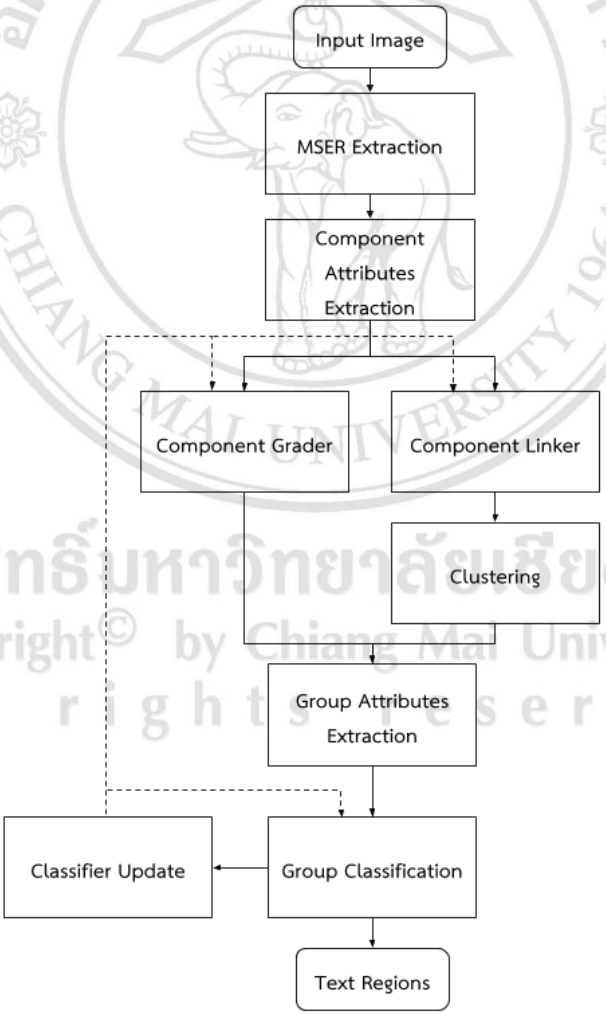
บทที่ 3

แนวคิดการแก้ไขปัญหาและขั้นตอนในการแก้ปัญหา

ดังที่ได้กล่าวมาข้างต้นแล้วว่าในการสร้างแบบจำลองในการตรวจจับตำแหน่งของข้อความบนภาพนั้น จะต้องใช้ปริมาณชุดฝึกสอนเป็นจำนวนมากโดยมนุษย์ ซึ่งต้องใช้เวลาและแรงงานในการเตรียมข้อมูลฝึกสอนเหล่านั้น จึงเป็นที่มาของแนวคิดในวิทยานิพนธ์นี้ที่ต้องการพัฒนาระบบระบุตำแหน่งข้อความที่สามารถเรียนรู้ และปรับปรุงการทำงานด้วยตัวเองได้ เพื่อลดระยะเวลาในการเตรียมชุดข้อมูลฝึกสอนโดยมนุษย์



รูปที่ 3.1 แผนผังการแก้ปัญหา

ซึ่งขั้นตอนการระบุตำแหน่งของข้อความที่พัฒนาขึ้นในงานวิจัยนี้นั้น ได้นำงานวิจัยของ Wiwatcharakoses และคณะ [8] ซึ่งเป็นวิธีที่ใช้ทั้งวิธีการเชิงพื้นที่ร่วมกับวิธีการเชิงพื้นผิวมาเป็นแนวทางในการพัฒนาและปรับปรุงการทำงาน โดยในงานนี้ได้ปรับปรุงในส่วนของการสกัดพื้นที่ MSER รวมถึงเพิ่มการทำงานในส่วนเรียนรู้และปรับปรุงการทำงานด้วยตัวเองอีกด้วย ซึ่งแผนภาพการทำงานจะเป็นไปดังรูปที่ 3.1

