of Mahalanobis Distance Learning for Nearest Neighbor . . . . .	12

Chapter	Page
2.5 Halton Sequence . . . . .	12
<b>IIITask-based Customization . . . . .</b>	<b>15</b>
3.1 Concept of Task-Based Customization . . . . .	15
3.2 Advantages and Drawbacks . . . . .	16
<b>IVCustomization for Classification . . . . .</b>	<b>17</b>
4.1 Customization for Classification by Transforming Input Space . .	17
4.1.1 Generating sequence of transformation for probabilistic classifiers . . . . .	17
4.1.1.1 Algorithm . . . . .	17
4.1.1.2 Numerical Results . . . . .	20
4.1.2 Applying with other classifiers . . . . .	21
4.1.2.1 Algorithm . . . . .	21
4.1.2.2 Numerical Results . . . . .	23
4.1.3 Overall Framework . . . . .	26
4.2 Customization for Classification by Patching . . . . .	27
4.2.1 Algorithms . . . . .	27
4.2.2 Numerical Results . . . . .	32
4.3 Customization for Classification with Nonlinear Dimension- ality Reduction . . . . .	35
4.3.1 Numerical Results . . . . .	36
<b>V Customization for Dimensionality Reduction . . . . .</b>	<b>44</b>
5.1 Framework of Customization for Dimensionality Reduction . . .	44
5.2 Combining Results from Linear Dimensionality Reduction . . . .	45
5.2.1 Algorithm . . . . .	45
5.2.2 Theoretical Results . . . . .	51
5.3 Numerical Results . . . . .	67
<b>VIConclusion and Future Work . . . . .</b>	<b>71</b>
6.1 Conclusion . . . . .	71

<b>Chapter</b>	<b>Page</b>
6.2 Future Work . . . . .	71
<b>References</b> . . . . .	<b>73</b>
<b>Biography</b> . . . . .	<b>76</b>

## List of Tables

Table	Page
2.1 First few terms in Halton sequence for two dimensional space. . . . .	13
4.1 Accuracy comparison between customization data, mapped customization data and mapped customization data using the proposed output function. . . . .	21
4.2 Accuracy from the generated classifier when using the algorithm that is applicable to non-probabilistic classifier. . . . .	23
4.3 Accuracy of each transformed subclassifier. . . . .	24
4.4 Accuracy of the combined classifier created from transformed subclassifiers. . . . .	25
4.5 Accuracy of each transformed subclassifier when using generated subclassifiers to determine the transformation. . . . .	25
4.6 Accuracy of the combined classifier created from transformed subclassifiers when using generated subclassifiers to determine the transformation. . . . .	26
4.7 Accuracy of patched subclassifiers from <i>pendigits</i> data. . . . .	33
4.8 Accuracy of patched subclassifiers from <i>optdigits</i> data. . . . .	33
4.9 Accuracy of combined classifier from patched subclassifier. . . . .	34
4.10 Accuracy of patched combined classifier. . . . .	35
4.11 Result from customization using linear dimensionality reduction algorithm. . . . .	37
4.12 Result from customization using the KMMLN algorithm. . . . .	37
5.1 Detail of all datasets used in the experiments of combining results from linear dimensionality reduction. The numbers of attributes shown are the numbers of attributes used in dimensionality reduction, omitting some useless attributes like index or specific name. . . .	67
5.2 Result from using the proposed frameworks with linear dimensionality reduction. Numbers whose values are minimal in their rows are typeset bold. The second column shows the reduced dimension. All of the equal values are different in higher precision. . . .	68

## List of Tables (cont.)

Table	Page
5.3 Classification accuracy after performing customization upon dimensionality reduction. (G) and (S) are the results from original data and customization data respectively, and (N) is the one created from both models with our algorithm. . . . .	70

## List of Figures

Figure	Page
2.1 Visualization for Adaptive Directed Acyclic Graph. . . . .	8
2.2 300 sampled data in two dimensions ( $[0, 1]^2$ ) and dispersion value measured with box space. The box shows the area which results in dispersion value. . . . .	14
4.1 Visualization of the problem in determining objective for mapping. The class of data are shown by depicting each of them as square and circle. The solid line is a likely decision boundary resulted from summing the objective value calculated from each data and the dashed line is the most suitable decision boundary. . . . .	19
4.2 Plotting of output function $s(x) = 1 - ((1 - p(x))^2)$ . . . . .	20
4.3 Visualization of results from transforming generated probabilistic classifiers. . . . .	38
4.4 Framework of customization for classification by Transforming Input Space. . . . .	39
4.5 Framework of customization for classification with nonlinear dimensionality reduction. . . . .	40
4.6 New classifier from the framework for customization of classification with nonlinear dimensionality reduction. . . . .	41
4.7 Visualization of generated dataset. . . . .	42
4.8 Embedding the data into a higher dimension. . . . .	43
5.1 Framework for customization for dimensionality reduction. . . . .	46
5.2 The new dimension reducer as the result of using the framework. . . . .	47