บทที่ 2
หลักการและทฤษฎีเบื้องต้นที่นำไปสู่แนวคิดของการวิจัย

2.1 แนวคิดในการรู้จำตัวอักษรภาษาไทย

ระบบการรู้จำตัวอักษรประกอบไปด้วย Optical Scanner สำหรับอ่านภาพตัวอักษร และซอฟต์แวร์สำหรับวิเคราะห์ภาพตัวอักษร ซึ่งระบบการรู้จำตัวอักษรโดยส่วนมากจะมีการรวมเอาฮาร์ดแวร์และซอฟต์แวร์รวมเข้าด้วยกันเพื่อใช้ในการรู้จำตัวอักษร ดังนั้นหลักการทำงานของการรู้จำตัวอักษรประกอบด้วย 3 กระบวนการหลัก ๆ คือ กระบวนการก่อนการรู้จำ กระบวนการรู้จำ และกระบวนการหลังการรู้จำ ซึ่งมีความสำคัญมากทั้ง 3 กระบวนการ ดังนั้นระบบจะทำงานได้ดีหากมีข้อมูลเข้ามาอย่างถูกต้อง และผ่านการรู้จำที่มีประสิทธิภาพในการเรียนรู้รูปแบบ Pattern และจัดแบ่งประเภทของกลุ่มวัตถุเหล่านั้นได้เหมาะสม พร้อมทั้งสามารถที่จะจัดเรียงเพื่อให้ได้ไฟล์ที่ประกอบไปด้วยตัวอักษรในรูปแบบ ASCII Code ที่สามารถเปลี่ยนแปลงแก้ไขได้ด้วยโปรแกรมจัดการข้อความต่าง ๆ ได้

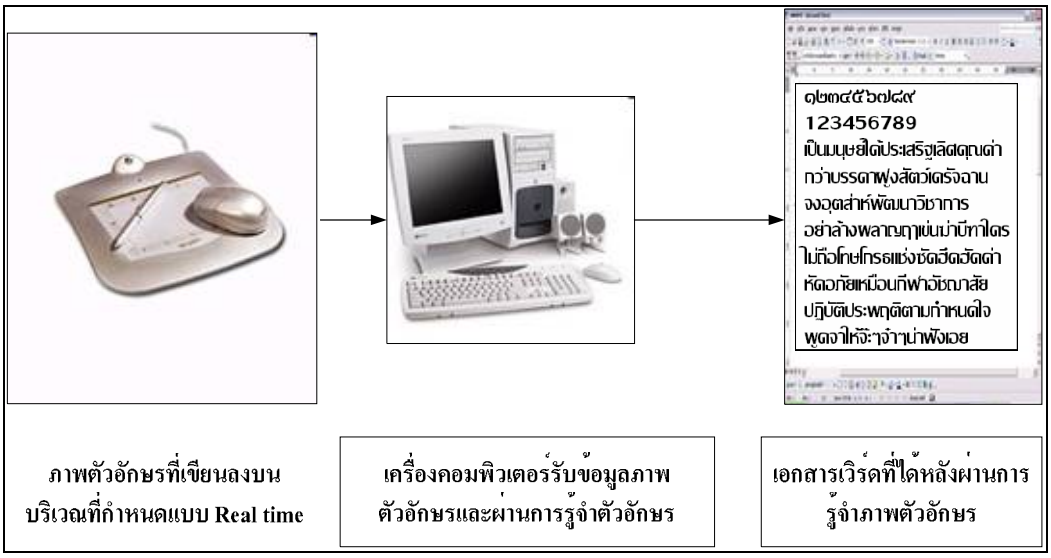
รูปที่ 2.1 ภาพแสดงแนวคิดในการรู้จำตัวอักษรภาษาไทยแบบออฟไลน์

การรู้จำตัวอักษรภาษาไทยแบ่งออกเป็น 2 ประเภทหลักคือ

1. Offline Thai Optical Character Recognition เป็นการรู้จำตัวอักษรภาษาไทยแบบออฟไลน์ของภาพตัวอักษรบนกระดาษที่ออกมาจากเครื่องพิมพ์ชนิดต่าง ๆ เช่นเครื่องพิมพ์อิงค์เจ็ต (Inkjet printer) เครื่องพิมพ์ดอตเมตริกซ์ (Dot metric printer) เครื่องพิมพ์เลเซอร์ (Laser printer) หรือเครื่องพิมพ์ดีด (typewriter) แล้วนำมาผ่านเครื่องสแกนเนอร์ จากนั้นจะได้ภาพเข้าสู่เครื่องคอมพิวเตอร์ออกมาเป็นภาพเอกสารไบนารี (Binary document images) แล้วผ่านเข้าสู่กระบวนการ

ก่อนการรู้จำโดยจะทำการจัดปรับภาพเอกสารก่อนการรู้จำให้มีคุณภาพที่ดีพร้อมส่งเข้าสู่กระบวนการรู้จำต่อไปเช่น กำจัดสัญญาณรบกวน ปรับภาพเอกสารที่เอียง เป็นต้น ดังรูปที่ 2.1

2. Online Thai Optical Character Recognition เป็นการรู้จำตัวอักษรภาษาไทยแบบออนไลน์ของตัวอักษรในขณะที่ทำการเขียนตัวอักษรนั้น ๆ ลงบนพื้นที่ที่กำหนดไว้แล้วจะทำการอ่านภาพตัวอักษรที่เขียนเข้าสู่กระบวนการรู้จำเป็นต้น ดังรูปที่ 2.2



รูปที่ 2.2 ภาพแสดงแนวคิดในการรู้จำตัวอักษรภาษาไทยแบบออนไลน์

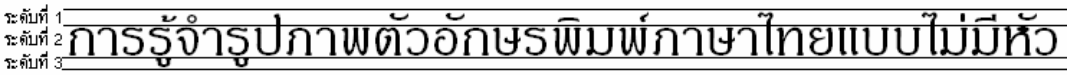
กระบวนการทำงานของการรู้จำตัวอักษรเริ่มต้นจากการวิเคราะห์ภาพและทำการแบ่งภาพออกเป็นแต่ละระดับ (Zone) พิจารณาเฉพาะบริเวณที่เป็นภาพตัวอักษรภาษาไทยประกอบไปด้วยภาพพยัญชนะ ภาพสระ ภาพวรรณยุกต์ ดังรูปที่ 2.3 จากนั้นจะแบ่งแต่ละระดับออกเป็นกรอบภาพตัวอักษรหรือเรียกว่า Character Block กระบวนการเหล่านี้เรียกว่ากระบวนการก่อนการรู้จำ และทำการส่งข้อมูลแต่ละกรอบภาพตัวอักษรเป็น Pattern เข้าสู่กระบวนการรู้จำโดยอาศัยการแบ่งกลุ่มของวัตถุออกมาเป็นกลุ่มประเภทต่าง ๆ แบ่งออกเป็น 2 วิธีหลัก ๆ คือ

<u>แบบที่ 1 Pattern Classification</u> เป็นการจัดแบ่งกลุ่มโดยที่ทราบว่าข้อมูลแต่ละ Pattern ที่เข้ามานั้นสามารถจัดแบ่งกลุ่มได้ตรงตามประเภทของ Pattern ที่กำหนดไว้

<u>แบบที่ 2 Pattern Clustering</u> เป็นการจัดแบ่งกลุ่มโดยไม่ทราบประเภทของ Pattern นั้น

สำหรับในขั้นตอนวิธีการของกระบวนการรู้จำของการแบ่งกลุ่มประเภทของ pattern ต่าง ๆ มีการนำเอาโครงข่ายประสาทเทียม (Neural Network) เข้ามาใช้ในกระบวนการนี้เรียกว่ากระบวนการรู้จำตัวอักษร ซึ่งเมื่อผ่านขั้นตอนของการรู้จำแล้ว ก็จะมีการจัดเรียงแต่ละภาพตัวอักษร

ที่รู้จำออกมาเป็นข้อมูล ASCII Code เพื่อทำการเรียงตัวอักษรที่ได้ส่วนนี้เรียกว่ากระบวนการหลังการรู้จำ



รูปที่ 2.3 การแบ่งระดับของตัวอักษรพิมพ์ภาษาไทย

สำหรับการหาค่า Histogram ในแนวนอน เพื่อแบ่งภาพตัวอักษรในแนวนอน (แยกบรรทัด)

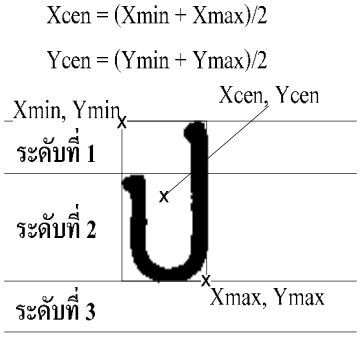
$$HorizontalHistogram(y) = \sum_{x} P(x, y) \tag{2.1}$$

สำหรับการหาค่า Histogram ในแนวตั้ง เพื่อแบ่งภาพตัวอักษรในแนวตั้ง (แยกตัวอักษรที่อยู่ในบรรทัดเดียวกัน)

$$VerticalHistogram(x) = \sum_{y} P(x, y) \tag{2.2}$$

เมื่อทำการแบ่งระดับของภาพตัวอักษรตามรูปที่ 2.3 แล้ว จะมีการแยกออกเป็นตัวอักษรเดี่ยวหรือเรียกว่ากรอบภาพตัวอักษร (Character Block) ตามตัวอย่างรูปที่ 2.4 โดยมีรายละเอียดดังนี้

- กลุ่มระดับบน คือกลุ่มตัวอักษรที่มีจุดศูนย์กลาง (Xcen, Ycen) อยู่ในบริเวณระดับที่ 1
- กลุ่มระดับกลาง คือกลุ่มตัวอักษรที่มีจุดศูนย์กลาง (Xcen, Ycen) อยู่ในบริเวณระดับที่ 2
- กลุ่มระดับล่าง คือกลุ่มตัวอักษรที่มีจุดศูนย์กลาง (Xcen, Ycen) อยู่ในบริเวณระดับที่ 3

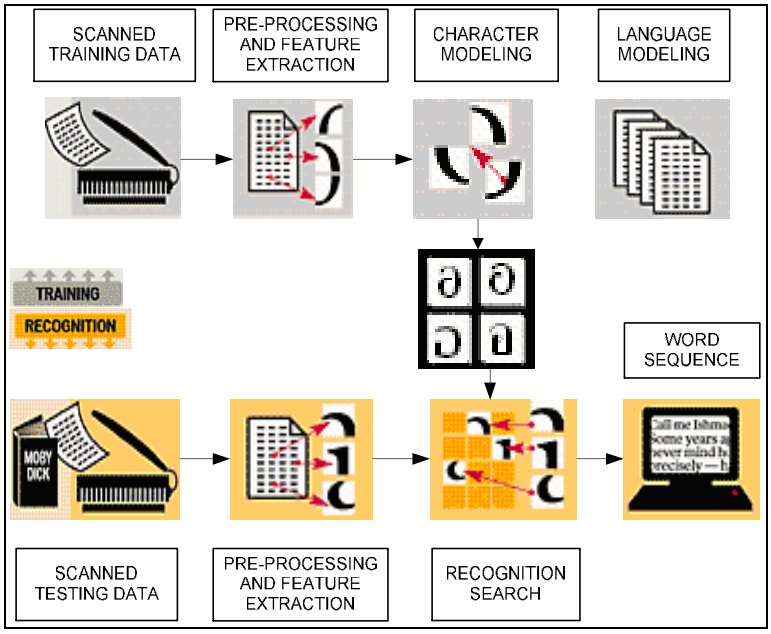


รูปที่ 2.4 การแบ่งกลุ่มระดับโดยการหาจุดศูนย์กลางของกรอบภาพตัวอักษรที่มีส่วนของการเหลื่อมล้ำเกินสองระดับ

ขั้นตอนในการรู้จำตัวอักษรภาษาไทยสามารถที่จะสรุปได้ดังรูปที่ 2.5 มีการประยุกต์ใช้งานโครงข่ายประสาทเทียม (Neural Network) สำหรับงานที่เป็น Characters Classification ซึ่งสามารถแบ่งการทำงานออกเป็น 2 ระยะคือ

1. ระยะของการเรียนรู้ (Training) เริ่มจากการสแกนข้อมูลภาพตัวอักษรจากเอกสารเข้าสู่เครื่องคอมพิวเตอร์ และผ่านเข้าสู่กระบวนการก่อนการรู้จำ (Preprocessing) ในขั้นตอนนี้อาจจะมีการดึงเอาคุณลักษณะพิเศษหรือลักษณะเด่นของวัตถุ (Pattern) ออกมา จากนั้นเข้าสู่กระบวนการรู้จำ (Character Modeling) โดยมีการจัดกลุ่ม Pattern ต่าง ๆ ที่มีความคล้ายคลึงกันมากพอให้อยู่ในกลุ่มประเภทเดียวกัน จากนั้นจะหาตัวแทนของตัวอักษรในกลุ่มนั้นแล้วแปลงเป็นรหัสแทนตัวอักษร ASCII Code เพื่อทำงานในกระบวนการหลังการรู้จำ (Language Modeling) และอาศัยแกรมมาเข้ามาช่วยในการจัดเรียงคำ

2. ระยะของการทดสอบการรู้จำ (Testing) เป็นการทดสอบในการนำไปใช้งานจริง ว่าการแบ่งกลุ่มข้อมูล Pattern ให้จัดอยู่ในประเภทเดียวกันนั้น ได้ผลออกมามีความถูกต้องมากน้อยเพียงใด โดยที่จะเอาชุดข้อมูลภาพตัวอักษรชุดใหม่ที่ยังไม่เคยใส่ในระบบในระยะของการเรียนรู้ มาเป็นข้อมูลในระยะของการทดสอบ แล้วเปรียบเทียบกับกลุ่มโดยดูจากตัวแทนของกลุ่มนั้น โดยจะได้ผลลัพธ์ออกมาเป็นกลุ่มที่ระบบการรู้จำแสดงออกมาว่ามีความคล้ายคลึงกับกลุ่มนั้นมากที่สุด



รูปที่ 2.5 แผนภาพการทำงานของระบบในการรู้จำตัวอักษร

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 HMM Topology Selection for On-line Thai Handwritten Recognition [15]

งานวิจัยนี้นำเสนอเกี่ยวกับการรู้จำลายมือเขียนตัวอักษรภาษาไทยแบบออนไลน์โดยใช้วิธีการของ Chain Code ในการทำ Preprocessing ส่วนขั้นตอนในการรู้จำใช้วิธีการของ Hidden Markov Model (Left-Right-Left) ซึ่งในงานวิจัยที่ผ่านมาวิธีการของ HMM นี้ได้มีการนำไปประยุกต์ใช้ในการรู้จำเสียง แต่สำหรับงานวิจัยนี้ได้ศึกษาการนำเอาวิธี HMM มาใช้ในการรู้จำรูปแบบที่เป็น pattern ลายมือเขียน หากมีการเปรียบเทียบลักษณะอินพุตที่เป็นข้อมูลเสียงและลายมือเขียนนั้น จะพบว่ามีลักษณะที่แตกต่างกันมากคือเสียงนั้นเป็นสัญญาณอนาล็อกที่มีลักษณะอินพุตประกอบไปด้วย Large Dimensional Vector of Continuous Real Values แต่ลายมือเขียนนั้นประกอบไปด้วย Small Dimensional Vector of Discrete Values

สำหรับการทดลองในครั้งนี้ข้อมูลจะถูกเก็บในรูปแบบ (x, y, t) ที่เกิดจากการ Sampling ในแต่ละครั้งโดย Sampling Rate เท่ากัน แต่ทั้งนี้ความเร็วในการเขียนลายมือเขียนมีผลทำให้ความสามารถในการรู้จำออกมาต่างกัน หากว่า Sampling Rate ต่ำกว่าความเร็วในการเขียนก็จะทำให้ลักษณะของลายเส้นบางส่วนที่สำคัญหายไป วิธีในการแทนค่าอินพุตจะใช้ Chain Code ประกอบไปด้วยค่าตั้งแต่ 1 – 8 ปัญหาที่พบต่อมาคือเนื่องจากบางครั้งตัวอักษรที่เขียนมีค่า Chain Code ออกมาเหมือนกัน แต่ต่างกันที่ความยาวของเส้นอักษรยกตัวอย่าง “บ” กับ “ป” จึงได้คิดวิธีปรับให้มีการบวกค่า 8 เข้าไป สำหรับการแทนค่าในช่วงกรอบล่างซ้ายและกรอบบนขวาของกรอบภาพตัวอักษร ต่อมาเป็นกระบวนการรู้จำที่อาศัยวิธี HMM โดยในการทดลองประกอบไปด้วย 3 โครงสร้างคือ Fully Connected (FC), Left Right (LR), และ Left-Right-Left (LRL)

กลุ่มข้อมูลที่ใช้ในการทดลองเป็นตัวอักษรภาษาไทย 38 ตัวอักษร ผลลัพธ์ที่ได้จากการทดลองสำหรับโมเดลแบบ LR = 82.56%, โมเดลแบบ FC=85.69% และโมเดลแบบ LRL=95.67%

2.2.2 Discriminative Training for HMM-Based Offline Handwritten Character Recognition [16]

งานวิจัยนี้มีการใช้ Discriminative Training ได้แก่ Maximum Mutual Information (MMI) พร้อมทั้งการนำเอา Composite Images ที่ทำการหมุนภาพและปรับเปลี่ยนมุมของภาพตัวอักษรลายมือเขียน และศึกษาวิธีของ Principal Component Analysis (PCA) โดยพัฒนาวิธีใหม่ขึ้นมาและเรียกว่า Block-based PCA ซึ่งจะเป็นการประยุกต์ใช้ PCA จัดการกับบริเวณที่เลื่อมล้ำกันในแนวตั้งสำหรับแต่ละกรอบตัวอักษร ในการทดลองครั้งนี้จะนำมาพัฒนาระบบการรู้จำลายมือเขียนตัวอักษรไทยที่อาศัย HMM-based System