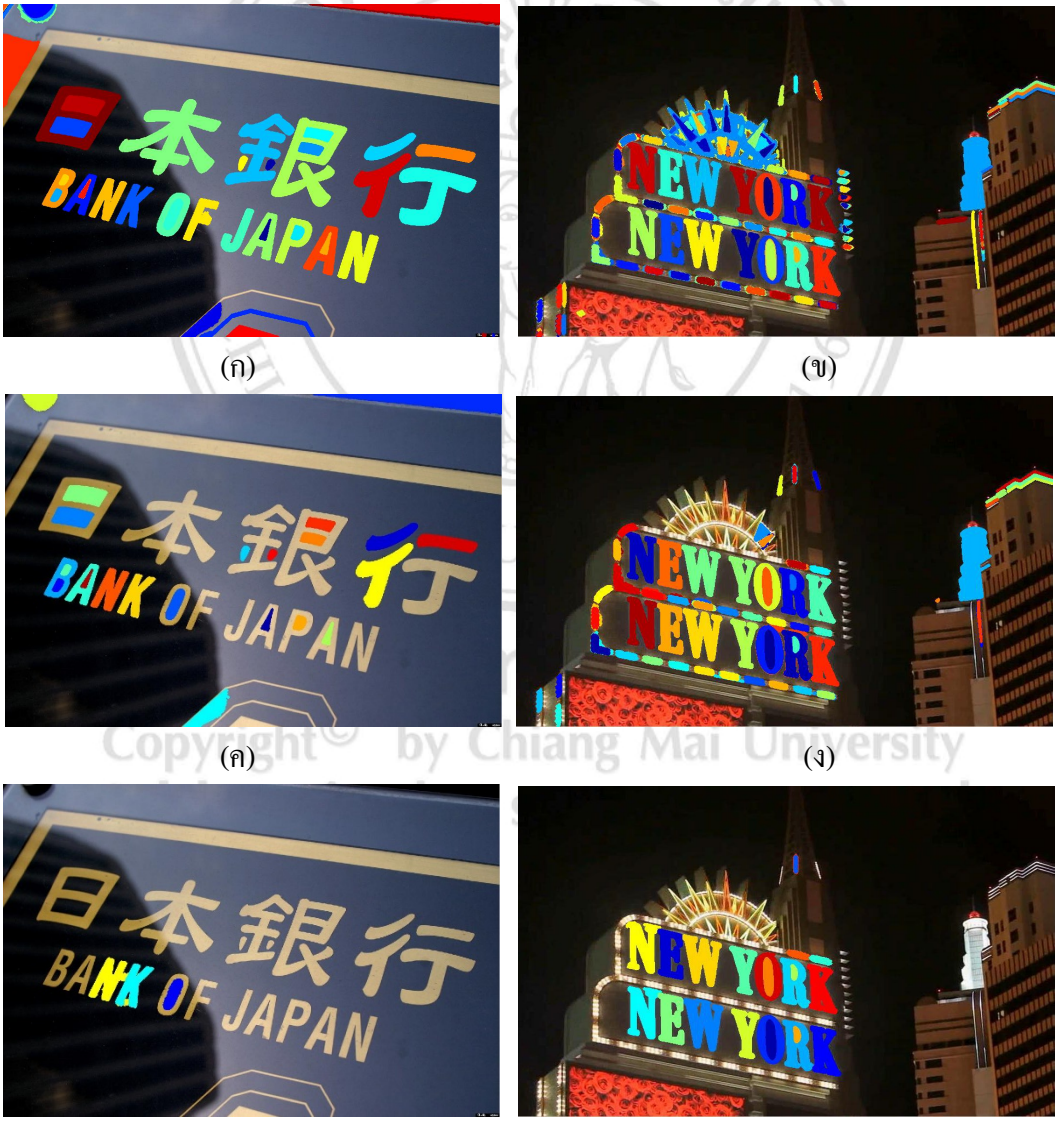
โดยในการทำงานเริ่มต้นจะสกัดหาพื้นที่คาดว่าจะเป็นข้อความโดยใช้ MSER ซึ่งจะได้ออกมาเป็นองค์ประกอบย่อย ๆ จากนั้นจึงจะสกัดคุณลักษณะของแต่ละองค์ประกอบเพื่อใช้ในการรวมกลุ่มและจำแนกองค์ประกอบว่าเป็นตัวอักษรหรือไม่ โดยคุณลักษณะที่สกัดออกมาจะเป็นข้อมูลเชิงขนาด, เชิงสี, เชิงตำแหน่ง และเชิงความกว้างของเส้น จากนั้นจะทำการรวมกลุ่มโดยส่วนให้คะแนนความสัมพันธ์ของคู่องค์ประกอบตัวเลือก (Component Linker) เพื่อหาความสัมพันธ์ของแต่ละคู่องค์ประกอบเพื่อเชื่อมองค์ประกอบที่คล้ายกันเข้าด้วยกัน โดยจะได้ออกมาเป็นกลุ่มของตัวอักษร และนอกจากนั้นองค์ประกอบแต่ละตัวจะนำเข้าไปในส่วนการให้คะแนนองค์ประกอบตัวเลือก (Component Grader) เพื่อวิเคราะห์ว่าองค์ประกอบตัวนั้น ๆ เป็นตัวอักษรหรือไม่ จากนั้นจึงจะนำสิ่งที่ได้จากทั้งส่วนเข้าไปในส่วนจำแนกกลุ่มของข้อความ (Group Classification) เพื่อหากลุ่มของตัวอักษรที่เป็นข้อความ และสุดท้ายจะเป็นในส่วนเรียนรู้เพิ่มเติมของตัวคัดกรอง (Classifier Update) โดยระบบจะเลือกตัวแทนข้อมูลจากคำตอบเพื่อนำไปฝึกสอนกับส่วนจำแนกทั้งสามคือ Component Grader, Component Linker และ Group Classification เพื่อเพิ่มความสามารถให้ระบบต่อไป โดยรายละเอียดของกระบวนการต่าง ๆ จะนำไปอธิบายในส่วนหัวข้อต่อไป

3.1. กระบวนการสกัดองค์ประกอบด้วย MSER (MSER Extraction)

เริ่มต้นระบบจะทำการสกัดพื้นที่ที่มีโอกาสจะเป็นข้อความ โดยใช้ MSER ในการทำงานเพื่อให้ได้พื้นที่ขนาดย่อย ๆ หรือก็คือองค์ประกอบตัวเลือก ซึ่งโดยปกติแล้วพารามิเตอร์ในการสกัด MSER ในทีนี้จะกล่าวถึงระดับการเพิ่มของค่าเกณฑ์ (Threshold Delta) และค่าความแตกต่างสูงสุดระหว่างพื้นที่ (Max Area Variation) ซึ่งหากกำหนดให้มีเงื่อนไขเข้มงวด พื้นที่ที่สกัดได้นั้นมีความเป็นไปได้ที่จะเป็นตัวอักษรสูง แต่ปริมาณพื้นที่ที่ได้จะมีน้อยคือจะไม่ได้ตัวอักษรทั้งหมดที่อยู่ในรูปภาพได้ และหากกำหนดให้มีเงื่อนไขที่ผ่อนผันเกินไปปริมาณพื้นที่ที่ได้จะมีมากเกินไปทำให้การทำงานของระบบอาจจะมีผลความแม่นยำต่ำ และใช้ระยะเวลาคำนวณสูง

