

การทำนายเวลาที่ใช้ในการพิมพ์ของผู้ฝึกโดยใช้
วิธีการค่าเฉลี่ยเคลื่อนที่ และแบบจำลองต้นไม้เอ็มไพร์พี

โดย

กิตติชน สุขรุจิ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

พ.ศ. 2551

Keystroke Time Interval Prediction
Using K-Means Clustering and Model Tree M5P

By

Kitichon Sukarujj

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science

Department of Computer Science

Faculty of Science and Technology

Thammasat University

2008

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

นาย กิติชน สุขขุจิ

เรื่อง

การทำนายเวลาที่ใช้ในการพิมพ์ของผู้ฝึกโดยใช้วิธีการค่าเฉลี่ยเคกลุ่ม
และแบบจำลองต้นไม้เอ็มไพรวีพี

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต


เมื่อ วันที่ 4 พฤษภาคม พ.ศ. 2552

ประธานกรรมการสอบวิทยานิพนธ์



(ดร.รัชต พิษวณิชย์)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์



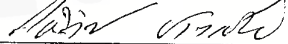
(ดร.เด่นดวง ประดับสุวรรณ)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม



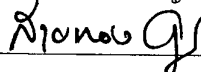
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

กรรมการสอบวิทยานิพนธ์



(ดร.กษิตศ ชาญเขียว)

คณบดี



(รองศาสตราจารย์ สายทอง อมรวีเชษฐ์)

บทคัดย่อ

งานวิจัยเสนอแนวทางการสร้างแบบจำลองเวลาที่ใช้ในการพิมพ์ของผู้ฝึกพิมพ์ดีดภาษาไทย เนื่องจากการสร้างแบบฝึกใหม่ หรือการจัดตำแหน่งแป้นพิมพ์ใหม่ ซึ่งจำเป็นต้องใช้มนุษย์มาทดลอง แต่ด้วยข้อจำกัดทางด้านเวลา และงบประมาณ ซึ่งเมื่อกลุ่มตัวอย่างผ่านการฝึกไปแล้วจะไม่สามารถใช้กลุ่มตัวอย่างนั้น ๆ มาทดลองได้อีก ผู้วิจัยจึงได้เสนอแนวทางการสร้างตัวแทนชาวนานาชาติมาทดลองแทนกลุ่มตัวอย่างดังกล่าว เพื่อให้ได้แบบฝึก หรือแป้นพิมพ์ที่เหมาะสมกับผู้ใช้แต่ละคนที่มีทักษะแตกต่างกัน

ในงานวิจัยนี้ได้เสนอวิธีการที่ใช้แบบจำลองต้นไม้ที่ใช้ข้อมูลอินพุตที่ถูกจัดกลุ่มด้วยอัลกอริทึมค่าเฉลี่ยเคลื่อนที่ เพื่อสร้างแบบจำลองที่เหมาะสมกับผู้ใช้ โดยเปรียบเทียบผลที่ได้จากวิธีการนำเสนอ กับผลการทดลองที่ได้จากแบบจำลองต้นไม้ปกติ โครงข่ายประสาทเทียม และการถดถอยเชิงเส้น โดยวัดจากรากที่สองของความผิดพลาดยกกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ผลการทดลองพบว่า วิธีการที่นำเสนอ นั้น มีค่า RMSE เฉลี่ยจากการทดสอบแบบไขว้ข้ามสิบกลุ่ม (10-fold cross validation) เท่ากับ 3.135 ซึ่งดีกว่าค่าที่ได้จากโครงข่ายประสาทเทียม และการถดถอยเชิงเส้นอย่างมีนัยสำคัญทางสถิติ

Abstract

This research proposes a new method of modeling typing time of each Thai typing learner. Because, in building new typing exercise or re-aligning keyboard layout, testing by human is needed. Due to time, cost and human constraint (that human can test only once), this research proposes a method in which an intelligent agent is trained instead of a user to construct typing lesson or keyboard layout that best fit with each user.