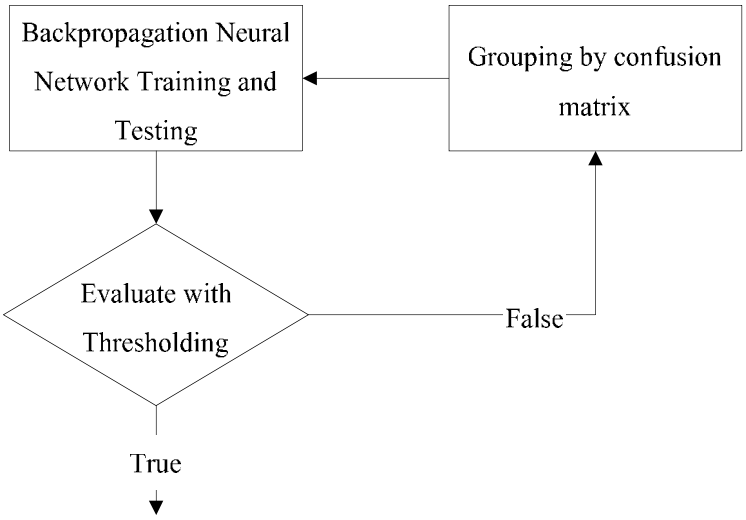
ข้อมูลที่ใช้ในการทดลองประกอบไปด้วย 77 ตัวอักษรที่เป็นพยัญชนะ สระ วรรณยุกต์ และตัวเลข โดยประกอบไปด้วยลายมือเขียนของ 20 คนที่มีวิธีการเขียนแต่ละตัวอักษรที่แตกต่างกันออกไป และ Training Set จะเลือกขึ้นมาตัวอักษรละ 120 รูปแบบจาก 20 คนที่เขียน จำนวนที่ใช้ Train และ Test ก็แบ่งเท่ากัน ซึ่งตัวอักษรจำนวน 64 ตัวอยู่ในระดับ Baseline และ 13 ตัวไม่อยู่ใน Baseline เช่น วรรณยุกต์ สระที่ปรากฏระดับบนและล่าง ผลลัพธ์จากการทดลองจะได้เปอร์เซ็นต์ของการรู้จำเท่ากับ 90.35%

2.2.3 The Clustering Technique for Thai Handwritten Recognition [17]

งานวิจัยนี้อธิบายถึงอัลกอริทึมสำหรับการจัดรวมกลุ่มตัวอักษรลายมือเขียนภาษาไทยที่มีลักษณะโครงสร้างคล้ายกันออกเป็นหมวดหมู่โดยเริ่มจากการทำ Vertical Stroke Detection โดยการพิจารณาพบว่าข้อมูลของตัวอักษรจะมีปรากฏในระดับแนวนอนมากกว่าแนวตั้ง จึงแบ่ง Character Area ออกเป็นขนาด 7x10 Blocks และมีการใช้ Global Feature ที่ได้จากการทำ Vertical Stroke ซึ่งจากการศึกษางานวิจัยนี้ ได้กำหนดขอบเขตเอาไว้แค่ในส่วนของการแบ่งกลุ่มเท่านั้น ไม่ได้ทำการรู้จำตัวอักษรออกมา

กลุ่มข้อมูลที่ใช้ทดลองคือตัวอักษรภาษาไทยที่ประกอบไปด้วยพยัญชนะ 44 ตัว สระ 17 ตัว วรรณยุกต์ 4 ตัว และสัญลักษณ์ 2 ตัว วิธีการแบ่งกลุ่มมุ่งไปที่ ลักษณะโครงสร้างตัวอักษรที่อาศัยการพิจารณาและแบ่ง Blocks ที่ไม่เท่ากันขึ้นอยู่กับ Vertical Stroke ที่ได้ จากรูปที่ 2.6 อัลกอริทึมสำหรับการจัดรวมกลุ่มตัวอักษรลายมือเขียนภาษาไทยที่มีลักษณะโครงสร้างคล้ายกันออกเป็นหมวดหมู่ เริ่มต้นจะมีการใช้ Confusion Matrix ในการวิเคราะห์ผลลัพธ์ที่ได้จาก Back-propagation Neural Network ตัวอักษรที่มีการรู้จำได้ต่ำ จะถูกจัดรวมกลุ่มเข้ากับตัวอักษรที่มีการรู้จำสูงสุด และมีการพิจารณาค่า Threshold ที่ระดับ ต่ำกว่า 70 หรือ ระหว่าง 70 กับ 90 หรือสูงกว่า 90

กลุ่มข้อมูลที่ใช้ในการทดลองได้แก่กลุ่มข้อมูล Isolate Character Set ของ NECTEC Corpus ซึ่งถูกเขียนโดยลายมือเขียนของ 68 คน แต่ละคนเขียนสองครั้ง และสแกนเข้าสู่คอมพิวเตอร์ด้วยค่าความละเอียด 200 dpi โดยมีจำนวนทั้งหมด 9,020 ตัว และแบ่งเป็น Training Set 5,940 ตัว และ Testing Set 3,080 ตัว โดยความสามารถในการจัดกลุ่มแบ่งออกเป็น 21 กลุ่มที่มีรูปแบบตัวอักษรคล้ายคลึงกันตามตารางที่ 2.1 เปอร์เซ็นต์ความสามารถในการจัดรวมกลุ่มโดยเฉลี่ยเท่ากับ 97.33%



รูปที่ 2.6 อัลกอริทึมสำหรับการจัดรวมกลุ่มตัวอักษรลายมือเขียนภาษาไทย

ตารางที่ 2.1 ความสามารถในการจัดกลุ่มของตัวอักษรที่คล้ายคลึงกันในแต่ละกลุ่ม ทั้ง 21 กลุ่ม

กลุ่มที่	**กลุ่มตัวอักษร**
1	ก ถ ภ
2	ข ฃ ฆ บ ม ย ษ
3	ค ต ด ศ
4	ง จ
5	ฉ ล ว อ า
6	ช ซ ธ ฐ ร ฮ
7	ฌ ฒ ณ ญ
8	ฏ ฎ
9	ฑ ท ห
10	น พ
11	ป
12	ผ
13	ฝ
14	ฟ
15	ฬ
16	ศ ส ษ
17	อุ อู

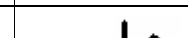
ตารางที่ 2.1 (ต่อ)

กลุ่มที่	กลุ่มตัวอักษร
18	อ๋ อ่
19	อิ อี อึ อื อั
20	อี๋ อ์ อ๊ อ้
21	โ ไ ใ

2.2.4 An Object-Oriented Expert System for Thai Character Recognition [18]

งานวิจัยส่วนมากในปัจจุบันมีเทคนิคในการจัดแบ่งโดยใช้ลักษณะโครงสร้างของตัวอักษร เช่นการแทน Code Value (Q-codes) สำหรับงานวิจัยนี้เสนอ Expert System สำหรับการรู้จำตัวอักษรไทย และใช้เทคนิค Knowledge Acquisition คือ ID3 Induction Algorithm เริ่มต้นพิจารณาที่ Character's center of gravity และ Q-codes ซึ่งเหมาะกับกลุ่มตัวอักษรที่เป็นลายเส้นต่อเนื่องกัน ตั้งแต่เริ่มต้นเขียนจนสิ้นสุดการเขียน 1 ตัวอักษร แล้วทำการกำหนด Basepoint ที่จุด Character's center of gravity จากนั้นจะมีการแบ่งกรอบตัวอักษรออกเป็น 9 ส่วน พิจารณาที่กรอบย่อยตัวอักษร 8 ส่วนโดยจะไม่นับรวมกรอบย่อยตัวอักษรที่มี Basepoint จากนั้นจะหาจุด center of gravity ของแต่ละส่วนที่ถูกแบ่งออกมา แล้วกำหนดค่าตาม Q-codes ดังตารางที่ 2.2 โดย Q-codes จะประกอบไปด้วย 16 ค่า ดังนั้นกรอบที่ถูกแบ่งออกทั้ง 9 กรอบจะมีค่าทั้งหมด 9 ค่าคือ Q1, Q2, Q3, …, Q9 แล้วใช้ Decision Tree เป็นกฎในการจัดแบ่งอักษรภาษาไทย และระบบการรู้จำพัฒนาด้วย Object Oriented Programming Language Smalltalk โดยที่ Object Oriented System Shell ที่ผ่านการรู้จำ Learning ที่ใช้ในการทดลองนี้มีชื่อเรียกว่า LEX-shell ที่ประกอบไปด้วย 4 ส่วนหลัก ๆ ได้แก่ ส่วนที่ 1 Acquisition Module จะอ่าน Training Set จาก Human Experts เข้ามา ส่วนที่ 2 The Knowledge Base จะจัดเก็บผลที่ได้จากการจัดแบ่งกลุ่มของตัวอักษรภาษาไทย ส่วนที่ 3 The Consulting Module เป็นหน้าจอส่วนติดต่อกับผู้ใช้งานระบบที่จะเชื่อมต่อเข้ากับส่วนที่ 4 The Inference Engine เพื่อผ่านการรู้จำและได้ผลลัพธ์ที่ต้องการออกมา สำหรับค่าความถูกต้องของการรู้จำนั้นขึ้นกับความสมบูรณ์และความถูกต้องในการ Training

ตารางที่ 2.2 แสดง Q-Code และรูปแบบแต่ละแบบ

Q-Code	รูปแบบ	Q-Code	รูปแบบ
0		3	
1		4	
2		5	

ตารางที่ 2.2 (ต่อ)

Q-Code	รูปแบบ	Q-Code	รูปแบบ
6		B	
7		C	
8		D	
9		E	
A		F	

2.2.5 Recognition of Handprinted Thai Characters Using the Cavity Features of Character Based on Neural Network [19]

งานวิจัยนี้อธิบายวิธี Cavity Feature และ Back-propagation Neural Network ในการรู้จำตัวอักษร ซึ่งจะใช้ Mathematic Morphology ในการหา Cavity Feature และได้ออกมาเป็น Pattern ซึ่ง Cavity Feature จะมี 6 ประเภทคือ East (E), West (W), North (N), South (S), Center (C) และ Hole (H) เมื่อเราได้ภาพตัวอักษรออกมาเป็นกรอบตัวอักษรแล้วจะทำการแบ่งออกเป็น 4 ส่วนย่อย ได้แก่ Region 1 = บนขวา Region 2 = บนซ้าย Region 3 = ล่างซ้าย Region 4 = ล่างขวา โดยที่ Feature vector ประกอบไปด้วย Cavity Image ในแต่ละพื้นที่ย่อย (H_1… H_4, C_1… C_4, N_1… N_4, S_1… S_4, E_1… E_4 and W_1… W_4) และจำนวนประเภทของ Cavity Feature คือ (N_H, N_C, N_N, N_S, N_E, N_W) และ Different levels คือ L_N แล้วผ่านเข้าสู่ Neural Network โดยที่พยายามแก้ปัญหาสำหรับตัวอักษรภาษาไทยที่เขียนด้วยตัวบรรจงและมีหัว

กระบวนการรู้จำแบ่งออกเป็น 3 ขั้นตอน ขั้นตอนแรกเริ่มจากแบ่งแยกออกจากประโยคเป็น 3 ระดับกลุ่มที่ต่างกัน จากนั้นพิจารณาที่ Cavity Feature ของแต่ละตัวอักษร แล้วคำนวณนับตัวเลขโดยใช้วิธีของ Euler Number ซึ่งขั้นตอนสุดท้ายจะใช้ Feature Code ของพื้นที่หลักที่หาออกมาส่งเข้าไปในกระบวนการรู้จำ Neural Network เพื่อแบ่งแยกแต่ละตัวอักษรออกมา แต่เนื่องจากตัวอักษรบางตัวก็ยังไม่สามารถบอกได้ว่าเป็นตัวอะไร จึงมีการใช้คุณลักษณะพิเศษแยกย่อยเช่น “ค” กับ “ด” ต่างกันที่หัวเข้า-ออก ก็ใช้ Cavity Feature พิจารณาแยกแยะออกมา และตัว “ข” กับ “บ” ก็พิจารณาระยะจากซ้าย-ขวา ความกว้างที่ไม่เท่ากัน

ผลการทดลองข้อมูลตัวอักษรไทยนั้นแบ่งออกเป็น 3 ระดับและแต่ละระดับรวมแล้วแบ่งเป็น 10 Classes ข้อมูลที่ใช้มาจากคน 40 คน แต่ละคนเขียน 80 ตัวอักษร จึงได้ข้อมูลตัวอักษรทั้งหมด 3,200 ตัว ซึ่งงานวิจัยนี้ไม่ได้ทำ Thinning character และไม่มีการทำ Edge Detection แต่มุ่งเน้นที่การพิจารณา Cavity Feature ผลของการรู้จำที่ได้คือ 98.3%

2.2.6 Off-line Handwritten Thai Characters from Word Script [20]

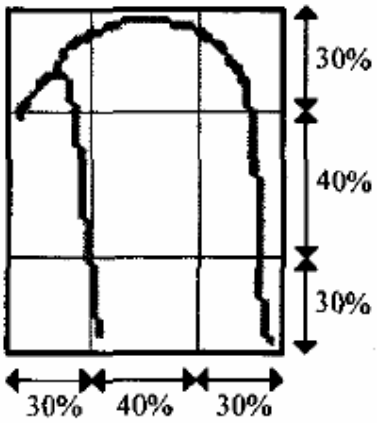
งานวิจัยนี้อาศัยโครงสร้างของ Loop ในการตัวอักษรไทยลายมือเขียนแบ่งออกเป็น 4 กลุ่มตัวอักษร และมีการใช้ Topological Properties of Stroke และลักษณะโครงสร้างของตัวอักษรภาษาไทย จากนั้นได้ออกมาเป็น Decision Tree หากแต่ว่ายังมีปัญหากับกลุ่มลายมือเขียนในรูปแบบที่ไม่มีหัว

2.2.7 Handwritten Thai Character Recognition Using Fuzzy Membership Function and Fuzzy ARTMAP [21]

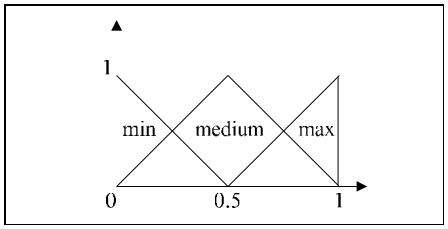
งานวิจัยนี้นำเสนอวิธีที่ใช้ในการรู้จำลายมือเขียนภาษาไทยแบบ Off-line โดยใช้หลักการของ Fuzzy Membership Function ในขั้นตอนของการทำ Feature Extraction เป็นการหาระดับความลึกของเส้นตัวอักษรจากกรอบตัวอักษร แล้วส่งผ่านเข้าสู่กระบวนการรู้จำของ Fuzzy ARTMAP โดยทั่วไปแล้วเอกสารภาษาไทยประกอบไปด้วย 5 ชนิดตัวอักษรได้แก่ พยัญชนะ สระ วรรณยุกต์ เลขอารบิก และสัญลักษณ์อื่น ๆ วิธีการเขียนตัวอักษรจะเขียนจากซ้ายไปขวา บนลงล่าง และแต่ละบรรทัดประกอบไปด้วย 4 ระดับได้แก่ above-upper level, upper level, central level และ lower level

ขั้นตอนการทำงานสำหรับการทดลองครั้งนี้แบ่งออกเป็น 6 ขั้นตอนคือ

1. สแกนเอกสารที่เขียนด้วยลายมือเขียนเข้าสู่คอมพิวเตอร์แล้วทำการแปลงเป็นรูปภาพ
2. กระบวนการก่อนการรู้จำจะแบ่งให้ได้กรอบตัวอักษรออกมา
3. การทำ Thinning Algorithm เพื่อลดขนาดของเส้นตัวอักษรให้บางลง
4. แบ่งกรอบตัวอักษรออกเป็น 9 ส่วน เริ่มจากซ้ายไปขวา และบนลงล่าง ตามรูปที่ 2.7
5. การทำ Feature Extraction ในแต่ละส่วนที่ถูกแบ่งออกจากกรอบตัวอักษรหนึ่ง โดยจะวัดค่าความลึกระหว่างเส้นตัวอักษรถึงขอบกรอบตัวอักษรที่ด้านนั้น ๆ และกำหนดระดับความลึกเป็น 5 ระดับและกำหนดตามค่าเปอร์เซ็นต์ความลึกวัดจากความกว้างของกรอบตัวอักษรได้แก่ ระดับที่ 1 มีค่า 0-20% ระดับที่ 2 มีค่า 21-40% ระดับที่ 3 มีค่า 41-60% ระดับที่ 4 มีค่า 61-80% และระดับที่ 5 มีค่า 81-100% แล้วจะทำการหาค่าเฉลี่ยออกมาสำหรับแต่ละ 9 ส่วนที่ถูกแบ่ง โดยแต่ละส่วนจะประกอบไปด้วย Left view, Top View, Right View และ Bottom View แล้วนำค่ามาพิจารณาใน Fuzzy Set DEPTH ดังแสดงตามรูปที่ 2.8
6. ขั้นตอนการรู้จำโดยใช้ Fuzzy ARTMAP Neural Network



รูปที่ 2.7 การแบ่งกรอบตัวอักษรออกเป็น 9 ส่วน



รูปที่ 2.8 The Membership Function of DEPTH

กลุ่มข้อมูลที่ใช้ในการทดลองมาจาก 2 แหล่งคือกลุ่มแรกมาจาก National Electronic and Computer Technology Center (On-line Thai Handwritten corpus) ซึ่งเป็นการรวบรวมจาก 70 คน โดยที่แต่ละคนเขียน 70 ตัวอักษร สำหรับกลุ่มที่ 2 มาจากการจัดเก็บเพิ่มเติมจากนักวิจัย ซึ่งประกอบไปด้วยลายมือเขียนจาก 4 คน ใน Training Set ประกอบไปด้วยการสุ่มขึ้นมา 50 ตัวอย่าง สำหรับการ Testing แบ่งออกเป็น 2 กลุ่มคือ กลุ่มที่ 1 เป็น 20 ตัวอย่างที่เหลือนอกเหนือจากการ Train ผลลัพธ์ของการรู้จำอยู่ที่ช่วง 39%-80% เฉลี่ยเท่ากับ 62.65% กลุ่มที่ 2 เป็นเอกสารภาษาไทยที่เขียนด้วยคน 4 คนแตกต่างกันไป ผลลัพธ์ของการรู้จำเฉลี่ยเท่ากับ 67.84%

