การจำแนกผู้ใช้โดยใช้เวลาระหว่างการกดแป้นพิมพ์

โดย

ฉัตรทิพย์ สุทธะลักษณ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2550

User Classification Using Keystroke Time Interval

By

Chatthip Sutthalux

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2550

มหาวิทยาลัยธรรมศาสตร์
วิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

ฉัตรทิพย์ สุทธะลักษณ์

เรื่อง

การจำแนกผู้ใช้โดยใช้เวลาระหว่างการกดแป้นพิมพ์

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต

เมื่อ วันที่ 5 พฤษภาคม 2551

ประธานกรรมการสอบวิทยานิพนธ์ _____

(ดร.รัชฎา คงคะจันทร์)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์ _____

(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

กรรมการวิทยานิพนธ์ _____

(ดร.เด่นดวง ประดับสุวรรณ)

กรรมการวิทยานิพนธ์ _____

(ดร.ชลวิช นัทธี)

คณบดี _____

(รองศาสตราจารย์สมชาย วิริยะยุทธกร)

บทคัดย่อ

พลวัตการพิมพ์เป็นหนึ่งในเทคนิคการวัดทางไบโอเมตริกซ์ (Biometric) ซึ่งใช้ช่วงเวลาของการกดแป้นพิมพ์เพื่อระบุตัวผู้ใช้ในปัจจุบันมีหลายวิธีที่ใช้ในการพิสูจน์ผู้ใช้โดยวัดจากระยะเวลาที่ใช้ในการพิมพ์งานวิจัยนี้ได้นำเสนอมุมมองที่แตกต่างของพลวัตการพิมพ์ด้วยข้อความอิสระที่ใช้ในการจำแนกผู้ใช้ด้วยการรวมตัวจำแนก วิธีการนี้ใช้จำแนกว่าเป็นผู้ใช้คนใดด้วยเวลาที่ใช้ในการพิมพ์ของแต่ละบุคคลซึ่งไม่เพียงแต่จำแนกผู้ใช้แต่ยังสามารถแสดงให้เห็นถึงลักษณะเฉพาะของแต่ละบุคคลอีกด้วยในท้ายที่สุดเราเปรียบเทียบความถูกต้องที่ได้จากวิธีการของเรากับวิธีการอื่นโดยใช้เทคนิคการวัดแบบไขว้ข้ามสิบกลุ่มโดยผลการทดลองแสดงให้เห็นว่าวิธีการรวมตัวจำแนกให้ผลลัพธ์ที่ดีกว่ารวมถึงง่ายต่อความเข้าใจมากกว่าวิธีการอื่นที่นำมาเปรียบเทียบในงานวิจัยนี้

Abstract

Keystroke Dynamic, in which keystroke time intervals are used to discriminate individuals, is one of biometric measurement techniques. Nowadays, there are a number of methods used to authenticate users by measuring time interval of typing. This paper presents a different aspect of free-text keystroke dynamic which can be used to classify users by ensemble of classifiers. This method does not only classifies a person who is producing his/her sequence of keystroke, but also reveals some specific characteristics of that person. Finally, we compare the accuracy obtained from our approach to other methods run in this research by 10-fold cross validation technique. The experimental results show that the ensemble of classifiers yields the better results and is comprehensible than other methods compared in this research.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สามารถลุล่วงไปได้ด้วยความกรุณาจากผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ อาจารย์ที่ปรึกษาที่ได้สละเวลาอันมีค่าในการให้คำแนะนำ และ สั่งสอน สำหรับการเขียนวิทยานิพนธ์ การเขียนบทความวิชาการ การตรวจและแก้ไขบทความ นอกจากนี้ ข้าพเจ้าขอขอบคุณอาจารย์ ดร.รัชฎา คงคะจันทร์ อาจารย์ ดร.เด่นดวง ประดับสุวรรณ จากภาควิชาวิทยาการคอมพิวเตอร์ และ อาจารย์ ดร.ชลวิช นัทธี จากสถาบันเทคโนโลยีนานาชาติสิรินธร มหาวิทยาลัยธรรมศาสตร์ ที่ได้ให้คำแนะนำในการแก้ไขปรับปรุงวิทยานิพนธ์ฉบับนี้ให้มีความถูกต้องสมบูรณ์มากยิ่งขึ้น ขอขอบคุณพี่ๆ เจ้าหน้าที่ภาควิชาวิทยาการคอมพิวเตอร์ทุกท่านที่ให้ความอนุเคราะห์และอำนวยความสะดวกในด้านทรัพยากรต่างๆ ที่ใช้ในการจัดทำวิทยานิพนธ์ฉบับนี้ ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ปริญญาโท พี่แนน บาส บอล ตี๋ พิงค์กี้ หนุ่ย เอก น้อย แอน รวมไปถึงเพื่อน ๆ ที่มหาวิทยาลัยรามคำแหง ที่ให้คำแนะนำ ให้คำปรึกษา ให้กำลังใจ แลกเปลี่ยนความคิดเห็นในการทำวิจัย และให้ความช่วยเหลือในทุกๆ ด้าน สุดท้ายข้าพเจ้าขอกราบขอบพระคุณมารดา บิดา น้องชาย รวมถึงหัวหน้า และเพื่อนร่วมงานที่สำนักงานผู้บังคับทหารอากาศดอนเมือง ที่คอยให้กำลังใจ และความให้ความห่วงใยอย่างต่อเนื่องตลอดจนกระทั่งวิทยานิพนธ์ฉบับนี้เสร็จ