(ก) (ข)

รูปที่ 3.2 (ก) และ (ข) ตัวอย่างรูปตั้งต้น



(ก) (ข)

(ค) (ง)

(จ) (ฉ)

รูปที่ 3.3 ตัวอย่างการใช้พารามิเตอร์ในการสกัด MSER แตกต่างกัน

จากรูปที่ 3.2 เป็นรูปภาพตั้งต้น และรูปที่ 3.3 จะแสดงตัวอย่างการสกัดหาพื้นที่ขีดสุดด้วยวิธีการ MSER และแต่ละพื้นที่ขององค์ประกอบตัวเลือกจะแสดงด้วยสีที่แตกต่างกันไป ซึ่งพารามิเตอร์ที่ใช้ในการสกัดจะแตกต่างกันโดยจะแสดงใน 3 รูปแบบ ซึ่งกำหนดให้ TD คือระดับการเพิ่มของค่าเกณฑ์ (Threshold Delta) และ MVA คือค่าความแตกต่างสูงสุดระหว่างพื้นที่ (Max Area Variation) ดังนี้

1) รูปที่ 3.3 (ก) และ (ข) ค่าพารามิเตอร์ที่ใช้คือ $TD = 18$ และ $MVA = 1.4$
2) รูปที่ 3.3 (ค) และ (ง) ค่าพารามิเตอร์ที่ใช้คือ $TD = 30$ และ $MVA = 1.3$
3) รูปที่ 3.3 (จ) และ (ฉ) ค่าพารามิเตอร์ที่ใช้คือ $TD = 45$ และ $MVA = 1.2$

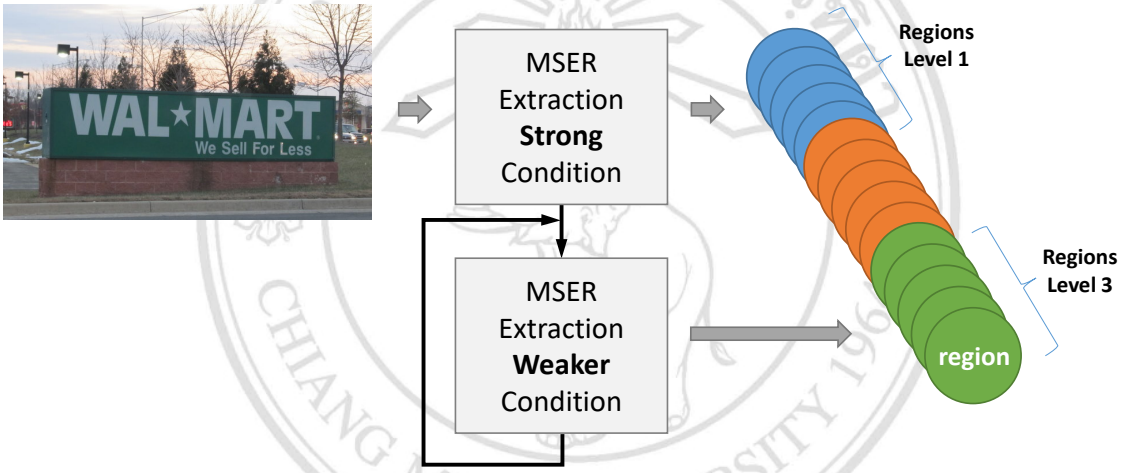
จะเห็นได้ว่าในรูปแบบที่ 1 องค์ประกอบตัวเลือกที่ออกมาในรูปตั้งต้น 3.2 (ก) มีปริมาณที่เหมาะสมและครอบคลุมตัวอักษรทั้งหมดที่อยู่ในรูป โดยเมื่อเทียบกับรูปตั้งต้น 3.2 (ข) ในรูปแบบเดียวกันนั้นองค์ประกอบตัวเลือกที่ได้มามีปริมาณมากเกินไป และได้ส่วนที่ไม่ใช่ตัวอักษรมาด้วย ซึ่งค่าพารามิเตอร์ที่เหมาะสมกับรูปตั้งต้น 3.2 (ข) นั้นควรจะใช้รูปแบบที่ 3 จึงจะเหมาะสมกว่า จากกรณีนี้จะเห็นได้ว่ารูปแต่ละรูปที่แตกต่างกันนั้น ในการสกัดหาพื้นที่ค่าพาริมิเตอร์ที่เหมาะสมในการสกัดจึงจะแตกต่างกัน