2.2.8 Experimental results of Using Rough Sets for printed Thai Character Recognition [22]

งานวิจัยนี้นำเสนอแนวทางในการรู้จำตัวพิมพ์อักษรไทยด้วยการใช้ Rough Set กลุ่มข้อมูลที่ใช้ในการรู้จำประกอบไปด้วยพยัญชนะไทย 42 ตัว ซึ่งไม่นับรวม 2 ตัวที่ไม่มีการนำมาใช้ โดยรูปแบบของตัวอักษรที่ใช้ในการทดลองมีทั้งหมด 7 ฟอนต์ได้แก่ Angsana, Browallia, Cordia, Dillenia, Eucrosia, Freesia, Iris และแต่ละฟอนต์มีขนาดต่างกัน 7 ขนาด จึงทำให้ได้ข้อมูลที่ใช้ทดสอบในการทดลองทั้งสิ้น 2,058 ตัวอักษร สำหรับงานวิจัยที่ผ่านมา ทฤษฎีของ Rough Set ได้มี

การนำไปประยุกต์ใช้ในกลุ่มของ Data Mining แต่ในงานวิจัยนี้ได้นำมาศึกษาและใช้งานร่วมกับการทำ Pattern Recognition เริ่มแรกจะกำหนด Attributes ของสมาชิกที่อยู่ใน Rough Set จากนั้นจะสร้าง Decision Making Rules ขึ้นมาจากค่า Attributes เหล่านั้นเพื่อใช้ในการรู้จำหรือการจัดแบ่ง Object

ภาพรวมของระบบประกอบไปด้วย 4 ขั้นตอนหลักได้แก่ ขั้นตอนที่ 1 การหา Attribute ของวัตถุจะเกิดขึ้นในกระบวนการก่อนการรู้จำมีการแบ่งกรอบภาพตัวอักษรที่มีขนาด 16x32 Pixels ออกเป็นกรอบย่อยขนาด 4x4 Pixels จะได้ทั้งหมด 32 กรอบย่อย ซึ่งจะเก็บค่าได้ทั้งหมด 32 Attributes ต่อหนึ่งภาพตัวอักษร ขั้นตอนที่ 2 เป็นการลดจำนวน Attributes ที่ไม่จำเป็นทิ้งไปด้วยวิธี Boolean Reasoning Algorithm และจะเหลือไว้เฉพาะ Attributes ที่จำเป็น ขั้นตอนที่ 3 เป็นส่วนของการสร้าง Decision rules จาก Attributes ที่ได้จากขั้นตอนที่ 2 ในรูปแบบของ If … then … ต่อจากนั้นขั้นตอนที่ 4 จะทำการ Classification โดยแบ่งกลุ่มจาก Rules ที่สร้างขึ้นมา และแก้ปัญหาด้วยวิธีการ Voting Algorithm

ผลการทดลองในงานวิจัยนี้ Testing Set ประกอบไปด้วยภาพตัวอักษรจำนวน 2,058 ตัวอักษรที่มีการพิมพ์จากเครื่อง HP LaserJet 6P ที่ค่าความละเอียด 300 dpi และสแกนเข้าสู่เครื่องคอมพิวเตอร์ด้วยเครื่อง HP Scan Jet 6100c ในรูปของ Binary black / white mode แล้วมีการสร้าง Decision Rules 3 Sets โดยแต่ละ rule สร้างมาจาก Training Set 3 Sets ที่ต่างกันออกไปคือกลุ่มแรกประกอบด้วยรูปแบบตัวอักษร 1 ฟอนต์ที่มีขนาดทั้งหมด 5 ขนาด รวมทั้งหมด 210 ภาพตัวอักษร กลุ่มที่สองประกอบด้วยรูปแบบตัวอักษร 7 ฟอนต์ที่มีขนาด 1 ขนาด รวมทั้งหมด 294 ภาพตัวอักษร กลุ่มที่สามประกอบด้วยรูปแบบตัวอักษร 7 ฟอนต์ที่มีขนาดทั้งหมด 7 ขนาด รวมทั้งหมด 588 ภาพตัวอักษร ผลที่ได้จากการรู้จำในกลุ่มที่ 1, 2 และ 3 คือ 46.20%, 63.15% และ 73.12% ตามลำดับ

2.2.9 Printed Thai Character Recognition using Fuzzy-Rough Sets [23]

งานวิจัยนี้นำเสนอวิธีการใช้ Fuzzy-Rough Sets ในการรู้จำตัวอักษรภาษาไทยซึ่งประกอบไปด้วยลำดับชั้นในการจัดแบ่งกลุ่ม 2 ลำดับชั้นคือ Coarse และ Fine Classification โดยขั้นตอนวิธีการที่ใช้แบ่งกลุ่มทั้ง 2 จะใช้วิธีการเดียวกัน แต่จะต่างกันที่ข้อมูลอินพุตที่ใช้ผ่านเข้าสู่กระบวนการในการแบ่งกลุ่ม

จะเห็นได้ว่าในการแบ่งกลุ่มลำดับชั้นที่ 1 (Coarse) จะใช้วิธี Projection ในการ Extract Feature ด้วยการหาค่า Projection ของทั้งกรอบภาพตัวอักษรออกมาทั้งในแนวแกน X และแนวแกน Y หลังจากนั้นจะได้ข้อมูล 2 ชุดสำหรับแต่ละภาพตัวอักษร ส่วนการแบ่งกลุ่มลำดับชั้นที่ 2 (Fine) ก็จะใช้วิธี Projection เช่นกัน แต่จะหาค่าออกมาแค่เพียงบางส่วนของภาพตัวอักษรที่มีลักษณะคล้ายกันมาก ซึ่งจะเป็นค่าชุดข้อมูลที่จัดอยู่ใน Coarse Class ที่ยังไม่สามารถระบุผลลัพธ์ออกมาได้ ในขณะเดียวกันจากการทดลองก็มี 40 Coarse Groups ที่ไม่จำเป็นต้องส่งข้อมูลต่อขึ้นไปยังลำดับ

ชั้นที่ 2 ในขั้นตอนต่อมาข้อมูลที่ผ่านการทำ Projection แล้วจะถูก Normalization เพื่อให้มีค่าอยู่ในช่วง 0-1 หลังจากนั้นข้อมูลซึ่งถูก Normalization แล้วจะถูกกำหนดให้เป็น Attributes ของ Object ที่ใช้ใน Rough Set จากนั้นหาค่า Attribute สูงสุดและต่ำสุด

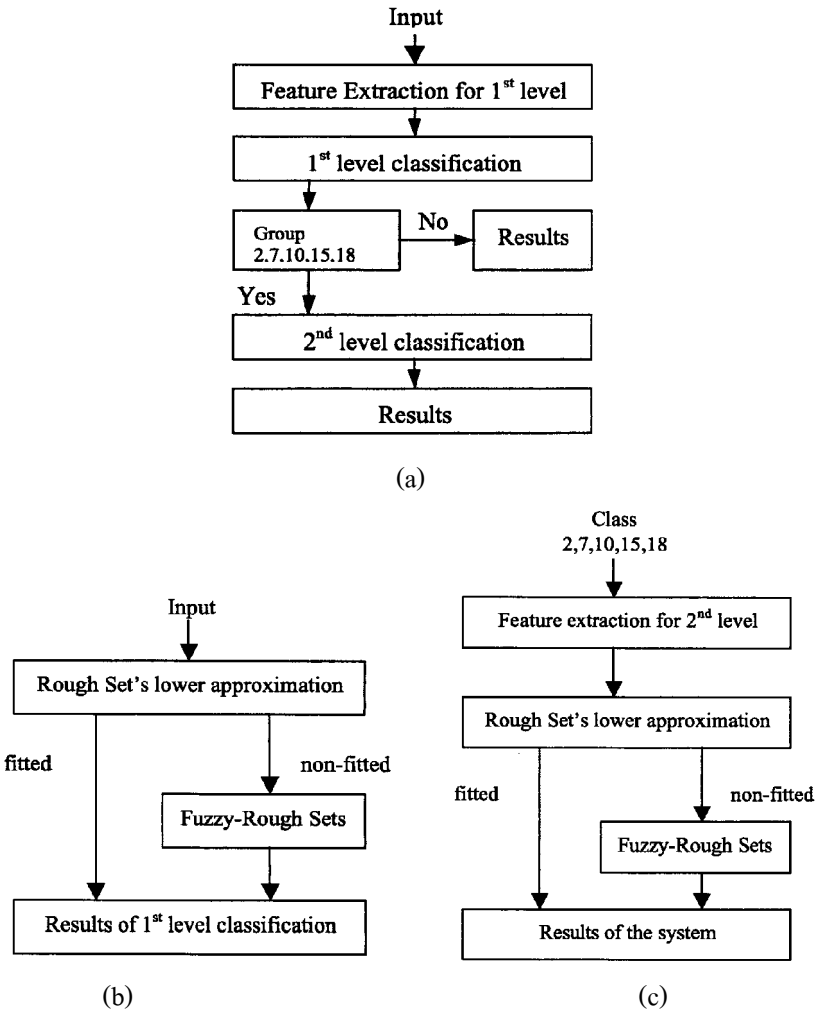
ภาพรวมของระบบแสดงดังรูปที่ 2.9 (a) เมื่อมีข้อมูลอินพุตเข้ามาจะมีการทำ Feature Extraction ของรูปแบบตัวอักษรทั้งตัวอักษรแล้วส่งเข้าสู่ขั้นตอนของการแบ่งกลุ่มลำดับชั้นที่ 1 ซึ่งประกอบด้วยขั้นตอนตามรูปที่ 2.9 (b) คือเมื่อเข้าสู่ 1st Level Classification จะใช้ Rough Set's Lower Approximation ในการกำหนดกลุ่มต่าง ๆ สำหรับการทดลองครั้งนี้มี 37 กลุ่มในการแบ่งกลุ่มลำดับชั้นแรกและในแต่ละกลุ่มจะมี 1 ตัวอักษรที่กำหนดไว้ แต่จะมีเพียงแค่กลุ่มที่ 2, 7, 10, 15, 18 ที่จะมี 2-3 ตัวอักษรที่ยังไม่สามารถระบุผลลัพธ์ออกมาได้ดังรูปที่ 2.10 โดยทั้ง 5 กลุ่มนี้จะส่งต่อเข้าไปยังการแบ่งกลุ่มลำดับชั้นที่ 2 ซึ่งประกอบด้วยขั้นตอนตามรูปที่ 2.9 (c) คือใน 2nd Level Classification จะใช้วิธี Rough Set's Lower Approximation ในการกำหนดกลุ่มซึ่งชุดข้อมูลในกลุ่ม 2, 7, 10, 18 นี้จะแบ่งได้ 2 Fine Groups แต่กลุ่มที่ 18 แบ่งได้ 3 Fine Groups และได้ผลลัพธ์ที่ต้องการออกมา สำหรับการทดลองในงานวิจัยนี้ได้กำหนดค่า Threshold เท่ากับ 0.8 หากมีกลุ่มที่ได้ค่ามากกว่าค่า Threshold จะเลือกจัดเข้ากลุ่มที่มีค่าที่ได้สูงสุด

ในการแบ่งกลุ่มจำนวนข้อมูลที่ใช้ในการทดลองคือกลุ่ม Training Set ประกอบไปด้วยรูปแบบตัวอักษร 4 ฟอนต์ แต่ละฟอนต์มีขนาด 4 ขนาด และ Testing Set ประกอบด้วยรูปแบบตัวอักษร 7 ฟอนต์ แต่ละฟอนต์มีขนาด 7 ขนาด และมีการรวมเอากลุ่มตัวอักษรของ Training Set เข้าไปด้วย ค่าความถูกต้องในการรู้จำเท่ากับ 89%

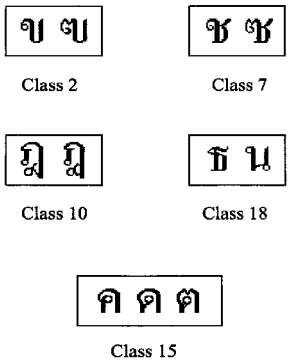
2.2.10 On Thai Character Recognition [24]

งานวิจัยนี้นำเสนอแนวทางการรู้จำตัวอักษรไทยโดยใช้การแบ่งกลุ่ม 2 ขั้นตอน ซึ่งจะใช้ Back-propagation Multi-Layered Perceptron และ Penalized Discriminant Analysis (PDA) กระบวนการก่อนการรู้จำจะนำเอาคุณลักษณะที่ต้องการออกมาจากข้อมูลด้วยการใช้วิธี Simlulated Light Sensitive (SLS) โดยมีการกำหนดค่า Threshold และเมื่อทำการ Projection กรอบภาพตัวอักษรแล้วพบว่าค่าที่ได้มีค่ามากกว่าค่า Threshold ก็จะกำหนดให้ Element State On แต่ถ้าไม่ใช่ก็จะกำหนดให้ Element State Off

จากนั้นเข้าสู่การแบ่งกลุ่มซึ่งแบ่งออกเป็น 2 ขั้นตอน ได้แก่ขั้นตอนที่ 1 จะแบ่ง Character Feature Vector ออกเป็นกลุ่มภาพตัวอักษร ทำการเปรียบเทียบกับ Representative Pattern ที่แบ่งออกเป็น 8 กลุ่มได้แก่ กลุ่มที่ 1 ประกอบด้วย “ก” “ถ” “ภ” “ห” “อ” “ฮ” กลุ่มที่ 2 ประกอบด้วย “ข” “ฃ” “ร” “ท” กลุ่มที่ 3 ประกอบด้วย “ค” “ด” “ต” “ศ” “ท” “ฑ” กลุ่มที่ 4 ประกอบด้วย “ฆ” “ม” กลุ่มที่ 5 ประกอบด้วย “ง” กลุ่มที่ 6 ประกอบด้วย “จ” “ว” “ล” “ส” กลุ่มที่ 7 ประกอบด้วย “ฉ” “น” “บ” “ป” “ผ” “ฝ” “พ” “ฟ” “ษ” “ย” “ฬ” กลุ่มที่ 8 ประกอบด้วย “ฌ” “ญ” “ฎ” “ฏ” “ฐ” “ฒ” “ณ”



รูปที่ 2.9 โครงสร้างของระบบรู้จำตัวอักษรโดยใช้ Fuzzy-Rough Sets (a) ภาพโดยรวมของระบบ (b) รายละเอียดของการจัดแบ่งกลุ่มลำดับชั้นที่ 1 (c) รายละเอียดของการจัดแบ่งกลุ่มลำดับชั้นที่ 2



รูปที่ 2.10 กลุ่มทั้ง 5 กลุ่มที่มีสมาชิกประกอบไปด้วยตัวอักษรมากกว่า 1 ตัวอักษรที่แตกต่างกัน

แล้วเข้าสู่ขั้นตอนที่ 2 เป็นการพัฒนาประสิทธิภาพของการจัดแบ่งภาพตัวอักษรย่อยในแต่ละกลุ่มจากขั้นตอนแรก หลักการที่ใช้ในการจัดแบ่งกลุ่มทั้ง 2 ขั้นตอนคือ Back-propagation Neural Network และ PDA

กลุ่มข้อมูลที่ใช้ในการทดลองคือภาพตัวอักษร 44 ตัวและมีการใส่สัญญาณรบกวนแบบ Salt and Paper เข้าไปรวมทั้งหมด 4,100 ภาพตัวอักษร และแบ่งอย่างละครึ่งคือ Training Set = 2,050 ภาพตัวอักษรและ Testing Set = 2,050 ภาพตัวอักษร กรอบภาพตัวอักษรที่จัดเก็บมีขนาด 32x32 Pixels พร้อมทั้งในการแบ่งออกเป็นกลุ่ม ๆ นั้นอาศัย Three-Layer Multi-Layer Perceptron (MLP) จากผลการทดลองพบว่าวิธีของ MLP มีการรู้จำดีกว่า PDA

2.2.11 Thai OCR: A Neural Network Application [25]

Thai OCR นับว่าเป็นหนึ่งในโปรแกรมคอมพิวเตอร์ของประเทศไทยที่ใช้ในการรู้จำตัวอักษรไทย งานวิจัยนี้นำเสนอระบบการรู้จำทั้งระบบเริ่มตั้งแต่ขั้นตอนก่อนการรู้จำ ขั้นตอนการรู้จำ และขั้นตอนหลังการรู้จำ ในการแก้ปัญหาที่อาจจะเกิดขึ้นในก่อน หรือหลังจากการรู้จำของ Thai OCR ในส่วนของผลจากการทดลองแสดงให้เห็นว่าวิธีที่ใช้ในการพัฒนา Thai OCR ทำให้ความสามารถในการรู้จำดีขึ้นอยู่ในช่วง 90% - 95% ทั้งนี้หลักการของ ANN เป็นวิธีการที่ใช้ในการรู้จำ Pattern และเป็นวิธีที่เรียนรู้การรู้จำคล้ายกับรูปแบบการเรียนรู้การรู้จำของมนุษย์ คุณลักษณะของตัวอักษรภาษาไทยประกอบไปด้วยเส้นโค้ง ซิกแซก และวงกลม รวมทั้งบางตัวอักษรก็มีลักษณะคล้ายกันเช่น “ก” “ถ” “ภ”, “ข” “ฃ” “ช” “ซ” “บ” “ษ” “ผ” “ฝ” “ป” “ฟ” “พ”, “ด” “ค” “ต” “ศ” เป็นต้น การทดลองประกอบด้วยตัวอักษรภาษาไทย 78 ตัวอักษรที่มีทิศทางการเขียนจากซ้ายไปขวา และในหน้ากระดาษจะมีทิศทางจากบนลงล่าง ซึ่งจะเขียนในบรรทัดเดียวกันแต่จะประกอบไปด้วย 4 ระดับ แล้วแต่ว่าประโยคนั้น มีตัวอักษรใดอยู่บ้าง ด้วยเหตุนี้ในการรู้จำจะยากและแตกต่างไปจากการรู้จำภาษาอื่น ๆ ทั่วไป เนื่องจากในหนึ่งบรรทัดแบ่งออกเป็นหลายระดับ นอกจากนี้ Thai OCR ยังสามารถรู้จำ English OCR ได้ด้วย ดังนั้นกลุ่มข้อมูลที่ใช้ในการรู้จำของโปรแกรม Thai OCR ประกอบไปด้วย ตัวอักษรภาษาไทย ตัวอักษรภาษาอังกฤษตัวพิมพ์เล็ก ตัวอักษรภาษาอังกฤษตัวพิมพ์ใหญ่ และตัวเลขอารบิก = 78 + 26 +26 + 10 = 140 ตัวอักษร แต่ในงานวิจัยครั้งนี้มุ่งเน้นไปที่การรู้จำตัวอักษรภาษาไทยเป็นหลัก