ข้าพเจ้าขอกราบขอบพระคุณเป็นอย่างสูง แก่ผู้ที่ได้กล่าวถึงและไม่ได้กล่าวถึงทุกท่าน ขอให้คุณพระศรีรัตนตรัยและสิ่งศักดิ์สิทธิ์ทั้งหลายในสากลโลก จงคุ้มครองให้ทุก ๆ ท่าน มีแต่ความสุขความเจริญ และมีสุขภาพพลานามัยสมบูรณ์แข็งแรงตลอดกาลเทอญ

ฉัตรทิพย์ สุทธะลักษณ์

นักศึกษาระดับบัณฑิตศึกษา

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยธรรมศาสตร์

พ.ศ. 2550

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... (2)

กิตติกรรมประกาศ..... (4)

สารบัญตาราง..... (6)

สารบัญภาพประกอบ ..... (7)

บทที่

1. บทนำ..... 1
    1.1 ความเป็นมาและความสำคัญของปัญหา 1
    1.2 วัตถุประสงค์ของการวิจัย 2
    1.3 ขอบเขตของงานวิจัย..... 2
    1.4 ประโยชน์ที่ได้จากงานวิจัย 3
    1.5 รายละเอียดของงานวิจัย 3
2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง..... 4
    2.1 การจำแนกประเภทข้อมูล (Data Classification) 4
        2.1.1 ต้นไม้ตัดสินใจ (Decision Tree)..... 5
        2.1.2 โครงข่ายงานประสาทเทียม (Artificial Neural Network)..... 14
    2.2 การประเมินตัวแบบ 24
        2.2.1 การประเมินตัวแบบด้วยวิธีไขว้ข้าม 10 กลุ่ม
        (10 - Fold Cross Validation) 24
        2.2.2 วิธีการประเมินตัวแบบโดยการใช้การรวมตัวจำแนก
        (Ensemble Classifier) 27
    2.3 งานวิจัยที่เกี่ยวข้อง 30

3. วิธีดำเนินงานวิจัย 36
3.1 ข้อมูลที่ใช้ในการทดลอง 36
3.2 การเก็บข้อมูลตัวอย่างที่ใช้ในการทดลอง 37
3.3 การจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ 38
3.4 การเรียนรู้ของต้นไม้ตัดสินใจ 39
4. วิธีการวัดประสิทธิภาพ และผลการทดลอง 42
4.1 การแทนต้นไม้ตัดสินใจต้นเดียวด้วยกฎ 42
4.2 การแทนต้นไม้โดยใช้การรวมตัวจำแนกด้วยกฎ 43
4.3 อธิบายวิธีการทดสอบกฎ 46
4.4 ผลการทดลอง 48
5. สรุปผลการศึกษาและข้อเสนอแนะ 50
5.1 สรุปผลการศึกษาและวิจารณ์ 50
5.2 ข้อเสนอแนะ 51
บรรณานุกรม 52
ประวัติการศึกษา 55

สารบาญตาราง

ตารางที่		หน้า
2.1	ตัวอย่างชุดข้อมูลสำหรับเรียนรู้ของคลาสเล่นกีฬาเทนนิส	6
2.2	ความสัมพันธ์ระหว่างเซลล์ประสาทกับเซลล์ประสาทเทียม.....	16
2.3	ตัวแปรต่าง ๆ พร้อมความหมาย.....	20
2.4	ชุดข้อมูลที่ได้ของวิธีแบ็กกิง	28
2.5	ชุดข้อมูลที่ได้ของวิธีบูสต์.....	29
3.1	ประโยคตัวอย่างที่ใช้สำหรับการพิมพ์เพื่อเก็บข้อมูลเวลาการพิมพ์ของผู้ใช้....	37
3.2	ตารางข้อมูลตัวอย่างเวลาระหว่าการกดแป้นพิมพ์สองอักขระ (Keystroke Interval Time)	38
4.1	ตัวอย่างข้อมูลที่นำมาทดสอบจากชุดข้อมูลสำหรับการทดสอบ.....	46
4.2	ตารางผลการทดสอบหาอัตราความถูกต้องในชุดข้อมูลสำหรับการทดสอบด้วยต้นไม้ตัดสินใจ.....	48

สารบาญภาพประกอบ

ภาพที่		หน้า
2.1	ตัวอย่างโครงสร้างต้นไม้เพื่อคาดเดาว่าจะเล่นกีฬาเทนนิสหรือไม่	7
2.2	ความสัมพันธ์ของค่าเอนโทรปี	9
2.3	เอาต์พุตต้นไม้ตัดสินใจจากการหาค่าการเพิ่มสารสนเทศขั้นแรก	11
2.4	เอาต์พุตต้นไม้ตัดสินใจจากการหาค่าการเพิ่มสารสนเทศในขั้นที่ 2	12
2.5	แสดงหลักการทำงานขั้นพื้นฐานของเซลล์ประสาทเทียม	15
2.6	เพอร์เซปตรอนแบบหลายชั้น	18
2.7	การทำงานของโครงข่ายประสาทเทียมแบบแพร่กระจายย้อนกลับ	19
2.8	สมการและกราฟแสดงการปรับเอาต์พุตเชิงเส้น	23
2.9	สมการและกราฟแสดงการปรับเอาต์พุตระหว่าง 0 ถึง 1	23
2.10	สมการและกราฟแสดงการปรับเอาต์พุตระหว่าง -1 ถึง 1	24
2.11	การแบ่งข้อมูลสำหรับการเรียนรู้และทดสอบ	25
2.12	การแบ่งข้อมูลจำนวน 10 ข้อมูลย่อย	26
2.13	การรวมตัวจำแนก	27
3.1	การจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ	38
3.2	การรวมคำตอบที่ได้จากต้นไม้ตัดสินใจ	40
4.1	ตัวอย่างต้นไม้ตัดสินใจ	42
4.2	การแทนต้นไม้ตัดสินใจแบบการรวมตัวจำแนกด้วยกฎ	43