ดังนั้นทางผู้จัดทำจึงเสนอวิธีการค้นหาพื้นที่แบบใหม่คือ “การสกัด MSER แบบหลายระดับความเข้มงวด (MSER Extraction with Multiple-threshold scheme)” โดยเป็นไปตามอัลกอริทึมและตัวอย่างการทำงานดังรูปที่ 3.4 และ 3.5 ตามลำดับ ซึ่งขั้นตอนการทำงานจะเริ่มต้นโดยสกัดหาพื้นที่จาก MSER โดยใช้เงื่อนไขเข้มงวดซึ่งจะถือว่าเป็นพื้นที่ที่มั่นใจว่าเป็นตัวอักษร จากนั้นจะสกัดหาพื้นที่ที่มีความมั่นใจน้อยตามระดับความเข้มงวดโดยใช้เงื่อนไขที่ผ่อนผันลงมา ทำให้ได้พื้นที่ที่คาดว่าจะเป็นตัวอักษรที่มีระดับความเข้มงวดที่ใช้หลากหลายระดับ จากนั้นจะทำการคัดพื้นที่ที่ทับซ้อนกันออกไปเนื่องจากยังมีพื้นที่บางส่วนที่มีรูปร่างคล้ายกันและทับซ้อนกันอยู่ และจำนวนองค์ประกอบตัวเลือกที่มากเกินไปก็เป็นปัจจัยหนึ่งในการเสียเวลาประมวลผลในขั้นตอนต่อ ๆ ไป ซึ่งในการกำจัดพื้นที่ทับซ้อนนั้นจะใช้ขนาดเป็นปัจจัยหลัก โดยเริ่มต้นจะทำการสร้างมาร์ก (Mask) หรือรูปลักษณะของพื้นที่นั้น ๆ และนำมาเทียบกับมาร์กหลัก โดยหากมีการซ้อนทับระหว่างพื้นที่ พื้นที่ที่มีขนาดเล็กกว่าจะถูกคัดทิ้ง แต่หากไม่ทับซ้อนกันก็จะนำพื้นที่นั้น ๆ ไปทำงานต่อไป และทำการปรับปรุงมาร์กหลักเพื่อใช้ในการเปรียบเทียบต่อไป โดยสุดท้ายพื้นที่ที่หาได้ทั้งหมดจะเป็นพื้นที่ขีดสุดที่มีการแบ่งระดับความเข้มงวดในการสกัด MSER และจะนำค่าระดับความเข้มงวดเป็นค่าคุณลักษณะเพื่อใช้ในการประมวลผลต่อไป

Algorithm 1 MSER Extraction with Multiple-threshold scheme

Declare: *Threshold* ← array of condition extraction start with strong to weak.
for each t in array *Threshold* **do**
    regions ← MSER objects extracted by using *Threshold*(t).
    for each r in array *regions* **do**
        create $mask_r$
        if not overlap($mask_r$,$mask_{main}$) **then**
            extract *regions*(r) attributes.
            $mask_{main}$ ← $mask_{main}$+$mask_r$ ▷ update $mask_{main}$

รูปที่ 3.4 รูป Pseudocode อัลกอริทึมของวิธีการสกัด MSER แบบหลายระดับความเข้มงวด



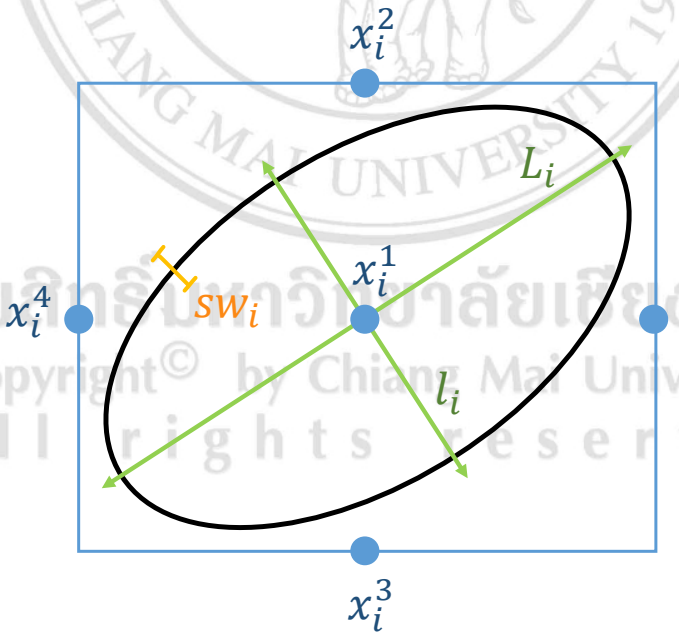
รูปที่ 3.5 ตัวอย่างการหาองค์ประกอบ MSER
ของวิธีการสกัด MSER แบบหลายระดับความเข้มงวด

3.2. การสกัดคุณลักษณะองค์ประกอบ MSER (Component Attribute Extraction)

จากกระบวนการก่อนหน้า พื้นที่องค์ประกอบตัวเลือกที่ได้มาจาก MSER จะนำมาสกัดหาคุณลักษณะต่าง ๆ ซึ่งจะเป็นข้อมูลเชิงขนาด, เชิงสี, เชิงตำแหน่ง และเชิงความกว้างของเส้น [22] เพื่อนำไปใช้ในกระบวนการจำแนกองค์ประกอบเพื่อคัดกรองสิ่งที่ไม่ใช่ตัวอักษรทิ้ง และหาความสัมพันธ์ระหว่างองค์ประกอบเพื่อจัดกลุ่มองค์ประกอบตัวเลือกให้เป็นกลุ่มเดียวกันต่อไป โดยคุณลักษณะที่ใช้มีดังนี้

1) พื้นที่ขององค์ประกอบตัวเลือก (Area) คือ จำนวนพิกเซลทั้งหมดที่เป็นสมาชิกของเซตองค์ประกอบตัวเลือก R_i ใด ๆ โดยจะแทนด้วยตัวแปร A_i

2) ตำแหน่งขององค์ประกอบตัวเลือก (Spatial Position) คือ ตำแหน่งขององค์ประกอบตัวเลือก R_i ใด ๆ โดยจะมีทั้งหมดห้าจุดคือ จุดศูนย์กลางมวล (Centroid) และ บริเวณกึ่งกลางของขอบทั้งสี่ด้านของกล่องขอบเขต (Bounding Box) ขององค์ประกอบตัวเลือกนั้น ๆ โดยกำหนดให้ตัวแปร $\mathbf{x}_i^{(j)}$ แทนค่าพิกัดขององค์ประกอบตัวเลือกแถวและหลักตามลำดับ ซึ่ง i จะเป็นลำดับขององค์ประกอบตัวเลือกใด ๆ และ j จะเป็นพิกัดของตำแหน่งต่าง ๆ โดยมีค่าตั้งแต่ $1-5$ คือ ตำแหน่งจุดศูนย์กลางมวล, บน, ล่าง, ซ้าย และขวาของกล่องขอบเขตตามลำดับ ดังแสดงในรูปที่ 3.6