กระบวนการ Thai OCR เริ่มต้นจากการรับภาพเอกสารเข้าสู่เครื่องคอมพิวเตอร์ แล้วผลลัพธ์ที่ได้คือเอกสารข้อความที่ผ่านการรู้จำ โปรแกรมจึงไม่ทำงานแค่การรู้จำ แต่รวมไปถึงกระบวนการก่อนการรู้จำที่ทำการปรับภาพอักษรที่เข้ามา กำจัดสัญญาณรบกวน และอักษรเอียง การแบ่งบรรทัด และได้ออกมาเป็น Isolated Thai Character แล้วส่งผ่านต่อเข้าไปในกระบวนการรู้จำที่เป็น ANN รับเอาอินพุตภาพเข้าไปด้วยขนาดที่กำหนดไว้ในขั้นตอนก่อนการรู้จำที่ปรับให้ทุกภาพตัวอักษรที่เข้ามาไม่ว่าขนาดเท่าไหร่ก็ตามให้มีขนาดเท่าที่กำหนดเหมือน ๆ กัน จากการศึกษา

พบว่าขนาดที่เหมาะสำหรับตัวพิมพ์อักษรไทยคือ 8x23 แต่เนื่องจากบางตัวอักษรมีขนาดกว้าง และบางตัวอักษรก็มีขนาดแคบ จึงมีการปรับขนาดเพิ่มอีก 3 ขนาดคือ 8x8, 8x12 และ 8x16 หลังจากนั้นขนาดของอินพุตจะได้รับการปรับก่อนที่จะส่งเข้าสู่ระบบในการ Train และ Test จากนั้นเมื่อรู้จำตัวอักษรออกมาแล้วก็จะส่งต่อเข้าสู่การทำ Post-processing เพื่อให้ได้เอกสารข้อความออกมาในรูป Text File ที่สามารถนำไปแก้ไขต่อได้ด้วยโปรแกรม Word Processing

สำหรับการทดลอง Thai OCR พัฒนาโดย NECTEC Software Technology Laboratory ภายใต้ Microsoft Windows มีการใช้ SNNS (Stuttgart Neural Network Simulator) บนเครื่อง Pentium 90 MHz ในการจำลองโมเดลระบบเครือข่าย กลุ่มตัวอักษรภาษาไทยมี 78 ตัวอักษรที่ใช้ในการทดลองประกอบไปด้วย 10 fonts ได้แก่ AngsanaUPC, BrowalliaUPC, CordiaUPC, DilleniaUPC, EucrosiaUPC, FreesiaUPC, IrisUPC, JasmineUPC, SB Busaba, TS Burrirum แต่ละฟอนต์มีขนาด 8 ขนาดที่แตกต่างกันตั้งแต่ 8 Point ถึง 22 Point (8, 10, 12, 14, 16, 18, 20, 22) และลักษณะ 4 แบบคือ Normal, Italic, Bold, Italic & Bold รวมทั้งหมด 2,496 ตัวอักษร ความสามารถในการรู้จำอยู่ในช่วง 90% - 95%

2.2.12 On-line Thai-English Handwritten Character Recognition using Distinctive Feature [26]

งานวิจัยนี้นำเสนอการรู้จำการแบ่งกลุ่มลายมือเขียนภาษาไทยและภาษาอังกฤษแบบออนไลน์ด้วยวิธีการดึงเอาคุณลักษณะที่แตกต่างระหว่าง 2 ภาษา (Distinctive Feature Extraction) ในรูปแบบของ Decision Tree Diagram ซึ่งก่อนที่จะเข้าสู่การรู้จำจะมีการแบ่งแยกกลุ่มระหว่างภาษาไทยและภาษาอังกฤษ เพื่อพัฒนาประสิทธิภาพของการรู้จำในการลดความซับซ้อนและความถูกต้องในการรู้จำ เนื่องมาจากการรู้จำตัวอักษรของแต่ละภาษา ก็จะมีโครงสร้างตัวอักษรของแต่ละภาษาที่แตกต่างกันไป ดังนั้นหากต้องการรู้จำภาษาใด ๆ ก็เป็นเรื่องสำคัญที่จะต้องศึกษาลักษณะโครงสร้างพิเศษของภาษานั้น ๆ

สำหรับแนวคิดในงานวิจัยครั้งนี้เน้นที่การรู้จำลายมือเขียนมากกว่าหนึ่งภาษาคือภาษาไทยและภาษาอังกฤษที่เขียนลงบนอุปกรณ์ PDA (Personal Digital Assistant) หลักการพื้นฐานของกระบวนการดึงเอาคุณลักษณะที่แตกต่างกัน มีจุดมุ่งหมายเพื่อดึงเอาคุณลักษณะที่แตกต่างกันออกมาเป็นตัวบ่งบอกความแตกต่างของตัวอักษรสำหรับการกำหนดรูปแบบในการแบ่งกลุ่ม ทั้งนี้ยังไม่มีงานวิจัยที่แบ่งแยกลักษณะของตัวอักษรภาษาไทยและภาษาอังกฤษ

ในงานวิจัยนี้เริ่มต้นศึกษาที่คุณลักษณะหลักที่แบ่งกลุ่มตัวอักษรทุกตัวได้แก่ ข้อที่ 1. Number of Islands คือจำนวนชิ้นส่วนของตัวอักษรเช่น “ต” ประกอบไปด้วย 1 ชิ้นส่วนของตัวอักษรแต่ “j” ประกอบไปด้วย 2 ชิ้นส่วนของตัวอักษร ข้อที่ 2. Number of Loops คือบริเวณที่เป็นเส้นทางการเขียนที่บรรจบกันเป็นลูป โดยแต่ละตัวอักษรสามารถมีได้มากกว่า 1 ลูปขึ้นไปหรือ

อาจจะไม่มีเลยก็ได้เช่น “C” ไม่มีลูป, “ข” มี 1 ลูป , “B” มี 2 ลูป และ “ฐ” มี 3 ลูป ข้อที่ 3 Loop Head จะมีการแบ่งกรอบตัวอักษรออกเป็นสามระดับในแนวนอนที่เท่ากันโดย Loop Head สามารถอยู่ในระดับใดก็ได้เช่น “ก” ไม่มี Loop Head , “P” มี Loop Head ที่ระดับบน, “อ” มี Loop Head ที่ระดับกลาง และ “เ” มี Loop Head ที่ระดับล่าง ข้อที่ 4. Loop Connection Point คือการที่มีเส้นตรงมาสัมผัสกับลูปโดยพิจารณาจุดสัมผัสที่เกิดขึ้นได้แก่ “กำ” ไม่มีจุดเชื่อมต่อ, “ถ” มีจุดเชื่อมต่อทางซ้าย, “ภ” มีจุดเชื่อมต่อทางขวา และบางตัวอักษรมีจุดเชื่อมต่อซ้าย-ขวา

ต่อจากนั้นจะพิจารณาคุณลักษณะรองสำหรับแต่ละกลุ่มเพื่อแบ่งแยกย่อยในบางกลุ่มตัวอักษรได้แก่ ข้อที่ 1. Active Region Feature จะทำการแบ่งกรอบตัวอักษรออกเป็นขนาด 2 x 2 แล้วทำการหา Endpoint ว่าอยู่ในพื้นที่บริเวณใดในทั้งหมด 4 ส่วน จะเรียกพื้นที่ที่มีจุด Endpoint ว่า Active Region โดยกำหนด Region 1, 2, 3, 4 จากขวาไปซ้าย และบนลงล่างเช่น “Q” มี Active Region 4 ข้อที่ 2. Loop-Width-to-Loop-Height Ratio เช่นการพิจารณาระหว่าง “o” “อิ” ในขั้นตอนสุดท้ายของการรู้จำ ข้อที่ 3. Character Ripples คือลักษณะเส้นตัวอักษรที่เป็นคลื่นหยักซึ่งพบได้ทั้งในภาษาอังกฤษและภาษาไทย โดยที่ภาษาอังกฤษจะพบเฉพาะคลื่นในแนวตั้งเช่น “m” แต่ภาษาไทยส่วนมากจะพบคลื่นในแนวนอนเช่น “ย”

งานวิจัยนี้ได้พิจารณา 6 แบบที่ใช้ในการแบ่งแยกระหว่างตัวอักษรภาษาไทยและตัวอักษรภาษาอังกฤษได้แก่ แบบที่ 1 Level of Character ที่มีการแบ่งออกเป็น 3 ระดับ โดยตัวอักษรภาษาอังกฤษจะอยู่ในระดับกลางเท่านั้น แต่ตัวอักษรภาษาไทยที่ประกอบไปด้วยพยัญชนะส่วนมากจะอยู่ในระดับกลางในขณะที่สระตัวอักษรภาษาไทยจะมีอยู่ปรากฏในทั้ง 3 ระดับ แบบที่ 2 Loop-to-Character-Area Ratio พิจารณาว่าพื้นที่ลูปของตัวอักษรไทยเมื่อเทียบกับพื้นที่ตัวอักษรทั้งหมดจะมีอัตราส่วนน้อยกว่าพื้นที่ลูปของตัวอักษรอังกฤษเมื่อเทียบกับพื้นที่ตัวอักษรทั้งหมดเช่น “น” กับ “B” แบบที่ 3 Loop-to-Character-Width Ratio เมื่อพิจารณาที่ความกว้างของลูปที่เกิดขึ้นกับความกว้างของตัวอักษรแล้ว ภาษาไทยนั้นมีความกว้างของลูปจะน้อยกว่า 40% ในขณะที่ภาษาอังกฤษมีความกว้างของลูปมากกว่า 50% เช่น “บ” กับ “P” แบบที่ 4 Loop-to-Character-Height Ratio เมื่อพิจารณาที่ความสูงของลูปที่เกิดขึ้นกับความสูงของตัวอักษรแล้ว ภาษาไทยนั้นมีความสูงของลูปจะน้อยกว่า 40% ในขณะที่ภาษาอังกฤษมีความสูงของลูปมากกว่า 50% เช่น “บ” กับ “P” แบบที่ 5 Length-of-Loop-Line-to-Character-Stroke-Length Ratio พิจารณาที่อัตราของความยาวของลูปเมื่อเทียบกับความยาวของเส้นตัวอักษรตั้งแต่จุดเริ่มจนถึงจุดปลายสำหรับตัวอักษรภาษาไทยจะน้อยกว่า 40% แต่ตัวอักษรภาษาอังกฤษจะมากกว่า 50% เช่น “ด” “R” แบบที่ 6 Loop Generated Point สำหรับตัวอักษรไทยรูปแบบข้อมูลจะพบว่าเริ่มต้นด้วยลูปที่เป็นส่วนหัวของตัวอักษร ในขณะที่รูปแบบข้อมูลของตัวอักษรภาษาอังกฤษจะพบลูปที่ส่วนท้ายของตัวอักษร

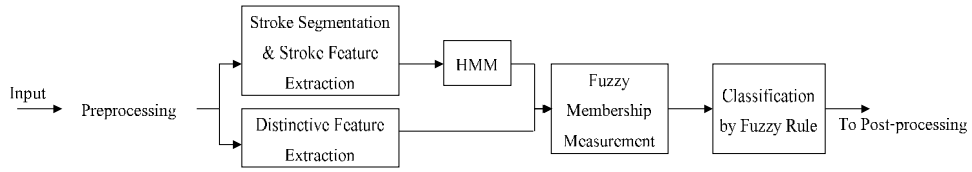
ข้อมูลที่ใช้ในการทดลองในงานวิจัยนี้ประกอบไปด้วยตัวอักษรภาษาไทย 70 ตัวได้แก่ พยัญชนะ 44 ตัวและสระ 26 ตัว และตัวอักษรภาษาอังกฤษ 52 ตัวได้แก่ตัวพิมพ์ใหญ่ 26 ตัวและ

ตัวพิมพ์เล็ก 26 ตัว ดังนั้นรวมทั้งหมด 122 ตัวอักษรที่ใช้ในการทดลอง และมีการสร้าง Decision Tree Diagram โดยกลุ่มแรกพิจารณากลุ่มตัวอักษรที่ไม่มีหัว จะพบว่ามีตัวอักษรภาษาไทยอยู่ 5 ตัว ซึ่งจะใช้ Active Region ในการแบ่งแยกย่อยต่อไป กลุ่มที่ 2 คือกลุ่มตัวอักษรที่มีลูปจะสามารถแบ่งต่อไปอีกได้สองกลุ่มคือกลุ่มที่ 3 และกลุ่มที่ 4 โดยที่ใช้ Loop-to-Character-Width Radio เป็นตัวพิจารณา ซึ่งกลุ่มที่ 3 จะมี Loop-to-Character-Width Radio น้อยกว่า 70% และพบว่ามีแต่ตัวอักษรภาษาไทย ส่วนกลุ่มที่ 4 จะมี Loop-to-Character-Width Radio มากกว่า 70% พบว่ามีทั้งตัวอักษรไทยและอังกฤษ ต่อจากนั้นนำกลุ่มที่ 4 มาพิจารณาต่อด้วย Loop-to-Character-Height Radio จะได้ว่ากลุ่มที่ 5 มี Loop-to-Character-Height Radio มากกว่า 80% ประกอบไปด้วยเฉพาะตัวอักษรภาษาอังกฤษ ส่วนกลุ่มที่ 6 มี Loop-to-Character-Height Radio น้อยกว่า 80% นั้นประกอบไปด้วยภาษาไทยและอังกฤษ จากนั้นนำกลุ่มที่ 6 มาพิจารณาด้วย Loop-Width-to-Loop-Height Ratio จะได้กลุ่มที่ 7 มี Loop-Width-to-Loop-Height Ratio น้อยกว่า 70% ประกอบไปด้วยเฉพาะตัวอักษรภาษาไทย และกลุ่มที่ 8 มี Loop-Width-to-Loop-Height Ratio มากกว่า 70% ประกอบไปด้วยตัวอักษรไทยและอังกฤษ

จากผลการทดลองสามารถรู้จำในการจัดแบ่งกลุ่มตัวอักษรภาษาไทยและตัวอักษรภาษาอังกฤษได้ 86.34% และ 95.42% ตามลำดับ และในการรู้จำทั้งภาษาไทยและภาษาอังกฤษรวมกันได้ 90.21%

2.2.13 On-line Thai Handwritten Character Recognition using Hidden Markov Model and Fuzzy Logic [27]

งานวิจัยนี้นำเสนอแนวทางการรู้จำตัวอักษรลายมือเขียนแบบออนไลน์ ที่มีการศึกษาวิธีในการรู้จำให้ดีขึ้นโดยการเพิ่ม การจัดแบ่งกลุ่มด้วย Fuzzy Logic เข้าไปก่อนที่จะใช้การรู้จำแบบ HMM เนื่องจากอัตราการรู้จำตัวอักษรมีผลมาจากการแบ่งกลุ่มตัวอักษรไทยก่อนเข้าสู่การรู้จำ ดังนั้นปัญหาหรือข้อบกพร่องที่ตามมาในขั้นตอนการจัดแบ่งกลุ่มคือความยากในการสร้างกลุ่มของกฎให้ครอบคลุมทุกรูปแบบวิธีการเขียนด้วยมือทั้งหมด โดยกฎที่กำหนดขึ้นจากการสังเกตพบว่าในหลายงานวิจัยเมื่อมีการจัดแบ่งกลุ่มแล้ว ยังคงมีตัวอักษรที่มีรูปแบบที่คล้ายคลึงกันอยู่มากเช่น “ค” กับ “ฅ” พบว่ามีแค่ส่วนที่เป็นรอยหยักที่ด้านบนของตัวอักษรที่ทำให้ตัวอักษรทั้งสองตัวต่างกัน หรือ “ก” “ภ” “ถ” ก็มีเพียงแค่จุดเริ่มต้นของตัวอักษรที่ต่างกัน เป็นเหตุให้งานวิจัยหลาย ๆ งานวิจัยที่ผ่านมาแก้ปัญหานี้ด้วยการอาศัยคุณลักษณะพิเศษเฉพาะในการจัดแบ่งกลุ่มย่อย ซึ่งในทางปฏิบัตินั้นเป็นเรื่องที่ยากในการสร้างกฎสำหรับการแบ่งกลุ่มเหล่านี้ขึ้นมาเพื่อที่จะให้ครอบคลุมทุกรูปแบบวิธีการเขียน ฉะนั้นในงานวิจัยนี้จึงนำเอา HMM มาใช้งานร่วมกับ Fuzzy เพื่อทำให้ได้ผลจากการรู้จำที่ดีขึ้น งานวิจัยนี้อธิบายแนวทางการใช้ Block Diagram ดังรูปที่ 2.11



รูปที่ 2.11 ระบบการรู้จำตัวอักษรลายมือเขียนแบบออนไลน์โดยใช้ HMM และ Fuzzy Logic

จากรูปที่ 2.11 เมื่อมีการรับค่าอินพุตที่เป็นลายมือเขียนเข้ามาแล้ว เริ่มต้นจะผ่านกระบวนการก่อนการรู้จำ หลังจากนั้นจะส่งข้อมูลของเส้นตัวอักษรลายมือเขียนที่ได้ไปยัง 2 กระบวนการได้แก่

- กระบวนการที่ 1 จะดึงเอาข้อมูลที่เกิดจากเส้นตัวอักษรทั้งตัวมาเข้าสู่การใช้ Stroke Segmentation & Stroke Feature Extraction ซึ่งจะพิจารณามุมระหว่างจุดข้อมูลบนเส้นตัวอักษรในการเขียนหากมีค่ามุมมากกว่าค่ามุม Threshold ที่กำหนดไว้ก็จะถือเอาจุดนั้นตัดออกมาจากเส้นตัวอักษรเดิมมาเป็นส่วนย่อยของตัวอักษร โดยที่หากกำหนดค่ามุม Threshold ต่ำ เส้นที่อยู่ในส่วนเดียวกันจะออกมาใกล้เคียงกับเส้นตรง ในขั้นตอนนี้ส่วนย่อยของเส้นตัวอักษรที่มีการตัดแบ่งแยกออกมาจะเรียกแต่ละส่วนว่า Stroke Segmentation

ต่อจากนั้นจะเก็บค่าลักษณะพิเศษของแต่ละส่วนแบ่งย่อย 4 ค่าได้แก่ จุดพิกัดศูนย์กลางของส่วนแบ่งย่อย มุมจากจุดเริ่มต้นถึงจุดปลายของส่วนแบ่งย่อย ค่าความแตกต่างของมุมสำหรับส่วนแบ่งย่อยนั้นกับส่วนแบ่งย่อยก่อนหน้า และความยาวของส่วนแบ่งย่อย ต่อจากนั้นใช้วิธี HMM