This research proposes an idea of using a tree model which uses clustered data by K-means algorithm to construct a regression model that best fit with each user. The experimental results comparing the ordinary Model Tree, Artificial Neural Network and Linear Regression were measured by Root Mean Square Error (RMSE). The average RMSE obtained from the proposed method is 3.135 centiseconds which is significantly better than other methods run in our experiments using 10-fold cross validation method.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดีอันเนื่องมาจากความกรุณาและอนุเคราะห์จาก ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ และดร.เด่นดวง ประดับสุวรรณ ผู้ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ให้คำแนะนำปรึกษา และข้อเสนอแนะในการทำวิจัย อีกทั้งตรวจทาน แก้ไข ข้อบกพร่องต่าง ๆ ผู้จัดทำจึงขอกราบขอบพระคุณที่ท่านอาจารย์ได้ให้ความกรุณาเสียสละเวลาอันมีค่ามารับเป็นที่ปรึกษาในครั้งนี้

ขอขอบพระคุณ ดร.รัชต พีชวณิชย์ ประธานกรรมการสอบวิทยานิพนธ์ และดร.กษิติศ ชาญเขียว กรรมการสอบวิทยานิพนธ์ ที่ให้ความกรุณาแนะนำ แก้ไข และปรับปรุงวิทยานิพนธ์ฉบับนี้ ขอกราบขอบพระคุณบิดา มารดา ผู้ให้ทั้งกำลังใจทรัพย์ และกำลังใจในการศึกษา พี่สาวที่คอยช่วยเหลือ ส่งเสริม สนับสนุน และเป็นกำลังใจตั้งแต่วันแรกที่ผู้วิจัยเข้าศึกษาในสถาบันแห่งนี้ จนถึงกระทั่งวันที่วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ เพื่อน ๆ ทุกคนที่ให้ความช่วยเหลือในการพิมพ์วิทยานิพนธ์ และกำลังใจในการจัดทำวิทยานิพนธ์นี้สำเร็จด้วยดี