รูปที่ 3.6 ตัวอย่างคุณลักษณะต่าง ๆ ขององค์ประกอบตัวเลือก

3) ค่าสี RGB และ HSV (RGB and HSV Color) คือ ค่าสีในช่องสัญญาณสีต่าง ๆ โดยใน RGB จะประกอบไปด้วยสีแดง เขียว และน้ำเงิน ส่วนใน HSV โดยจะประกอบไปด้วยค่าโทนสี (Hue) ค่าความเข้มสี (Saturation) และค่าความสว่างของสี (Value) ซึ่งในที่นี้จะใช้ค่ามัธยฐาน (Median) ของค่าสีในแต่ละช่องสีของทุกพิกเซลในองค์ประกอบตัวเลือก R_i ใด ๆ โดยจะแทนด้วยตัวแปร $\mathbf{C}_i = [r_i, g_i, b_i, h_i, s_i, v_i]$ โดย ค่า r g b h s และ v แทนด้วยสีแดง เขียว น้ำเงิน ค่าโทนสี ค่าความเข้มสี และค่าความสว่างของสีตามลำดับ

4) ความยาวแกนหลัก (Length of Major Axis) คือ ค่าความยาวระหว่างสองจุดใด ๆ ที่ห่างกันมากที่สุดขององค์ประกอบตัวเลือก R_i ใด ๆ โดยจะแทนด้วยตัวแปร L_i ดังแสดงในรูปที่ 3.4

5) ความยาวแกนรอง (Length of Minor Axis) ค่าความยาวที่มากสุด เมื่อพิจารณาเส้นที่ตั้งฉากกับแกนหลักในองค์ประกอบตัวเลือก R_i ใด ๆ โดยจะแทนด้วยตัวแปร l_i ดังแสดงในรูปที่ 3.4

6) ความหนาเฉลี่ยของเส้น (Average Stroke Width) [4] คือ ค่าที่แสดงถึงความกว้างเฉลี่ยขององค์ประกอบตัวเลือก R_i ใด ๆ โดยจะแทนด้วยตัวแปร sw_i ดังแสดงในรูปที่ 3.4 ซึ่งโดยปกติสามารถคำนวณได้จากบทที่ 2 แต่ในที่นี้เพื่อลดความซับซ้อนในการคำนวณ จะใช้การประมาณค่าความกว้างเฉลี่ยจากการนับพิกเซล โดยจะนับพิกเซลของเส้นขอบทั้งในแนวตั้งและแนวนอนทั้งหมด จากนั้นจึงนำมาหาค่าเฉลี่ยของทั้งองค์ประกอบตัวเลือก

7) ความแปรปรวนของความหนาของเส้น (Stroke Width Variance) คือ จากข้อ 7 จะสามารถนำค่าความกว้างเฉลี่ยของเส้นมาหาค่าความแปรปรวนของความหนาของเส้นในแต่ละองค์ประกอบตัวเลือก R_i ใด ๆ ได้ โดยจะแทนด้วยตัวแปร swv_i

8) จำนวนรู (Number of Hole) คือ การนับจำนวนรูที่เกิดขึ้นขององค์ประกอบตัวเลือก R_i ใด ๆ ซึ่งปกติสิ่งเป็นตัวอักษรมักจะมีจำนวนรูค่อนข้างน้อย โดยจะแทนด้วยตัวแปร NH_i

9) จำนวนพิกเซลที่ติดกับขอบรูป (Border Touch) คือ การนับจำนวนพิกเซลทั้งหมดที่ติดกับขอบรูปขององค์ประกอบตัวเลือก R_i ใด ๆ โดยจะแทนด้วยตัวแปร BT_i

10) อัตราส่วนของขนาด (Ratio) คือ ค่าอัตราส่วนระหว่างค่าความยาวแกนหลักเทียบกับความยาวแกนรอง ขององค์ประกอบตัวเลือก R_i ใด ๆ โดยจะแทนด้วยตัวแปร RT_i

11) ค่าระดับความเข้มงวด (Threshold Delta) คือ ค่าระดับค่าเกณฑ์ที่ใช้ในการสกัดหาพื้นที่ขององค์ประกอบตัวเลือก R_i ใด ๆ โดยจะแทนด้วยตัวแปร TD_i

12) ค่าความทึบ (Solidity) คือ ค่าอัตราส่วนระหว่างจำนวนพิกเซลเทียบกับค่าพื้นที่ในเส้นรอบรูป (Convex Hull) ขององค์ประกอบตัวเลือก R_i ใด ๆ ซึ่งยิ่งองค์ประกอบตัวเลือกนั้น ๆ เป็นทึบมากเท่าไหร่ ค่านี้ก็จะยิ่งเข้าใกล้ 1 มากขึ้นเท่านั้น โดยจะแทนด้วยตัวแปร SLD_i

13) ความยาวเส้นรอบวง (Perimeter) คือ ค่าอัตราส่วนระหว่างจำนวนพิกเซลรอบนอกขององค์ประกอบเทียบกับพื้นที่ของกล่องขอบเขตขององค์ประกอบตัวเลือก R_i ใด ๆ โดยจะแทนด้วยตัวแปร PER_i