- กระบวนการที่ 2 จะมีการดึงเอาลักษณะพิเศษออกมาจากตัวอักษรเพื่อจัดแบ่งกลุ่มตัวอักษรภาษาไทย นับว่าเป็นวิธีการที่งานวิจัยนี้นำเสนอขึ้นมาให้มีการทำงานควบคู่กับ HMM เพิ่มไปด้วย ซึ่งจะมีคุณลักษณะพิเศษอยู่ 3 อย่างที่ใช้ในการพิจารณาคืออย่างแรก Character Head เป็นส่วนหัวมีลักษณะเป็นลูปที่เกิดขึ้นที่จุดเริ่มต้นของตัวอักษร สังเกตได้ว่าเป็นสิ่งที่แบ่งแยก “ก” เป็นตัวที่ไม่มีหัวออกจาก “ถ” ที่มีลูปเกิดขึ้นในทิศทางตามเข็มนาฬิกา และ “ภ” ที่มีลูปเกิดขึ้นในทิศทางทวนเข็มนาฬิกา อย่างที่สอง Starting and End-point Locations เช่น “ร” กับ “ว” หรือ “อ” กับ “ฮ” พบว่าแม้จะมีจุดเริ่มแบบเดียวกัน แต่มีจุดปลายของตัวอักษรอยู่ที่บริเวณต่างกัน อย่างที่สาม Curl and Notch of Character เช่น “ค” กับ “ฅ” หรือ “ม” กับ “ฆ” หรือ “ฎ” กับ “ฏ” เป็นต้น จากนั้นจะพิจารณา Fuzzy Membership Measures จากค่า Posteriori Probability สำหรับ Output ที่ออกมาจาก HMM และค่า Character Head Measure จะมาจากการแบ่งตัวอักษรแต่ละตัวออกเป็นสองส่วนคือส่วนหัวและตัวอักษร โดยพิจารณาจากการใช้การแบ่ง Sub-Stroke และค่า Starting and End-point Locations Measure ซึ่งจะกำหนดแบ่งกรอบตัวอักษรออกเป็น 9 ส่วน คือกรอบที่มีขนาด 3x3 จากนั้นพิจารณาเป็น 2 ส่วนคือแนวนอนและแนวตั้ง โดยพื้นที่แนวนอนประกอบด้วย Left Variable, Center Variable, Right Variable ส่วนพื้นที่แนวตั้งประกอบด้วย Top Variable, Middle

Variable, Bottom Variable และค่า Curl and Notch Measure จะพิจารณาจากรอยหยักที่เกิดบนแต่ละส่วนของทั้งตัวอักษรและสามารถที่จะแบ่งกลุ่มได้สามกลุ่มคือกลุ่มที่ 1 รอยหยักที่ด้านบนตัวอักษรได้แก่ “ฃ”, “ต”, “ฆ”, “ฑ”, “ฒ”, “ซ” และ “ต” กลุ่มที่ 2 รอยหยักที่ด้านล่างตัวอักษรได้แก่ “ฏ”, “ฐ”, “ผ”, “ฝ”, “พ”, “ฟ” และ “ฬ” กลุ่มที่ 3 รอยหยักที่ด้านซ้ายของตัวอักษรได้แก่ “ย” ต่อจากนั้นเป็นขั้นตอนของการจัดแบ่งกลุ่มตาม Fuzzy Rule โดยพิจารณารวมทั้ง 3 ค่าข้างต้นที่ได้ของแต่ละตัวอักษรค่าอินพุตที่เข้ามา และพิจารณาจากค่าที่น้อยที่สุดที่ได้ออกมาเพื่อหาผลลัพธ์ที่ได้จากการรู้จำตัวอักษรนั้น

ผลการทดลองตัวอักษรที่ใช้ในการทดลองมาจาก 44 ตัวอักษรไทย โดยกลุ่มข้อมูลใน Training Set ประกอบด้วยจำนวนตัวอักษรทั้งสิ้น13,608 ตัวอักษรจากคนที่เขียนลายมือเขียน 37 คน สำหรับข้อมูลที่ใช้ใน Testing Set ประกอบด้วย 7,664 ตัวอักษรจากคนที่เขียนลายมือเขียน 21 คน ผลจากการทดลองเมื่อเทียบกับ HMM มีความสามารถในการรู้จำได้ 89.3% แต่วิธีที่ศึกษาและนำเสนอด้วย HMM + Fuzzy นั้นมีความสามารถในการรู้จำได้ถึง 92.1%

2.2.14 Online Thai Handwritten Character Recognition using Hidden Markov Models and Support Vector Machines [28]

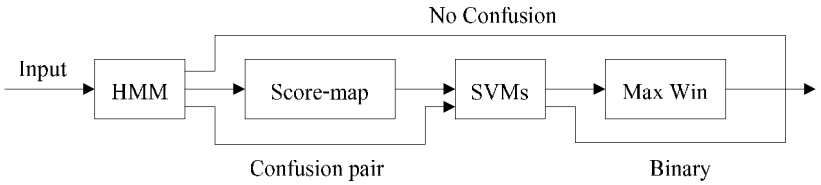
งานวิจัยนี้นำเสนอแนวทางของวิธี Support Vector Machines (SVM) เข้ามาใช้งานร่วมกับ Hidden Markov Models (HMM) โดยพิจารณาจาก Confusion Matrix ของ HMM หากพบว่า Confused Candidates มีค่ามากกว่าหนึ่งค่าแสดงว่ายังไม่สามารถรู้จำกลุ่มนั้นออกมาได้ ก็จะใช้วิธี SVM เข้าทำงานร่วมด้วยเพื่อการเพิ่มประสิทธิภาพในการรู้จำ

ระบบประกอบไปด้วยขั้นตอนก่อนการรู้จำ ลักษณะของข้อมูลที่ได้มาจากลายมือเขียนแบบออนไลน์นั้นจะมีข้อมูลที่ได้มาคือลำดับของจุดพิกัดที่เกิดจากการเคลื่อนที่ของปากกาเขียนที่รับเข้ามาจากอุปกรณ์รับข้อมูลเช่น Tablet อย่างไรก็ตามไม่ใช่ว่าทุกจุดจะถูกเก็บและส่งเข้ามา แต่จะพยายามลดจำนวนจุดที่รับเข้ามา ดังนั้นในกระบวนการนี้จะลดจุดพิกัดที่ซ้ำกันบางส่วนออกไปหรือนับได้ว่าเป็นจุดที่เปลี่ยนแปลงน้อยที่สุดออก เพื่อที่จะหาระยะทางของเส้นตัวอักษรจากลายมือเขียนให้ได้ระยะที่สั้นที่สุด จากนั้นจะเข้าสู่กระบวนการ Feature Extraction ลำดับข้อมูลของแต่ละตัวอักษรจะถูกแบ่งออกเป็น Sub-Stroke ด้วยวิธี Stroke Segmentation และจะจัดเก็บค่า 4 ค่าออกมาได้แก่ จุดพิกัดศูนย์กลางของส่วนแบ่งย่อย มุมจากจุดเริ่มต้นถึงจุดปลายของส่วนแบ่งย่อย ค่าความแตกต่างของมุมสำหรับส่วนแบ่งย่อยนั้นกับส่วนแบ่งย่อยก่อนหน้า และความยาวของส่วนแบ่งย่อยแล้วส่งต่อเข้าสู่ HMM ที่มีโครงสร้างโมเดล Left-to-Right with No Skip และแต่ละ State Output ที่แบ่งแยกออกมาได้คือ Gaussian Components ของ Score-map เนื่องจาก SVM เป็น Binary Classifier ดังนั้นจึงมีการประยุกต์เอา SVM มาใช้ใน Multi-Classification โดยเลือกใช้ Max Win Algorithm เพราะจำนวนของกลุ่มที่ได้มีขนาดไม่ใหญ่ พิจารณาจาก HMMs Confusion matrix และ

ผลลัพธ์จากการรู้จำจะจัดอยู่ในกลุ่มที่มีค่า Maximum Number of Votes ทั้งนี้ความยาวของขนาดอินพุตที่รับเข้ามามีค่าไม่คงที่แน่นอน และวิธีการใช้ SVM จะต้องกำหนดให้ค่าอินพุตที่เข้ามานั้นมีขนาดที่คงที่เท่ากัน

งานวิจัยนี้จึงนำเสนอการทำ Normalization of the Feature Vectors ที่ประกอบด้วย 4 ขั้นตอนได้แก่ Sequence Length Normalization, Normalizing the Score-maps for Training and Testing, Score-space Whitening และ Spherical Normalization ในระบบที่นำเสนอในงานวิจัยนี้เป็นไปตามรูปที่ 2.12 HMMs ใช้ในการจัดแบ่งลำดับข้อมูลที่ได้รับเข้ามาจากอินพุตในขั้นตอนแรกและวิเคราะห์คาดคะเน Model Parameter สำหรับ Score-map ออกมา จากนั้น SVMs with Score-space Kernel จะเข้ามาช่วยในการแบ่งกลุ่มที่ยังไม่สามารถระบุผลลัพธ์ในการรู้จำในขั้นตอนที่สองโดยมีการใช้ Max Win Algorithm ใน Multi-class ถ้าจำนวน Input Candicate of SVMs มีระบุมากกว่า 2 Classes

สำหรับผลการทดลองครั้งนี้ข้อมูลที่ใช้คือตัวอักษรภาษาไทย 42 ตัวอักษร โดยไม่นับรวม “ฃ” และ “ฅ” เพราะปัจจุบันไม่มีการนำมาใช้แล้ว ในกลุ่ม Training Set ประกอบด้วยลายมือเขียนจาก 31 คน รวมทั้งหมด 14,557 ตัวอักษร และกลุ่ม Testing Set ประกอบด้วยลายมือเขียนจากใน Training Set จำนวน 10,844 ตัวอักษร และรวมกับ 7,812 ตัวอักษรจากลายมือเขียนของ 62 คน ผลจากการรู้จำโดยใช้วิธีการของ HMM = 89.9% แต่สำหรับการทดลองครั้งนี้มี 2 วิธีคือ HMM/SVM/LR (HMM/SVM with Likelihood Ratio Score-space) = 88.1% และ HMM/SVM/SLR (HMM/SVM with Symmetric Likelihood Ratio Score-space) = 92.5%



รูปที่ 2.12 ระบบการรู้จำโดยอาศัยวิธี Support Vector Machines และ Hidden Markov Models

2.2.15 Character Recognition System for Cellular Phone with Camera [29]

งานวิจัยนี้นำเสนอระบบการรู้จำตัวอักษรด้วยกล้องถ่ายรูปที่นำไปประยุกต์ใช้งานบน Mobile Device อย่างเช่น PDA และโทรศัพท์มือถือที่มีกล้องดิจิตอลสี โดยเริ่มจากการพัฒนาระบบการรู้จำตัวอักษรด้วยกล้องถ่ายรูปสำหรับเครื่อง PC โดยอาศัยเทคโนโลยีต่าง ๆ เช่น Image Enhancement, Local Adaptive Binarization และ Blob Coloring ในการดึงภาพพื้นที่ตัวอักษรและกำจัดสัญญาณรบกวนที่อาจจะเกิดจากการจับภาพจากกล้องถ่ายรูป หลังจากนั้นจะแปลงระบบ OCR

System บนเครื่อง PC ให้อยู่ในรูปแบบ Embedded OCR System สำหรับโทรศัพท์มือถือ โดยที่มี Function การทำงานมากมายที่พัฒนาขึ้นมาโดยเฉพาะที่ใช้งานบนอุปกรณ์สื่อสารเคลื่อนที่ ที่ไม่ได้มีการคำนวณเกิดขึ้นจริงเนื่องมาจากพื้นที่ของหน่วยความจำและทรัพยากรของเครื่องที่มีอยู่อย่างจำกัด ในงานวิจัยนี้มุ่งเน้นไปที่ปัญหาต่าง ๆ เหล่านี้และกลไกการทำงานของวิธีการรู้จำที่นำเสนอ ในอดีตการใช้งานของโทรศัพท์คือการพูดคุยสนทนาและต่อมาการสื่อสารทางมือถือได้รับความนิยมมากขึ้น จุดมุ่งหมายกลายมาเป็นการจัดการและจัดเก็บข้อมูลส่วนตัวบนโทรศัพท์มือถือ รวมไปถึงการพัฒนารวมเอากล้องดิจิตอลเข้ากับโทรศัพท์มือถือหรืออุปกรณ์สื่อสารแบบเคลื่อนที่ไปมาได้ อย่างเช่น PDA และสิ่งที่พัฒนาตามมาก็คือโปรแกรมประยุกต์ต่าง ๆ เช่น Navigation System, Smart Tour Guide, Robot Automatic Traveling เป็นต้น อย่างไรก็ตามในการจับภาพ (Capture) ของกล้องยังคงมีสัญญาณรบกวนที่เกิดจากความสว่างและความคมชัดต่ำอันเนื่องมาจากระดับแสงที่ไม่คงที่ จึงทำให้ยากในการหาบริเวณของพื้นที่ตัวอักษร และรู้จำตัวอักษรนั้น ๆ

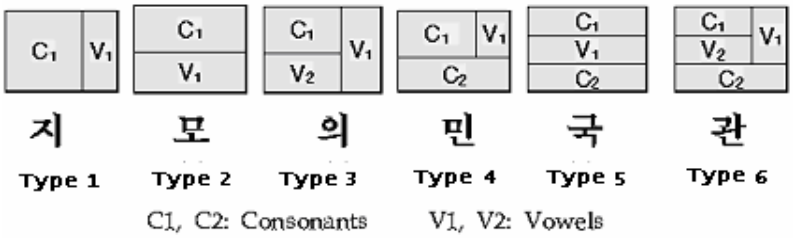
ในงานวิจัยนี้เป็นการรู้จำตัวอักษรจากการถ่ายภาพเอกสารด้วยกล้องดิจิตอล และนำเสนอวิธี Locally Adaptive Binarization รวมทั้งการทำ Image Enhancement Algorithm ก่อนที่จะทำ Binarization เพื่อช่วยในการปรับสภาพแวดล้อมของแสงที่เกิดให้คงที่สม่ำเสมอ ระบบการรู้จำตัวอักษรสำหรับภาพจากกล้องถ่ายรูปบนมือถือนั้นประกอบไปด้วย 3 ขั้นตอนหลักได้แก่ Preprocessing, Segmentation และ Recognition โดยที่กระบวนการ Preprocessing นั้นจะแตกต่างจากการจัดการกับภาพที่สแกนเข้ามาสู่คอมพิวเตอร์ตรงที่ภาพที่ถ่ายจากกล้องนั้นค่อนข้างมีความแตกต่างและความสว่างของภาพที่ได้ไม่เท่ากันทำให้สิ่งเหล่านั้นเกิดเป็นสัญญาณรบกวนที่ทำให้ยากต่อการค้นหาพื้นที่บริเวณของตัวอักษรออกมาได้ ประสิทธิภาพของกระบวนการก่อนการรู้จำจึงขึ้นอยู่ที่ความสามารถในการกำจัดเอาสัญญาณรบกวนเหล่านี้ออกไปได้มากน้อยเพียงใด ดังนั้นในขั้นตอนของกระบวนการก่อนการรู้จำนี้แบ่งออกเป็นการแปลงภาพที่ถ่ายมาให้เป็น Gray Level Image แล้วมีการทำ Image Enhancement จากนั้นเข้าสู่การทำ Binarization โดยใช้วิธี Locally Adaptive Threshold เพื่อที่จะสามารถแยกภาพตัวอักษรออกจากพื้นที่ข้างหลังได้ส่วนหนึ่ง ต่อจากนั้นจะทำการกำจัดสัญญาณรบกวนด้วยกระบวนการ Blob Coloring โดยวิเคราะห์จากตำแหน่งและขนาดของ Blob ที่เกิดขึ้น และเมื่อผ่านกระบวนการ Preprocessing แล้วก็จะเป็นกระบวนการ Segmentation ซึ่งจะดึงเอาเฉพาะกรอบภาพตัวอักษรที่ใช้ในการรู้จำตัวอักษรแต่ละตัวออกมา โดยจะแบ่ง Binary Image ที่ได้ออกเป็นบรรทัด จากนั้นแบ่งออกเป็นแต่ละคำและแบ่งออกเป็นแต่ละตัวอักษร โดยในงานวิจัยนี้ใช้วิธี Projection ในการทำ Segmentation โดยเริ่มจากการทำ Projection ในแนวนอนเพื่อให้ได้แต่ละบรรทัดแยกออกมาจากกัน จากนั้นแบ่งแต่ละคำออกจากกันโดยอาศัยการทำ Projection ในแนวตั้งจะได้แต่ละคำออกมา และใช้การทำ Projection แนวตั้งในแต่ละคำก็จะได้เป็นกรอบตัวอักษรแต่ละตัวออกมา

ตัวอักษรที่ใช้ในการรู้จำสำหรับงานวิจัยนี้คือตัวอักษรภาษาเกาหลีที่มีชื่อเรียกว่า “Hangul” โดยโครงสร้างของภาษาประกอบไปด้วยพยัญชนะและสระ แบ่งแยกลักษณะโครงสร้างของตัวอักษรออกเป็น 6 รูปแบบตามรูปที่ 2.13

ขั้นตอนของกระบวนการรู้จำ แต่ละกรอบภาพตัวอักษรที่ได้จากการทำ Segmentation จะมีการดึงเอา Feature ออกมาแล้วส่งเข้าสู่ระบบการรู้จำที่อาศัยวิธี Neural Network โดยที่จะมีการทำ Normalization เพื่อให้กรอบภาพตัวอักษรแต่ละภาพมีขนาดเท่ากัน จากนั้นทำ Feature Extraction ด้วยการใช้ Mesh Feature คือการรวมพิกเซลของแต่ละส่วนของตัวอักษรและใช้ Chain Code ที่ประกอบด้วย 8 ทิศทาง โครงข่ายประสาทเทียมที่ใช้ในการทำ Type Classification นั้นประกอบไปด้วย Input Node จำนวน 256 โหนด Hidden Node จำนวน 100 โหนด และ Output Node จำนวน 6 โหนด ซึ่งเป็นกลุ่มโครงสร้างของตัวอักษรเกาหลีทั้ง 6 รูปแบบ จากนั้นจะส่งกรอบภาพตัวอักษรในแต่ละกลุ่มเข้าไปยังระบบการรู้จำตัวอักษร โดยอาศัยโครงข่ายประสาทเทียม ทั้งนี้ขนาดของ Input Vector ทั้ง 6 กลุ่มนั้นจะแตกต่างกันไปดังนี้ 314, 120, 76, 54, 520, 301 โหนดตามลำดับและ Output ที่ได้จากการรู้จำมีจำนวน 55 Neurons ในแต่ละกลุ่มตามรูปที่ 2.14