กิติชน สุขรุจิ

ภาควิชาการวิทยาการคอมพิวเตอร์ มหาวิทยาลัยธรรมศาสตร์

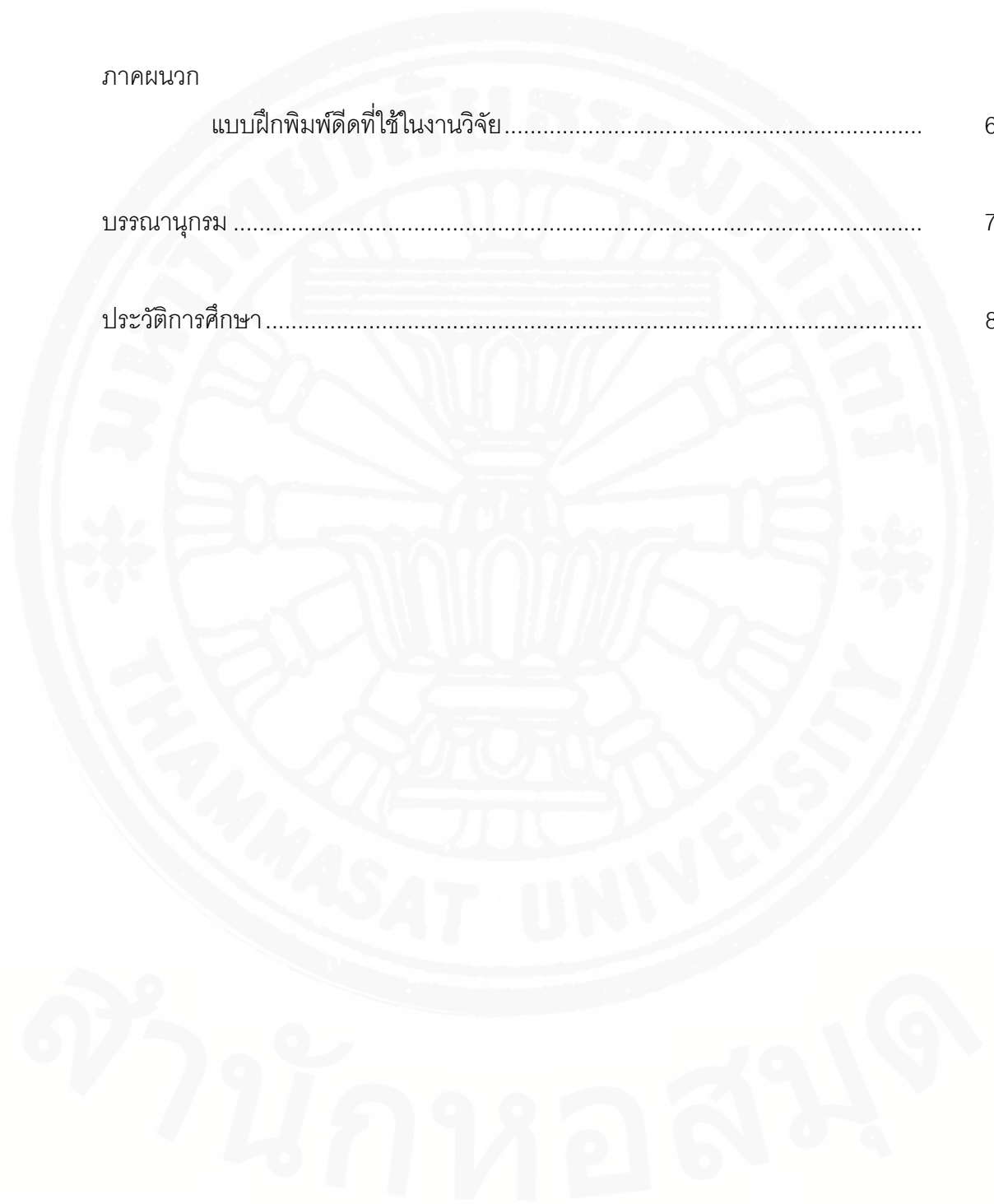
พ.ศ. 2551

สารบัญ

	หน้า
บทคัดย่อ	(1)
กิตติกรรมประกาศ.....	(3)
สารบัญตาราง.....	(7)
สารบัญภาพประกอบ.....	(8)
บทที่	
1. บทนำ.....	1
1.1 ความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์และขอบเขตของการวิจัย	4
1.3 ขอบเขตของงานวิจัย.....	4
1.4 ประโยชน์ที่คาดว่าจะได้รับ	5
1.5 รายละเอียดของวิทยานิพนธ์.....	5
2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	6
2.1 ทฤษฎีที่เกี่ยวข้อง.....	6
2.1.1 การทำงานของสมองและทักษะในการทำงาน.....	6
2.1.2 โครงข่ายประสาทเทียม	7
2.1.3 การวิเคราะห์การถดถอย	12
2.1.4 แบบจำลองต้นไม้เอนิมไฟว์พี.....	15
2.1.5 การแบ่งกลุ่มข้อมูล	19
2.1.6 ข้อมูลอนุกรมเวลา	20