14) ลำดับของระยะห่างจากจุดศูนย์กลางมวล (Order of Centroid Distance) เป็นค่าที่ได้จากการคำนวณฟังก์ชันระยะห่างจากจุดศูนย์กลางมวล (Centroid Distance Signature) เริ่มจากกำหนดให้ $\mathbf{P}_i = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n]$ โดยที่ $\mathbf{p}$ คือตำแหน่งของพิกเซลที่อยู่บนเส้นรอบรูปที่เรียงต่อกันขององค์ประกอบตัวเลือก R_i ใด ๆ ตั้งแต่ 1 ถึง n ซึ่งเป็นจำนวนพิกเซลทั้งหมดที่อยู่บนเส้นรอบรูปขององค์ประกอบตัวเลือกนั้น ๆ และระยะห่างจากจุดศูนย์กลางมวลไปยังพิกเซลแต่ละจุด d_k ใด ๆ โดยที่ k มีค่าตั้งแต่ 1 จนถึง n ก็สามารถคำนวณได้จากสมการ 3.1 และสุดท้ายก็จะได้ลำดับของระยะห่างจากจุดศูนย์กลางมวลไปยังพิกเซลรอบรูปทุกจุดคือ $\mathbf{D}_i = [d_1, d_2, \ldots, d_n]$

$$d_k = \left\| \mathbf{x}_k^1 - \mathbf{p}_k \right\|_2 \tag{3.1}$$

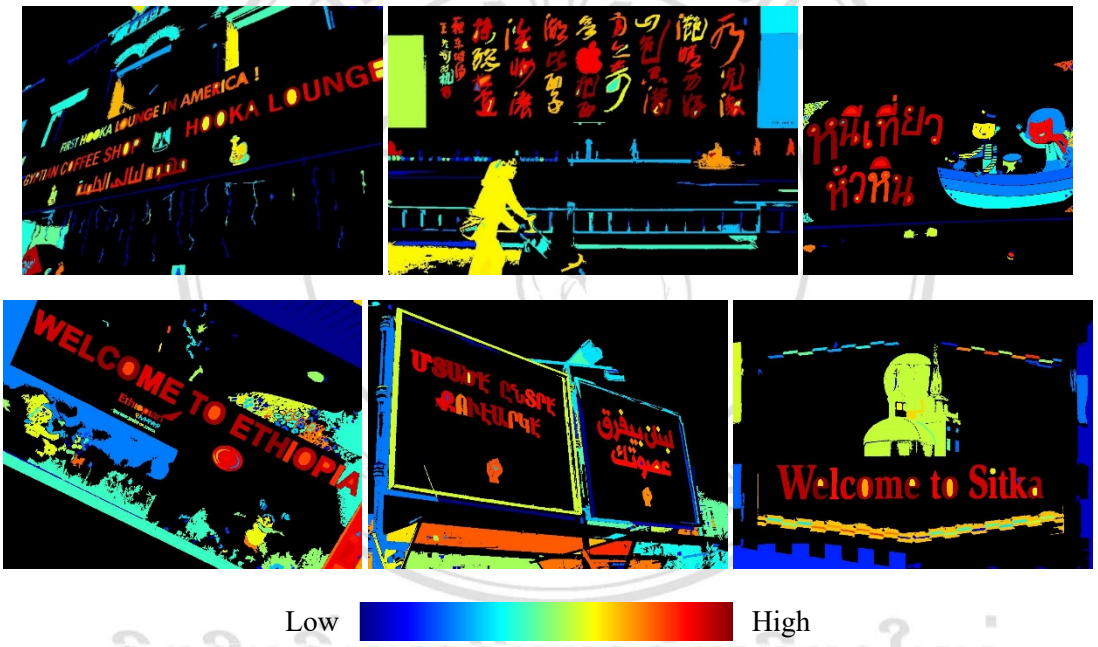
เนื่องจากองค์ประกอบตัวเลือกที่ได้มาจากขั้นตอนก่อนหน้าจะมีทั้งส่วนที่เป็นข้อความและไม่ใช่ข้อความ ดังนั้นในงานนี้จึงนำเสนอ “ส่วนวิเคราะห์องค์ประกอบตัวเลือก” (Analysis Candidate Components) เพื่อใช้จำแนกและจัดกลุ่มองค์ประกอบตัวเลือกเหล่านั้น โดยจะประกอบไปด้วยส่วนการทำงานสองส่วนคือ ส่วนการทำงานที่ให้คะแนนองค์ประกอบตัวเลือก (Component Grader) และส่วนการทำงานที่ให้คะแนนความสัมพันธ์ระหว่างคู่ขององค์ประกอบตัวเลือก (Component Linker) ซึ่งรายละเอียดจะอธิบายในส่วนถัดไป

3.3. การให้คะแนนองค์ประกอบตัวเลือก (Component Grader)

Component Grader เป็นส่วนให้คะแนนเพื่อหาว่าองค์ประกอบแต่ละตัวนั้นเป็นตัวอักษรหรือไม่ โดยจะวิเคราะห์แค่เฉพาะลักษณะรูปร่างของแต่ละองค์ประกอบย่อยอย่างเพียงเดียว และผลลัพธ์ที่ได้จะแบ่งออกเป็น 2 คลาส Text และ Non-Text ซึ่งคุณลักษณะที่ใช้จะนำมาจากการสกัดคุณลักษณะองค์ประกอบ MSER ของกระบวนการก่อนหน้า โดยจะใช้คุณลักษณะที่ 8 – 13 ซึ่งคุณลักษณะในงานนี้จะไม่ได้ขึ้นกับภาษาใด ๆ เพื่อให้เหมาะกับแนวคิดที่สามารถระบุตำแหน่งของข้อความ โดยเป็นอิสระต่อภาษาใด ๆ แต่นอกเหนือจากงานนี้หากต้องการรู้จำเฉพาะภาษาก็สามารถเพิ่มคุณลักษณะที่มีผลกับรู้จำตัวอักษรเช่น CNN (Convolutional Neural Network) หรือ HOG (Histogram of Oriented Gradient) เป็นต้น