งานวิจัยนี้มุ่งเน้นที่จะพัฒนาและประยุกต์วิธีการรู้จำตัวอักษรบนเครื่อง PC ไปใช้ในการรู้จำบนอุปกรณ์สื่อสาร มือถือ โดยภาพรับเข้ามาจากกล้องถ่ายรูปที่มีอยู่บนตัวอุปกรณ์เหล่านั้น แต่เนื่องด้วยหน่วยความจำและขีดความสามารถในการคำนวณของอุปกรณ์เหล่านั้นมีขีดจำกัด อาทิเช่นในการคำนวณ Sigmoid Function นั้นเป็น Exponential Function ซึ่งอุปกรณ์มือถือไม่สามารถที่จะคำนวณได้ จึงต้องทำการพัฒนาให้อยู่ในรูปแบบ Embedded OCR System โดยการแปลง Real Number Operation ไปเป็น Integer Operation พิจารณาที่การลดอัตราการรู้จำที่ไม่ถูกต้อง และมีการเปลี่ยน 64 Bits Decimal Data เป็น 32 Bits Integral Data นอกจากนี้ยังต้องมีการทำ Approximation of Sigmoid Function

จากการทดลองนี้มีการใช้ระบบการรู้จำตัวอักษรบน WIPI (Wireless Internet Platform for Interoperability) ภายใต้การพัฒนาและได้รับความร่วมมือจาก SK Telecom Korean Company กลุ่มข้อมูลที่ใช้ในการทดลองมาจาก ETRI (Electronics and Telecommunications Research Institute) Database โครงข่ายประสาทเทียม MLP (Multi-layer perceptron) ประกอบด้วย 256 Input Neurons, 100 Hidden Neurons และ 6 Output Neurons ในการแบ่งกลุ่มโครงสร้างออกเป็น 6 กลุ่ม Training Set ประกอบไปด้วย 11,260 ตัวอักษร Testing Set ประกอบไปด้วย 6,756 ตัวอักษร ค่าความละเอียดของกล้องดิจิตอลเท่ากับ 1,280 x 1,024 Pixels จากผลการทดลองในการแบ่งกลุ่มตามรูปแบบ 6 รูปแบบนั้นมีค่าความถูกต้องโดยเฉลี่ย 99.19% และค่าเฉลี่ยของการรู้จำตัวอักษรของทั้ง 6 รูปแบบเท่ากับ 96.82%

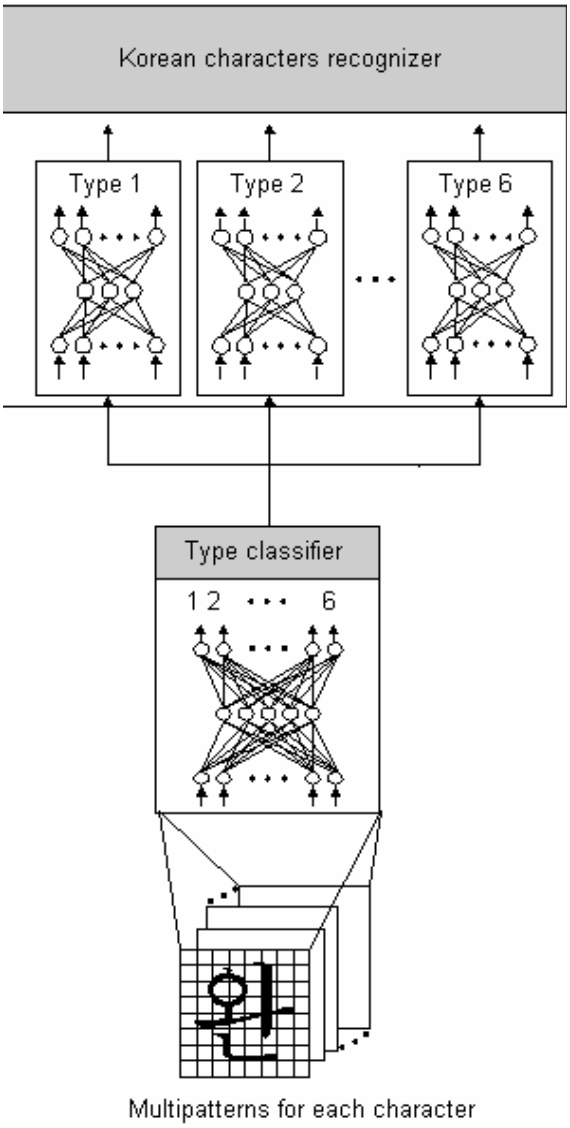


รูปที่ 2.13 รูปแบบตัวอักษรเกาหลีทั้ง 6 รูปแบบ

2.2.16 Multi-Modular Architecture Based on Convolutional Neural Network for Online Handwritten Character Recognition [30]

ในงานวิจัยนี้มีการนำเอาสถาปัตยกรรมของโครงข่ายประสาทเทียมมาใช้ในการรู้จำลายมือเขียนแบบออนไลน์ที่เป็นภาษาลาติน โดยที่สถาปัตยกรรมที่ใช้หลัก ๆ มีอยู่ 2 แบบที่ได้รับการพัฒนาและทำให้สมบูรณ์ แบบแรกคือ TDNN (Time Delay Neural Network) เป็นกระบวนการที่ทำงานเกี่ยวกับ Online Feature Extraction แบบที่สองคือ SDNN (Space Displacement Neural Network) ที่อาศัยการสร้างเส้นทางของลายมือเขียนจากปากกาแบบ Offline Bitmaps ขึ้นมาใหม่ นอกจากนั้นได้มีการรวมโครงสร้างสถาปัตยกรรมทั้งสองรวมกันเกิดเป็น SDTDNN (Space Displacement and Time Delay Neural Network) โดยสามารถที่จะใช้กลไกใหม่นี้มารู้จำได้ทั้งแบบออนไลน์และออฟไลน์และเมื่อเป็นเช่นนี้จึงทำให้ดูเหมือนว่าอัตราการรู้จำนั้นจะมีประสิทธิภาพสูงขึ้น ซึ่งโดยทั่วไปแล้วการรู้จำลายมือเขียนก็แบ่งออกเป็น 2 แบบคือ ออนไลน์และออฟไลน์ ทำให้รูปแบบของวิธีการได้มาซึ่งข้อมูล Input ก็จะแตกต่างกันไปตามวิธีการรู้จำ

จุดประสงค์หลักในการศึกษาครั้งนี้มีอยู่ 2 ข้อคือ จุดประสงค์ข้อแรกคือเพื่อให้สถาปัตยกรรมโครงข่ายประสาทเทียมตัวใหม่นี้มีความซับซ้อนน้อยกว่า Multi-Layer Perceptron (MLP) และในขณะเดียวกันสามารถนำมาใช้งานได้กับลายมือเขียนที่มีรูปแบบไม่เป็นมาตรฐานและอาจจะมีสัญญาณรบกวนได้มากกว่ารูปแบบการรู้จำทั่ว ๆ ไป เพื่อให้เหมาะสม งานวิจัยนี้จึงศึกษาพัฒนาและทดสอบของจากระบบ CNN (Convolutional Neural Network) จุดประสงค์ข้อที่สองคือเค้าโครงของระบบการรู้จำแบบออนไลน์ได้ศึกษาในทางตรงกันข้ามรวมเอาระหว่างการนำเสนอตัวอักษรแบบ Static และ Dynamic เนื่องมาจากการเก็บข้อมูลจากเส้นทางของปากกาในการเขียนลายมือเขียนแบบ Static นั้นจะได้ข้อมูลเส้นตัวอักษรที่ต่อเนื่องกัน ในขณะเดียวกันแบบ Dynamic ก็จะมีการเก็บ Pattern เดียวกันแต่จะมีการเคลื่อนที่แบบไม่ต่อเนื่องซึ่งจะมีช่วงระยะการเคลื่อนที่ที่ขึ้นกับความเร็วของการเคลื่อนที่ของปากกาแบบไม่แน่นอน ดังนั้นจะได้รูปแบบที่เป็น Pattern ที่ได้มาจากวิธีการทั้งสองวิธี ซึ่งข้อมูลทั้งสองวิธีที่ได้มาจะเป็นสิ่งที่ทำให้การรู้จำลายมือเขียนนั้นมีความถูกต้องมากยิ่งขึ้น



รูปที่ 2.14 กระบวนการรู้จำโดยอาศัยโครงข่ายประสาทเทียมสำหรับระบบการรู้จำตัวอักษรด้วยกล้องถ่ายรูปที่นำไปประยุกต์ใช้งานบน Mobile Device

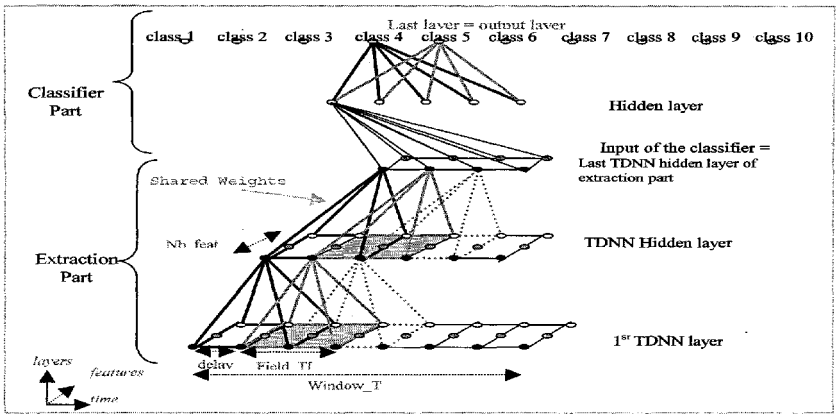
สำหรับโครงสร้างการรู้จำแบบ CNN นั้นมีพื้นฐานสถาปัตยกรรมมาจาก Multi-Layer Perceptron with Back-propagation Learning ที่เป็น Fully-Connection โดยที่โครงสร้างของ Input นั้นสามารถเป็นรูปแบบอะไรก็ได้ โดยจะไม่มีผลกระทบกับผลลัพธ์ที่ได้ออกมาของการ Training ส่วน CNN ในชั้นของ Hidden Neuron จะเชื่อมต่อกับ Subset of Neuron ที่มาจาก Preceding Layer ซึ่งในงานวิจัยนี้นำเอารูปแบบ 2 อย่างของ CNN คือ TDNN มาใช้กับข้อมูลแบบออนไลน์และ SDNN มาใช้จัดการข้อมูลแบบออฟไลน์ จากรูปที่ 2.15 เป็น TDNN Architecture เป็น Neural Network with Temporal Shift จะมีการพิจารณาสิ่งต่าง ๆ เช่น Size of the Receptive Fields, Number of Layers, Constraint on the Weight Sharing และ Learning Algorithm ซึ่งจากรูปประกอบ

ไปด้วย 2 ส่วนหลักคือส่วนล่าง Lower Layer เรียกว่า Extraction Part โดยจะแปลงลำดับของ Feature Vectors ให้เป็นลำดับของ Higher Order Feature Vectors ส่วนที่สองคือ ส่วนบน เรียกว่า Classifier Part เป็นโครงสร้าง MLP ที่รับเอา Input ที่เป็น Output ออกมาจาก Extraction Part

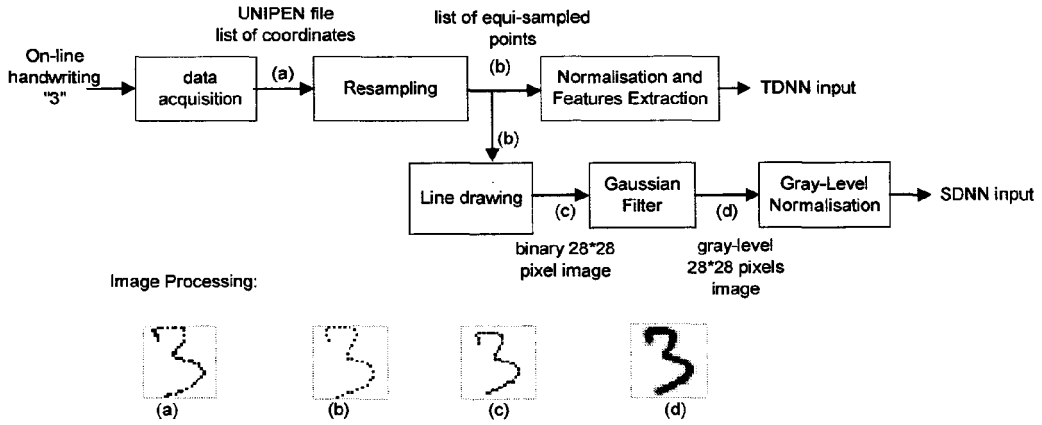
ข้อมูลที่ใช้ในการทดลองแบบออนไลน์มาจากฐานข้อมูลของ Unipen และ IRONOFF โดยที่ข้อมูลของ Input ได้มีการกำหนดรับเข้ามา 50 จุดคงที่สำหรับแต่ละลายมือเขียนที่รับมาจากความเร็วของการเคลื่อนที่ของปากกา จากนั้นผ่านเข้าสู่ Preprocessing ตามรูปที่ 2.16 เริ่มจากรับเอาลายมือเขียนแบบออนไลน์เข้ามา จากนั้นทำ Resampling จำนวน 50 จุดสำหรับแต่ละ Input แล้วแบ่งข้อมูลเป็น 2 ชุดเพื่อส่งต่อเข้าไป 2 วิธีต่อไปนี้ได้แก่วิธีที่ 1 ทำการ Normalized Features จากแต่ละจุดจะได้ค่า Position 2 ค่า, Direction 2 ค่า, Curvature 1 ค่า, Pen Status 1 ค่า แล้วส่งต่อเข้าไปให้เป็น Input ของ TDNN วิธีที่ 2 คือส่งเข้าสู่การทำ Line Drawing ได้รูปภาพแบบไบนารีขนาด 28x28 Pixels เข้าสู่ Gaussian Filter จะได้รูปภาพแบบ Gray-Level ขนาด 28x28 Pixels ออกมา จากนั้นจะทำ Gray-Level Normalization เพื่อส่งต่อเป็น SDNN Input ในขณะเดียวกันก่อนที่จะส่งข้อมูล Input ต่อเข้าไปยัง TDNN ก็มีการจัดแบ่งออกเป็น 3 ประเภทคือ Digits, Lowercase และ Uppercase ซึ่งจะช่วยให้ประสิทธิภาพในการรู้จำดีขึ้น 16% สำหรับ SDNN Input นั้นจะมีค่าระหว่าง -1 ถึง 1

ฐานข้อมูลที่ใช้เป็นข้อมูลในการทดลองการรู้จำแบบออฟไลน์สำหรับ SDNN คือฐานข้อมูลเดียวกันกับที่ใช้ใน TDNN แต่อยู่ในรูปแบบ Offline Image ผลการทดลองที่ได้ TDNN มีอัตราการรู้จำสูงกว่า MLP โดยเฉลี่ย 1-0.8% ในขณะเดียวกันก็มีการลดความซ้ำซ้อนของจำนวน Weights และขนาดของระบบนั้นเป็น Low Storage Capacities

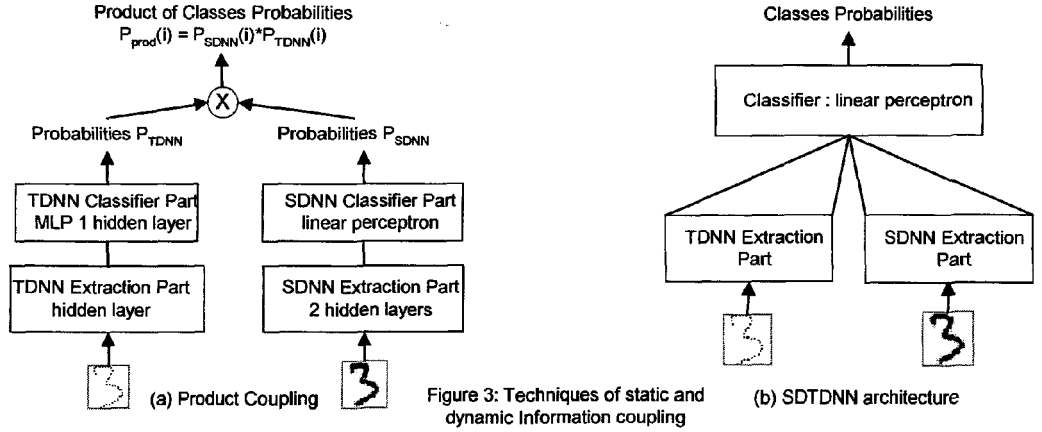
สำหรับผลการทดลองที่ได้ SDNN มีอัตราการรู้จำสูงกว่า MLP โดยเฉลี่ย 1-3% ในการศึกษาครั้งนี้ได้รวมวิธี TDNN เข้ากับ SDNN ตามรูปแบบสถาปัตยกรรมรูปที่ 2.17 โดยที่ข้อมูล Training TDNN จะแยกจากการ Training SDNN และจากการทดลองผลจากการรู้จำของ TDNN/SDNN เพิ่มขึ้นจาก 97.9% เป็น 98.2%



รูปที่ 2.15 สถาปัตยกรรม TDNN



รูปที่ 2.16 ขั้นตอนก่อนการรู้จำสำหรับใช้ในการรู้จำลายมือเขียนแบบออนไลน์ที่เป็นภาษาลาติน



รูปที่ 2.17 วิธีการของ Static และ Dynamic Information Coupling

2.3 ทฤษฎีโครงข่ายประสาทเทียมที่ใช้ในการรู้จำ

2.3.1 Fuzzy ARTMAP

เริ่มแรกของการรู้จำของ Neural Network มีการใช้อัลกอริทึมแบบ Back Propagation ซึ่งพบว่ามีปัญหาที่พบในเรื่องของระยะเวลาในการเรียนรู้และทดสอบนั้นใช้เวลานานรวมทั้งขีดความสามารถในการเรียนรู้ Pattern Input กล่าวคือในการ Trained จะมีการกำหนดจำนวน Neurons และค่าของWeight ไว้ไม่มีการเปลี่ยนแปลง จึงไม่สามารถที่จะเรียนรู้ Pattern ใหม่ที่ผ่านเข้ามาได้ และหากให้มีการเรียนรู้ Pattern รูปแบบใหม่ก็จะลืมการรู้จำ Pattern แบบเก่าไป ปัญหาตรงนี้เรียกว่า "Plasticity/stability dilemma"