2.1.7	วิธีการทดสอบแบบไขว้ข้าม 10 กลุ่ม.....	22
2.2	งานวิจัยที่เกี่ยวข้อง.....	23
3.	วิธีการดำเนินการวิจัย.....	29
3.1	ขอบเขตการทดลอง.....	29
3.1.1	แบบเรียนที่ใช้ทดลอง.....	29
3.1.2	กลุ่มผู้ทดลอง.....	29
3.2	เครื่องมือและซอฟต์แวร์ที่ใช้ในการพัฒนาระบบ.....	30
3.3	ขั้นตอนการทดลอง.....	30
3.3.1	การออกแบบสร้างแบบจำลอง.....	30
3.3.2	การจัดเตรียมข้อมูล.....	33
3.3.3	การสร้างแบบจำลอง.....	39
3.3.4	การเรียนรู้ของโครงข่ายประสาทเทียม.....	43
3.4	วิธีการวัดผลการทดลอง.....	45
3.5	สมมติฐานการวิจัย.....	46
4.	ผลการทดลอง.....	48
4.1	ผลการทดลองสร้างแบบจำลองแบบที่ 1.....	48
4.2	ผลการทดลองสร้างแบบจำลองแบบที่ 2.....	50
4.2.1	ผลการทดลองของแบบจำลองที่แบ่งกลุ่มข้อมูล แยกตามแถวของคู่อักขระ (Row-Based).....	51
4.2.2	ผลการทดลองของแบบจำลองที่แบ่งกลุ่มข้อมูล ด้วยค่าเฉลี่ยเคกลุ่ม (K-Means).....	53
4.3	ผลการทดลองสร้างแบบจำลองแบบที่ 3.....	56
5.	สรุปผลการศึกษาและข้อเสนอแนะ.....	58

5.1 สรุป	58
5.2 ข้อเสนอแนะ	59
ภาคผนวก	
แบบฝึกพิมพ์ดีดที่ใช้ในงานวิจัย.....	62
บรรณานุกรม	77
ประวัติการศึกษา.....	80



สารบัญตาราง

ตารางที่		หน้า
3.1	รายละเอียดของเครื่องมือ และซอฟต์แวร์ที่ใช้ในการพัฒนาระบบ	30
3.2	ระยะเวลาการพิมพ์ของผู้ใช้	34
3.3	ค่าข้อมูลที่ป้อนเข้าวิเคราะห์สร้างแบบจำลองด้วย WEKA.....	41
4.1	ค่าความผิดพลาดกำลังสองที่ทำนายด้วยแบบจำลองแบบที่ 1	49
4.2	ค่าความผิดพลาดกำลังสองที่ทำนายด้วยแบบจำลองแบบที่ 2 (Row-Based)	51
4.3	ค่าความผิดพลาดกำลังสองที่ทำนายด้วยแบบจำลองแบบที่ 2 (K-Means)....	54
4.4	ค่าความผิดพลาดกำลังสองที่ทำนายด้วยแบบจำลองแบบที่ 3	57

สารบัญภาพประกอบ

ภาพที่		หน้า
2.1	ฟังก์ชันการคำนวณในโหนดของโครงข่ายประสาทเทียม	9
2.2	ฟังก์ชันกระตุ้นแบบต่าง ๆ ของโครงข่ายประสาทเทียม	10
2.3	โครงข่ายประสาทเทียมแบบหลายชั้น	11
2.4	เส้นการถดถอยเชิงเส้น.....	13
2.5	รูปแบบของแบบจำลองการทำงานสมอง.....	27
3.1	ขั้นตอนการทดลองสร้างแบบจำลองรูปแบบที่ 1.....	31
3.2	ขั้นตอนการทดลองสร้างแบบจำลองรูปแบบที่ 2.....	31
3.3	ขั้นตอนการทดลองสร้างแบบจำลองรูปแบบที่ 3.....	32
3.4	รูปแบบของแบบจำลองที่พัฒนา	33
3.5	ระยะเวลาการพิมพ์.....	33
3.6	การแบ่งกลุ่มข้อมูลตามตำแหน่งแถวตัวอักษร	37
4.1	เปรียบเทียบค่าเวลาการพิมพ์ของผู้ใช้กับแบบจำลองแบบที่ 1	50
4.2	เปรียบเทียบค่าเวลาการพิมพ์ของผู้ใช้กับแบบจำลองแบบที่ 2 (Row-Based)	52
4.3	เปรียบเทียบค่าเวลาการพิมพ์ของผู้ใช้กับแบบจำลองแบบที่ 2 (K-Means)	54