ซึ่งข้อมูลที่ใช้ในการฝึกสอนจะเป็นตัวอย่างองค์ประกอบตัวเลือกที่เป็นตัวอักษร และไม่ใช่ตัวอักษร โดยในการฝึกสอนครั้งแรกนั้นจะใช้ชุดข้อมูลฝึกสอนที่มีตัวอย่างไม่มาก ทำให้อาจจะไม่ครอบคลุมพยัญชนะ สระ หรือรูปแบบอักษร (Font) ของแต่ละภาษาทั้งหมด ดังนั้นในงานวิจัยนี้จึงนำเสนอวิธีการที่สามารถค้นหาข้อมูลฝึกสอนที่ไม่ได้มีอยู่ในชุดฝึกสอนแรกเพิ่มเติมขณะทำงาน และเรียนรู้การทำงานด้วยข้อมูลเหล่านั้นด้วยตัวเองได้

โดยในรูปที่ 3.7 จะเป็นการแสดงค่าคะแนนขององค์ประกอบตัวเลือกแต่ละตัวที่ได้จากการทำงานของ Component Grader ซึ่งจะเห็นได้ว่าองค์ประกอบตัวเลือกที่เป็นตัวอักษรจะมีค่าคะแนนสูงเมื่อเทียบกับองค์ประกอบตัวเลือกที่ไม่ได้เป็นตัวอักษร โดยสีแดงคือมีค่าคะแนนสูง และสีน้ำเงินคือค่าคะแนนต่ำ

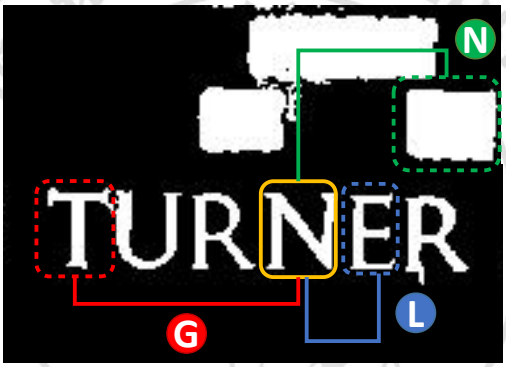


รูปที่ 3.7 ตัวอย่างค่าคะแนนขององค์ประกอบตัวเลือกที่ได้จาก Component Grader

3.4. การให้คะแนนความสัมพันธ์ขององค์ประกอบตัวเลือก (Component Linker)

Component Linker เป็นส่วนให้คะแนนเพื่อหาความสัมพันธ์ของแต่ละคู่ขององค์ประกอบตัวเลือกว่ามีความสัมพันธ์อยู่ในกลุ่มเดียวกันหรือไม่ จากการวิเคราะห์ค่าความต่างระหว่างคู่ขององค์ประกอบนั้น ๆ โดยจะแบ่งความสัมพันธ์ออกเป็น 3 คลาสตามรูปที่ 3.8 คือ

1) “Link (L)” ความสัมพันธ์ขององค์ประกอบตัวเลือกที่อยู่ในคำเดียวกันและอยู่ติดกัน
2) “Group (G)” ความสัมพันธ์ขององค์ประกอบตัวเลือกที่อยู่ในคำเดียวกันแต่ไม่ได้อยู่ติดกัน
3) “Non-Group (N)” ความสัมพันธ์ขององค์ประกอบตัวเลือกที่ไม่อยู่ในคำเดียวกัน



รูปที่ 3.8 ตัวอย่างคลาสของความสัมพันธ์ของแต่ละคู่ขององค์ประกอบตัวเลือก

รูปที่ 3.9 ตัวอย่างเส้นเชื่อมระหว่างคู่ขององค์ประกอบตัวเลือกที่อยู่ในคลาส Link

โดยผลลัพธ์ที่ได้นั้นสามารถนำมาจัดกลุ่มขององค์ประกอบตัวเลือกได้จากการนำคู่ขององค์ประกอบเชื่อมด้วยคลาส Link หรือคลาสที่คู่ขององค์ประกอบตัวเลือกที่อยู่ในคำเดียวกันและอยู่ติดกันมาจัดให้อยู่ในกลุ่มเดียวกัน และจากรูปที่ 3.9 เป็นตัวอย่างของการเชื่อมคู่ขององค์ปะกอบของคลาส Link นั้นจะเห็นได้ว่าสามารถเชื่อมองค์ประกอบได้ทั้งแนวตั้งและแนวนอน ซึ่งเหมาะกับการทำงานกับภาษาต่าง ๆ ที่มีความหลากหลายทางด้านทิศทางของการเรียงตัวอักษรในข้อความ

เพื่อที่จะสร้างส่วนให้คะแนนของ Component Linker หรือตัวจำแนกความสัมพันธ์ระหว่างคู่องค์ประกอบ R_i และ R_j ใด ๆ ได้นั้น ต้องนำคุณลักษณะระหว่างองค์ประกอบมาวิเคราะห์ว่ามีความสัมพันธ์เชื่อมต่อกันหรือไม่ โดยนิยามความต่างของคุณลักษณะในมิติต่าง ๆ ระหว่างคู่องค์ประกอบตัวเลือกใด ๆ ที่ใช้ในงานนี้เพื่อใช้เป็นคุณลักษณะของ Component Linker มีดังนี้