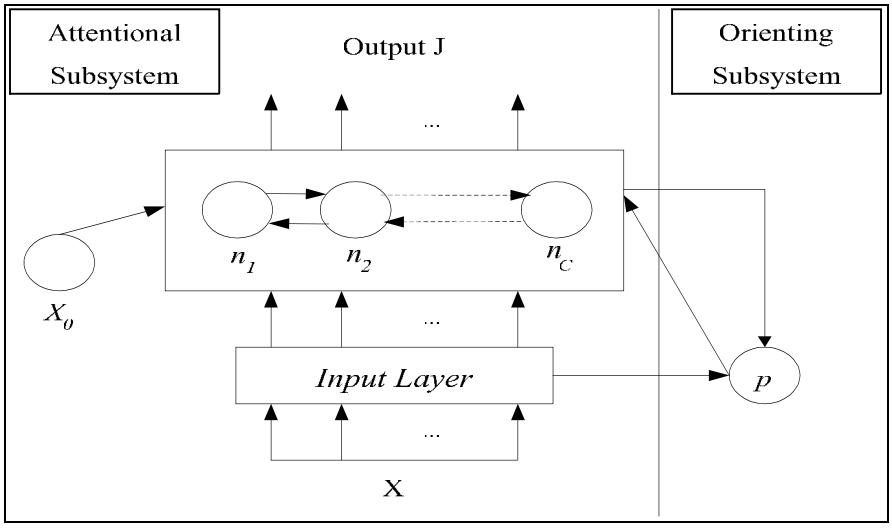
ในระยะเวลาต่อมาได้มีการคิดค้น ART เป็น Neural Network ชนิดหนึ่งที่ออกแบบโดย Grossberg ในปี ค.ศ. 1976 เพื่อนำมาใช้แก้ปัญหา "Plasticity/stability dilemma" เวอร์ชั่นแรกของ

ART คือ ART-1 คิดค้นโดย Carpenter และ Grossberg ในปี ค.ศ. 1987 ใช้งานในการจัดแบ่งข้อมูลไบนารี จากนั้นมาไม่นานก็มี ART-2 เกิดขึ้นมาโดยมีการขยายขีดความสามารถคือจัดการกับข้อมูลอนาล็อกได้ด้วย โดยทั้ง ART-1 และ ART-2 มีการรู้จำข้อมูลแบบไม่มีการเรียนรู้ (unsupervised clustering data) ต่อมามีการพัฒนามาเป็น ARTMAP ซึ่งจัดว่าเป็นอัลกอริทึมที่มีการเรียนรู้ใช้งานกับข้อมูลไบนารี หลังจากนั้นก็มี Fuzzy ARTMAP ที่สามารถพัฒนาให้ใช้งานกับข้อมูลอนาล็อกได้ โดยที่ทั้ง ARTMAP และ Fuzzy ARTMAP สามารถใช้วิธีในการรู้จำแบบมีการเรียนรู้ (Supervised Learning Classification)

ปัจจุบันนักวิจัยต่าง ๆ ได้ทำการพัฒนาเพิ่มประสิทธิภาพของรูปแบบการรู้จำของ Neural Network อาทิเช่น Adaptive Hamming Net (AHN), Gaussian ART (GA), Simplified Fuzzy ARTMAP (SFAM) และ Simplified ART (SFART)

2.3.1.1 ART-1

จัดว่าเป็นเวอร์ชั่นแรกของ ART-based Network ที่ได้รับการคิดค้นโดย Carpenter และ Grossberg โดยมุ่งไปที่การรู้จำข้อมูลไบนารีแบบไม่มีการเรียนรู้ (unsupervised clustering of binary data) โดยแบ่งออกเป็นสองส่วนหลักคือส่วนแรกเรียกว่า Attentional subsystem จะประกอบด้วย 1 Layer และส่วนที่สองเรียกว่า Orienting Subsystem ซึ่งใช้ในการตรวจสอบพิจารณาจากค่า Match Function เปรียบเทียบกับค่า Vigilance เพื่อหา Node ที่ชนะ โดยมีสถาปัตยกรรมของ ART ดังรูปที่ 2.18



รูปที่ 2.18 สถาปัตยกรรมของ ART

กำหนดให้ ค่า $T(X,W_j)$ คือ *Choice Function*

$$T(X,W_j)=\frac{\left\|X\cap W_j\right\|}{\alpha+\left\|W_j\right\|} \tag{2.3}$$

และค่า $M(X,W_j)$ คือ *Match Function*

$$M(X,W_j)=\frac{\left\|X\cap W_j\right\|}{\left\|X\right\|} \tag{2.4}$$

สมการในการเปรียบเทียบหา Match Criterion กับค่า Vigilance ρ

$$M(X,W_j)\geq\rho \tag{2.5}$$

จากสมการที่ 2.5 ถ้า Match Function มีค่ามากกว่าหรือเท่ากับค่า Vigilance แสดงว่าโหนดนั้นเป็น Winner Node จะมีการยอมให้ Input จัดเข้ากลุ่ม Node นั้น แต่หาก Match Function มีค่าน้อยกว่าค่า Vigilance จะไม่มีการจัดให้ Input เข้ากลุ่ม Node นั้น

$$a\cap b=\left(a_1ANDb_1,a_2ANDb_2,...,a_cANDb_c\right) \tag{2.6}$$

$$\left\|a\right\|=\sum_{i=1}^{D}a_i \tag{2.7}$$

2.3.1.2 Fuzzy ART

ต่อมา Carpenter และ Grossberg ได้พัฒนาให้ ART-1 สามารถทำงานกับชุดข้อมูลอินพุตที่เป็นข้อมูลอะนาล็อกได้ด้วย โดยอาศัยทฤษฎีของ Fuzzy Set และมีการใช้ Complement Coding กับชุดข้อมูลอินพุต กำหนดให้

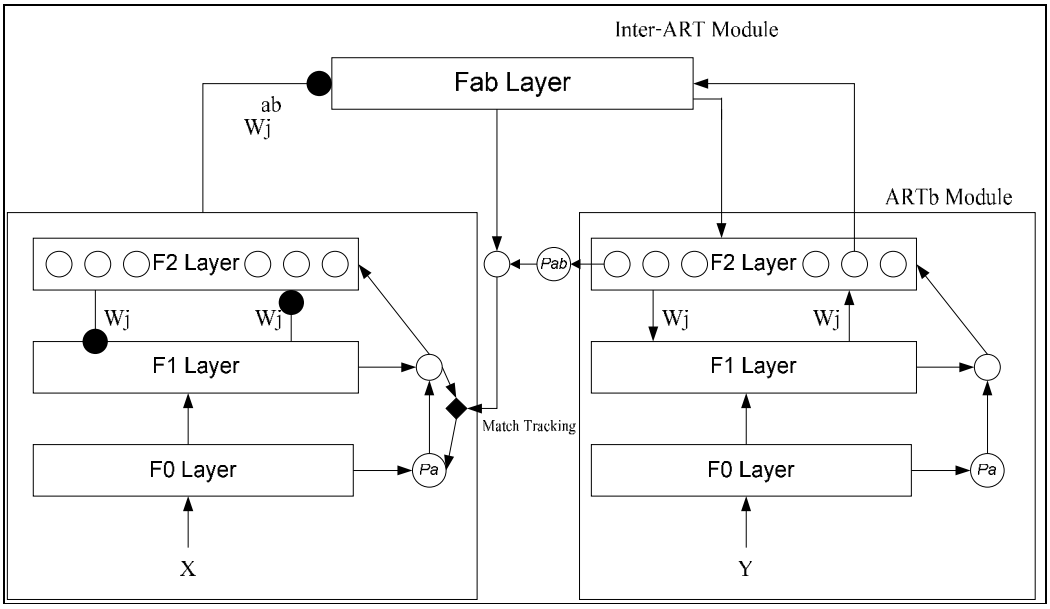
Fuzzy AND operation $\wedge: a\wedge b=\left(\min(a_1,b_1),\min(a_1,b_1),...,\min(a_D,b_D)\right)$ (2.8)

$\left\|.\right\|$ คือ $\left\|a\right\|=\sum_{i=1}^{D}a_i$ (2.9)

$$X=\left(X,X^C\right)=\left(x_1,x_1,...,x_D,1-x_1,1-x_2,...,1-x_D\right) \tag{2.10}$$

2.3.1.3 ARTMAP

เนื่องจาก ART และ Fuzzy ART เป็นวิธีการรู้จำแบบไม่มีการเรียนรู้ (Unsupervised Clustering Methods) จึงทำให้มีการพัฒนา ARTMAP ขึ้นมาเพื่อใช้สำหรับวิธีการรู้จำแบบมีการเรียนรู้ (Supervised Learning Classification) สถาปัตยกรรมของ ARTMAP ประกอบไปด้วย 2 ART คือ ART_a และ ART_b หลักการทำงานคือ Pattern Input จะถูกส่งเข้าไปยัง ART_a และ Label Input จะถูกส่งไปยัง ART_b ดังรูปที่ 2.19



รูปที่ 2.19 สถาปัตยกรรมของ ARTMAP, Fuzzy ARTMAP

2.3.1.4 Fuzzy ARTMAP

พัฒนาต่อเนื่องมาจาก ARTMAP เพื่อให้สามารถรับค่าข้อมูลอินพุตที่เป็นอนาล็อกเข้าไปได้ โดยที่ค่าที่ใช้ในการพิจารณาจะคล้ายกับแนววิธีพิจารณาเหมือน Fuzzy ART และมีโครงสร้างสถาปัตยกรรมเหมือนกับ ARTMAP ดังรูปที่ 2.19

หลักการทำงานของ Fuzzy ARTMAP มีขั้นตอนดังต่อไปนี้

<u>ขั้นตอนที่ 1</u>

Input Vector: ในแต่ละ Input Vector จะมีขนาด n-dimensional vector และมีการรวมเอา Complement Coding เข้าไปด้วย

$$x = (A, A^C) = \{A_1\, A_2\, A_3\, \ldots A_n\, 1\text{-}\, A_1\, 1\text{-}\, A_2\, 1\text{-}\, A_3 \ldots 1\text{-}\, A_n\} \qquad (2.11)$$

$$|x| = \left|\left(A, A^C\right)\right| = \sum_{i=1}^{n} A_i + \left(n - \sum_{i=1}^{n} A_i\right) = \text{n} \tag{2.12}$$

กำหนดให้ x แทน Input Vector ของกระบวนการรู้จำ ART_a

$A = A_1 \ldots A_n$ แทน ค่า Input Pattern

$A^C = 1\text{-}A_1\ 1\text{-}A_2\ 1\text{-}A_3 \ldots 1\text{-}A_n$ แทน ค่า Complement Coding ของค่า Input Pattern

∴ ขนาดของ 1 Vector Input สำหรับ 1 Pattern (ภาพตัวอักษร) = 2n

<u>ขั้นตอนที่ 2</u>

Weight Vector: ในแต่ละ Category จะมีการกำหนดค่า Weight Vector

$$W_1 = W_2 = W_3 = \ldots = W_n = 1 \tag{2.13}$$

กำหนดให้ $W_1\ W_2\ W_3 \ldots W_n$ แทน Weight Vector ของชั้น F_{12} Layer

<u>ขั้นตอนที่ 3</u>

Parameters:

- A Choice Parameter (α) จะมีค่าเข้าใกล้ 0 ∴ กำหนดให้ $\alpha = 0$
- A Learning Parameter (β) จะมีค่าได้ 3 ค่าตามความเร็วของการเรียนรู้ดังนี้
  - ○ β=1 เรียกว่า Fast Learning ส่งผลให้ค่าของ Weight Vector มีการลดลงเปลี่ยนแปลงไปอย่างรวดเร็วขึ้นอยู่กับ Input และ Weight Vector
  - ○ 0< β <1 แสดงว่ามีการเรียนรู้แบบค่อยเป็นค่อยไป ส่งผลให้ค่าของ Weight Vector มีการลดลงอย่างเรื่อย ๆ เป็นสัดส่วนกับค่า β ที่เลือกใช้
  - ○ β=0 แสดงว่าไม่มีการเรียนรู้เลย ส่งผลให้ค่าของ Weight Vector ไม่มีการเปลี่ยนแปลงตลอดการเรียนรู้

สำหรับการเลือก Neuron ที่ชนะของ ART ได้มาจากการคำนวณค่า Choice Function เพื่อวัดค่าความคล้ายคลึงกันระหว่าง Input และ Weight Vector หาได้จากสมการต่อไปนี้

$$T_j(x) = \frac{|x \wedge W|}{|W|} \tag{2.14}$$

$$Winner = T_J = \max\{T_j(x)\} \tag{2.15}$$

กำหนดให้ $T_j(x)$ แทนค่าความคล้ายคลึงกันระหว่าง Input (x) และ Weight (W) จะมีค่าอยู่ตั้งแต่ 1 ถึง 0

$|W|$ คือค่า Norm ของ Weight Vector หาได้จาก $|W| = \sum_{i=1}^{n} W_i$

T_J คือค่าที่ใช้ในการเลือก Neuron ที่เหมาะสมและชนะของ ART ซึ่งถ้ามีค่ามากกว่า 1 Neuron ที่ชนะจะเลือกโหนด Neuron ที่มีค่า j น้อยที่สุด

x แทน Input Vector ของกระบวนการรู้จำ ART_a

j คือ Neuron แต่ละตัวที่ Layer นั้น

n แทนจำนวนของ Element ใน Weight Vector

ขั้นตอนที่ 4

- A Vigilance Parameter (ρ) คือค่าความคล้ายคลึงระหว่าง Input และ Weight Vector ว่ามีความคล้ายคลึงกันมากพอที่จะจัดให้ Input นั้นไปรวมอยู่ในกลุ่ม Category นั้นหรือไม่ โดยจะกำหนดค่า = 1 เมื่อเหมาะสมที่มีความคล้ายคลึงกันมากที่สุดที่จะจัดไว้ในกลุ่มเดียวกัน มีการตรวจสอบตามสมการต่อไปนี้

$$\frac{|x \wedge W_J|}{|x|} \geq \rho \tag{2.16}$$

จากนั้นตรวจสอบ Vigilance Criterion ของส่วน Map Field ตามสมการดังต่อไปนี้

$$\frac{|y_{ART_b} \wedge W^J_{ART_b}|}{|y_{ART_b}|} \geq \rho_{ab} \tag{2.17}$$

และถ้าเป็นสมการเป็นเท็จ จะทำการปรับค่า Vigilance ให้ $\rho > \frac{|x \wedge W_J|}{|x|}$

ขั้นตอนที่ 5

ในขณะที่มีการเรียนรู้นั้นจะคำนวณหาค่าจาก Category Choice Function ระหว่าง Input กับทุก ๆ Weight ของ Neurons จากนั้นจะทำการนำค่าที่ได้ที่มีค่ามากที่สุดมาทำการเปรียบเทียบกับค่า Vigilance Parameter (ρ) ซึ่งจะเกิดเหตุการณ์ที่เป็นไปได้ 2 กรณีคือ

1. Resonance

"If match function of the chosen node meets (equal or more than) the vigilance criterion"

$$f(x,W) \geq \rho \tag{2.18}$$

หากพบว่ามี Neuron ที่มีค่าที่ได้ออกมาจาก Category Choice Function มีค่ามากกว่าค่า Vigilance อยู่มากกว่า 1 Neuron จะทำการเลือก Neuron ลำดับแรกสุดที่ชนะ โดยทำการตรวจสอบ Vigilance Criterion

2. Reset

"If match function of the chosen node meets (less than) the vigilance criterion"

$$f(x,W) < \rho \tag{2.19}$$

จากนั้นจะมีการสร้าง Neuron ขึ้นมาใหม่สำหรับ Input Node ใหม่ที่รับเข้าสู่กระบวนการรู้จำและกำหนดค่า Weight ขึ้นมาเริ่มต้นตามขั้นตอนที่ 2

ขั้นตอนที่ 6

Learning: ส่วนของการเรียนรู้นี้จะเกิดขึ้นเมื่อขั้นตอนที่ 4 มีการเกิดกรณี Resonance หลังจากนั้นจะทำการปรับค่าของ Weight Vector สำหรับ Neuron ที่ชนะ (Winner) และยอมให้จัดเอา Input ที่เข้ามารวมไว้ในกลุ่มเดียวกันได้ ดังสมการต่อไปนี้

$$W^{new} = \beta(x \wedge W^{old}) + (1-\beta)W^{old} \tag{2.20}$$

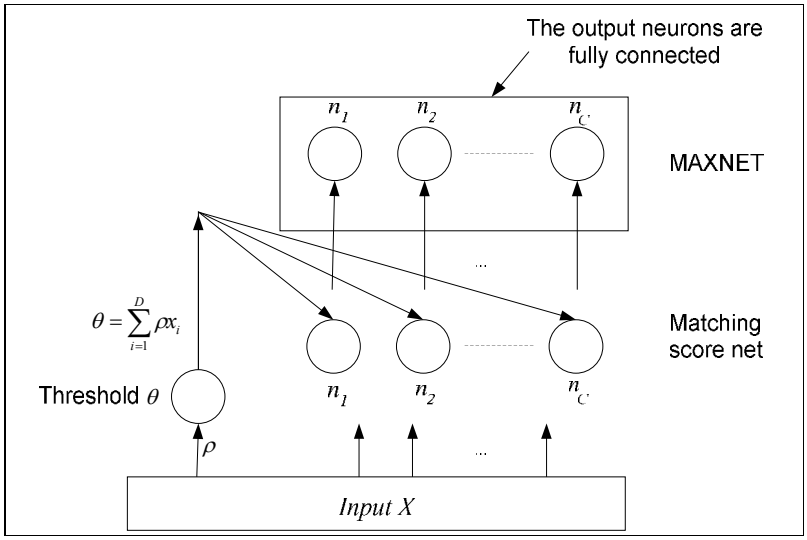
2.3.1.5 Adaptive Hamming Net (AHN)

พัฒนามาจาก ART Models ออกแบบโดย Cheng-An Hung และ Sheng-Fuu สถาปัตยกรรมของ AHN แสดงดังรูปที่ 2.20 ประกอบไปด้วย 2 เลเยอร์ จำนวนของ Neurons ของ Hidden Layer จะเท่ากับจำนวน Neurons ของ Output Layer ซึ่ง ทุก Input จะเชื่อมต่อเข้ากับทุก Neurons ใน Hidden Layer แต่สำหรับในแต่ละ Neuron ที่ j ใน Hidden Layer จะเชื่อมต่อไปที่ Neuron ที่ j ใน Output Layer เท่านั้น ซึ่งใน Hidden Layer จะพิจารณาค่า Matching Score Net ว่า Input Pattern มีความเหมือนพอที่จะจัดเข้ากลุ่ม Neuron นั้น ๆ ได้หรือไม่ โดยที่อาจจะมี Neurons ที่คล้ายกับ Input เกิด Activated มากกว่า 1 Neuron ซึ่งหลังจากนั้นใน Output Layer จะใช้การหา

MAXNET ในการคัดเลือกว่า Neuron ไหนจะชนะซึ่งจะเป็น Neuron ที่มีค่า Matching Score Net มากที่สุด

2.3.1.6 Gaussian ARTMAP

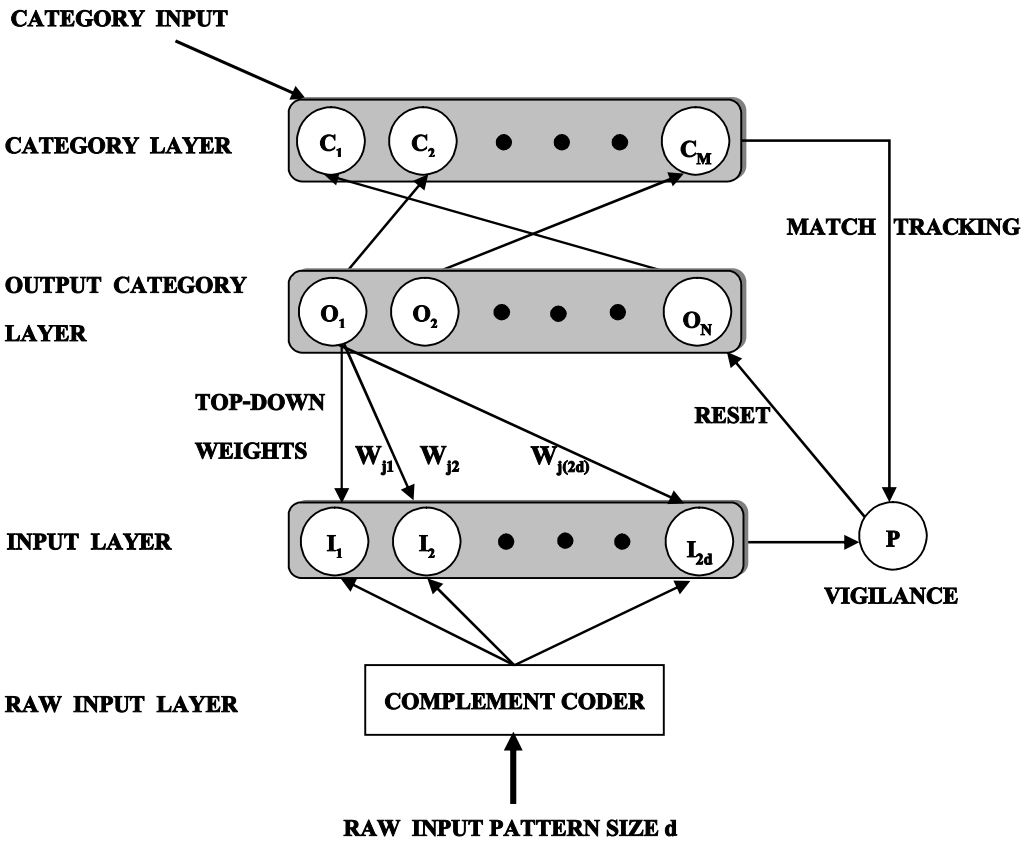
Williamson ได้คิดวิธีในการแก้ปัญหาที่พบใน Fuzzy ARTMAP 2 อย่างคือ ข้อมูลสัญญาณรบกวนที่เกิดขึ้นในการรู้จำ และการแบ่งกลุ่มนำเสนอแต่ละประเภทอย่างไม่มีประสิทธิภาพ ด้วยปัญหาสองอย่างนี้ จึงนำเอาการกระจายแบบ Gaussian เข้ามาใช้ในพิจารณา Choice Function, Match Function, และ Classification Function ข้อดีคือการเพิ่มค่าของ Choice Function และ Matching Function จะมีค่าที่สามารถแบ่งกลุ่มให้รูปแบบของ Input เข้าจัดกลุ่มได้ตรงตามแต่ละประเภทได้อย่างเหมาะสม โดยไม่มีผลมาจากสัญญาณรบกวนที่อาจจะเกิดขึ้นและทำให้แบ่งกลุ่มผิด นอกจากนี้การกระจายแบบ Gaussian เป็นรูปแบบของการกระจายที่มีการนำมาใช้งานส่วนใหญ่ แต่สำหรับขอบเขตของ GA ก็คือการใช้งานจะมีค่าการกระจายแบบ Gaussian อย่างอิสระดังนั้นจึงจำเป็นต้องมี Diagonal Ellipsoids ครอบคลุมช่วงของข้อมูลอินพุต ข้อดีที่เห็นได้ชัดว่ามีประสิทธิภาพมากกว่า Fuzzy ARTMAP ก็คือการรู้จำแบ่งกลุ่มได้มีจำนวน Nodes ที่รู้จำได้ถูกต้องมากขึ้นตรงตามค่า Output ที่ต้องการและทำการกำหนดไว้



รูปที่ 2.20 สถาปัตยกรรมของ AHN

2.3.2 Simplified Fuzzy ARTMAP (SFAM)

เป็น ART ชนิดหนึ่งที่มีการปรับวิธีในการรู้จำให้ง่ายขึ้นจาก Fuzzy ARTMAP โดยจะมีการลดส่วนของ ART_b ออกไป คงเหลือไว้เพียงแต่ ART_a ที่มีลักษณะการทำงานคล้ายกับใน Fuzzy ARTMAP ในการเข้าใจการทำงานโครงสร้างภายในเป็นไปตามฟัซซี่ลอจิก (Fuzzy Logic) สถาปัตยกรรมของ SFAM ดังรูปที่ 2.21 โครงสร้างของ SFAM จะคล้ายกับการเชื่อมต่อ Weight, Input Layer และ Output Layer นอกจากนี้จะมี Category Layer ที่มีการรับเอา Category Input เข้ามาไว้ ลักษณะพิเศษนี้ทำให้ SFAM สามารถแก้ปัญหากรณีที่ข้อมูลอินพุตที่เข้ามาเป็น Nonlinear Decision Region คือการจัดประเภทกลุ่มชุดข้อมูลโดยไม่มีรูปแบบที่แน่นอน ทั้งนี้จะมีการทำ Complement Coding ให้กับ Input Vector ก่อนที่จะเข้าสู่ SFAM เนื่องจากเป็นการทำการ Normalize และเพิ่มขยายข้อมูล Input ให้มีขนาดเป็นสองเท่าของชุดข้อมูลเดิม เพื่อช่วยให้ SFAM ได้ทำการตัดสินใจได้ดีขึ้น



รูปที่ 2.21 สถาปัตยกรรมของ Simplified Fuzzy ARTMAP (SFAM)

2.4 ทฤษฎีเบื้องต้นของงานวิจัยนี้

2.4.1 ครอสคอร์รีเลชัน (Cross-Correlation)

ในการจัดแบ่งกลุ่มของรูปแบบต่าง ๆ ออกเป็นกลุ่มเดียวกัน จะอาศัยการวัดความคล้ายคลึงกันระหว่างรูปแบบของชุดข้อมูลต่าง ๆ เหล่านั้น ครอสคอร์รีเลชันคือวิธีมาตรฐานอีกวิธีหนึ่งที่ใช้ในการวัดและประเมินความคล้ายคลึงกันว่าชุดข้อมูลต่าง ๆ มีความสัมพันธ์กันมากน้อยเพียงใด ซึ่งถ้ากล่าวถึงค่าคอร์รีเลชันก็คือค่าครอสคอร์รีเลชันที่มีค่า Delay = 0 โดยค่าความสัมพันธ์ที่ได้เมื่อทำการ Normalize ค่าครอสคอร์รีเลชันแล้ว จะได้ค่าความสัมพันธ์อยู่ระหว่าง 0 ถึง 1 พิจารณาจากสมการต่อไปนี้

$$r(x,y) = \sum_i x_i y_i \tag{2.21}$$

$$r_p(x,y) = \frac{\sum_i x_i y_{i-p}}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}} \tag{2.22}$$

$$R(x,y) = \max_p \left(r_p(x,y) \right) \tag{2.23}$$

กำหนดให้ $r(x,y)$ แทนค่า Correlation (at elay p = 0) without normalized

$r_p(x,y)$ แทนค่า Normalized Cross-correlation at delay = p

$R(x,y)$ แทนค่า Normalized Cross-correlation function ระหว่างสองชุดข้อมูล x, y

x_i แทนชุดข้อมูลรูปแบบ x_i

y_i แทนชุดข้อมูลรูปแบบ y_i

i = 0, 1, 2, 3, …, N-1

N คือจำนวนของ element ที่อยู่ใน Pattern ชุดข้อมูล x_i และ y_i

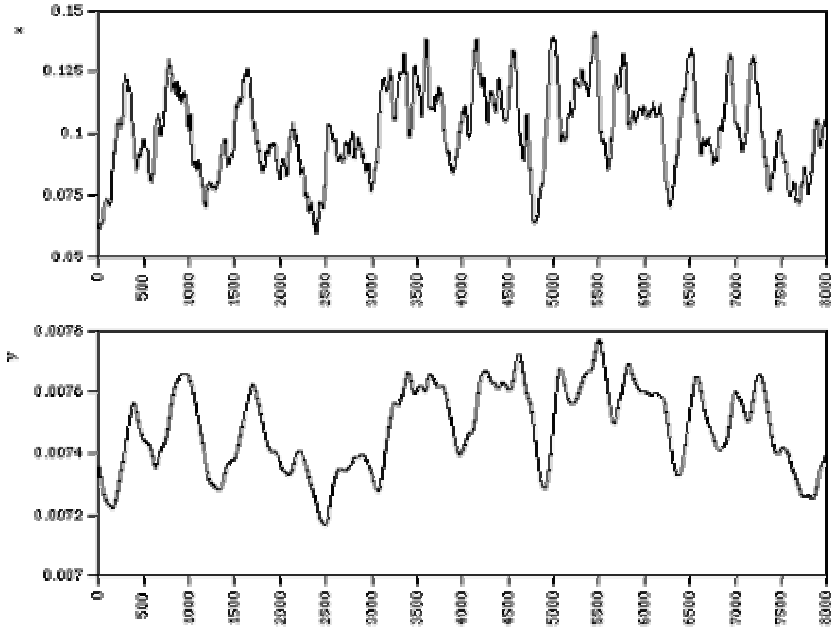
p คือค่า delay ที่มีการเลื่อนในแต่ละตำแหน่งของการหาค่าความสัมพันธ์ออกมา โดย p = 0, 1, 2, 3, …, N-1

max คือการคำนวณหาค่ามากที่สุดที่ได้

ในการประยุกต์ใช้งานในการจัดกลุ่มของวัตถุที่เป็นสัญญาณคลื่นอนาล็อกหรือเรียกว่ารูปคลื่นแบบต่อเนื่องทางเวลา (Continuous-time Waveform) ดังรูปที่ 2.22 หรืออาจจะเป็นชุดข้อมูลแบบลำดับไม่ต่อเนื่อง (Discrete-time Sequence) หรือชุดข้อมูลแบบเวคเตอร์ดังรูปที่ 2.23

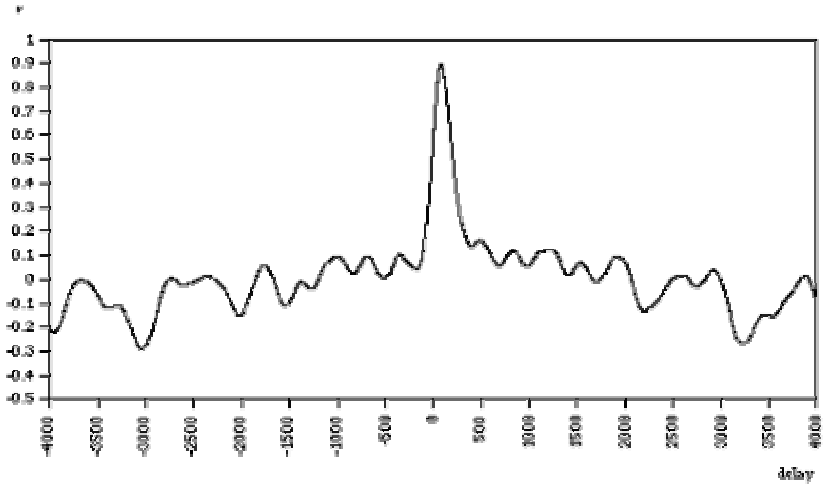
สำหรับการพิจารณารูปคลื่นแบบต่อเนื่องทางเวลาจะคำนวณตามสมการ ซึ่งเป็นการหาค่าความคล้ายคลึงกันของคลื่นตามช่วงเวลา หรือเรียกว่าเป็นการหาค่าครอสคอร์รีเลชันที่ค่า Delay ต่าง ๆ และสำหรับการหาค่าคอร์รีเลชันก็คือการหาค่าครอสคอร์รีเลชันที่ค่า delay = 0 จากรูปที่ 2.10 มีการทำ Normalized โดยค่าครอสคอร์รีเลชันจะมีค่าอยู่ระหว่าง -1, 1โดยที่ถ้า r = 1 คือมีความคล้ายคลึงกันมาก แต่ถ้า r = 0 คือไม่มีความคล้ายคลึงกัน และถ้า r = -1 คือมีความแตกต่างกันระหว่างชุดข้อมูลทั้งสอง

จากรูปที่ 2.23 ยังไม่มีการทำ Normalized ดังนั้นค่าที่ได้ก็จะมีค่าสูงไปด้วย ตามขนาดของเวคเตอร์ และพบว่าสมาชิกของเวคเตอร์ที่มีค่าเป็น 0 นั้นจะเข้ามารบกวน ส่งผลกระทบต่อความคล้ายคลึงกัน แต่การวัดค่าด้วยครอสคอร์รีเลชันจะได้ผลออกมาเป็น 1 แสดงว่าชุดข้อมูลสองเวคเตอร์มีความเหมือนกัน แต่หากพิจารณาค่าคอร์รีเลชันจะพบว่ามีค่าเป็น 3/5 = 0.6 เมื่อทำการ Normalized ซึ่งจะเห็นว่าการพิจารณาเฉพาะค่าคอร์รีเลชันจะได้ค่าความคล้ายกันออกมา ณ ค่าของ delay = 0 ซึ่งจะเห็นว่าการใช้การคำนวณค่าครอสคอร์รีเลชันจะได้ผลลัพธ์ของการแสดงค่าความเหมือนหรือคล้ายคลึงกันได้ออกมาอย่างถูกต้องและเหมาะสม



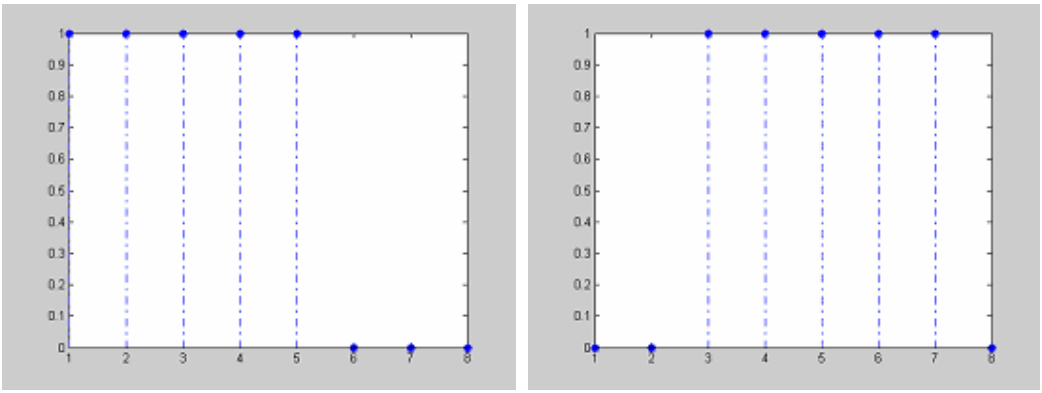
(ก)

รูปที่ 2.22 (ก) รูปคลื่น x, y แบบต่อเนื่องทางเวลา (ข) ค่าครอสคอร์รีเลชันที่ได้ประมาณ 0.9 แสดงว่ารูปคลื่น x, y มีความสัมพันธ์คล้ายคลึงกันมากที่ delay = 40 จากทั้งหมด 4000 delay

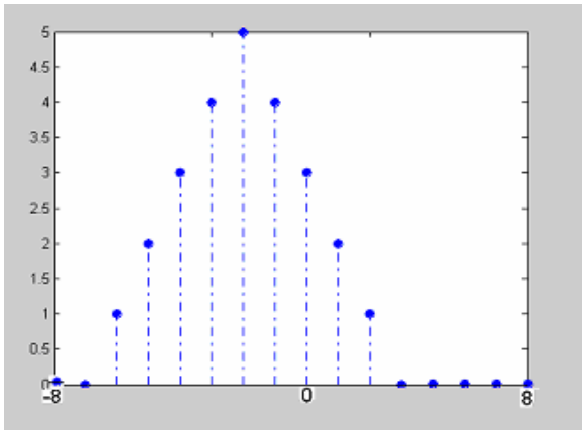


(ข)

รูปที่ 2.22 (ต่อ)



(x) (y)



(r)

รูปที่ 2.23 (x) ชุดข้อมูล x และ (y) ชุดข้อมูล y (r) ค่าครอสคอร์รีเลชัน (เมื่อทำการ normalized) ที่ได้ประมาณ 1 (r =1) แสดงว่าชุดข้อมูล x, y มีความสัมพันธ์เหมือนกันที่ delay = -2