1) ระยะขจัดระหว่าง Component (Normalized Spatial Distance) คือ ระยะห่างระหว่างตำแหน่งขององค์ประกอบตัวเลือกสององค์ประกอบตัวเลือกใด ๆ ซึ่งจะถูกหารด้วยความยาวแกนหลักที่สั้นที่สุดระหว่างคู่องค์ประกอบตัวเลือก R_i และ R_j ใด ๆ เพื่อให้ระยะห่างเป็นไปจริงตามขนาดของรูป โดยจะแทนด้วยตัวแปร SPD_{ij} และสามารถคำนวณได้จากสมการที่ 3.2

$$SPD_{ij} = \frac{\left\| \mathbf{x}_i^1 - \mathbf{x}_j^1 \right\|_2}{\min(l_i, l_j)} \tag{3.2}$$

2) ค่าความต่างของความเข้มสีระหว่าง Component (Color Difference) คือ การเปรียบเทียบความต่างของคุณลักษณะสีของแต่ละองค์ประกอบตัวเลือก R_i และ R_j ใด ๆ โดยจะแทนด้วยตัวแปร CLD_{ij} และสามารถคำนวณได้จากสมการที่ 3.3

$$CLD_{ij} = \left\| \mathbf{C}_i - \mathbf{C}_j \right\|_2 \tag{3.3}$$

3) อัตราส่วนของความหนาของเส้น (Stroke Width Ratio) คือ ค่าอัตราส่วนสำหรับเปรียบเทียบความต่างของความหนาของเส้นเฉลี่ยระหว่างองค์ประกอบตัวเลือก R_i และ R_j ใด ๆ โดยจะแทนด้วยตัวแปร SWR_{ij} และสามารถคำนวณได้จากสมการที่ 3.4

$$SWR_{ij} = \left| \log\left(\frac{sw_i}{sw_j} \right) \right| \tag{3.4}$$

4) อัตราส่วนของขนาด (Size Ratio) คือ ค่าอัตราส่วนระหว่างพื้นที่ของกล่องขอบเขตของทั้งสององค์ประกอบตัวเลือก R_i และ R_j ใด ๆ โดยจะแทนด้วยตัวแปร SZD_{ij} ซึ่งพื้นที่ในที่นี้คำนวณจากแกนหลักและแกนรองของกล่องขอบเขตขององค์ประกอบตัวเลือกนั้น ๆ และสามารถคำนวณได้จากสมการที่ 3.5

$$SZD_{ij} = \left| \log\left(\frac{L_i \times l_i}{L_j \times l_j} \right) \right| \tag{3.5}$$

5) ค่าความต่างของความยาวแกนหลักระหว่าง Component (Major Axis Difference) คือ การเปรียบเทียบความต่างระหว่างค่าความยาวแกนหลักของแต่ละองค์ประกอบตัวเลือก R_i และ R_j ใด ๆ โดยจะแทนด้วยตัวแปร MJD_{ij} และสามารถคำนวณได้จากสมการที่ 3.6

$$MJD_{ij} = \left| \log\left(\frac{L_i}{L_j} \right) \right| \tag{3.6}$$

6) ค่าความต่างของความยาวแกนรองระหว่าง Component (Minor Axis Difference) คือ การเปรียบเทียบความต่างระหว่างค่าความยาวแกนรองของแต่ละองค์ประกอบตัวเลือก R_i และ R_j ใด ๆ โดยจะแทนด้วยตัวแปร MID_{ij} และสามารถคำนวณได้จากสมการที่ 3.7

$$MID_{ij} = \left| \log\left(\frac{l_i}{l_j} \right) \right| \tag{3.7}$$

7) อัตราส่วนความแปรปรวนของความหนาของเส้นที่น้อยสุดระหว่าง Component (Minimum Stroke Width Ratio) คือ ค่าอัตราส่วนระหว่างค่าความแปรปรวนของความหนาของเส้นเทียบกับค่าความหนาของเส้นขององค์ประกอบตัวเลือกแต่ละตัว โดยจะเลือกค่าอัตราส่วนที่น้อยที่สุดระหว่างองค์ประกอบตัวเลือก R_i และ R_j ใด ๆ และจะแทนด้วยตัวแปร $MiSWR_{ij}$ และสามารถคำนวณได้จากสมการที่ 3.8

$$MiSWR_{ij} = \min\left(\frac{swv_i}{sw_i}, \frac{swv_j}{sw_j} \right) \tag{3.8}$$

8) อัตราส่วนความแปรปรวนของความหนาของเส้นที่มากสุดระหว่าง Component (Maximum Stroke Width Ratio) คือ ค่าอัตราส่วนระหว่างค่าความแปรปรวนของความหนาของเส้นเทียบกับค่าความหนาของเส้นขององค์ประกอบตัวเลือกแต่ละตัว โดยจะเลือกค่าอัตราส่วนที่มีค่ามากที่สุดระหว่างองค์ประกอบตัวเลือก R_i และ R_j ใด ๆ และจะแทนด้วยตัวแปร $MxSWR_{ij}$ และสามารถคำนวณได้จากสมการที่ 3.9

$$MxSWR_{ij} = \max\left(\frac{swv_i}{sw_i}, \frac{swv_j}{sw_j}\right) \tag{3.9}$$

9) ค่าความต่างของระดับความเข้มงวดระหว่าง Component (Threshold Delta Difference) คือ การเปรียบเทียบความต่างระหว่างค่าระดับความเข้มงวดของแต่ละองค์ประกอบตัวเลือก R_i และ R_j ใด ๆ โดยจะแทนด้วยตัวแปร TDD_{ij} และสามารถคำนวณได้จากสมการที่ 3.10

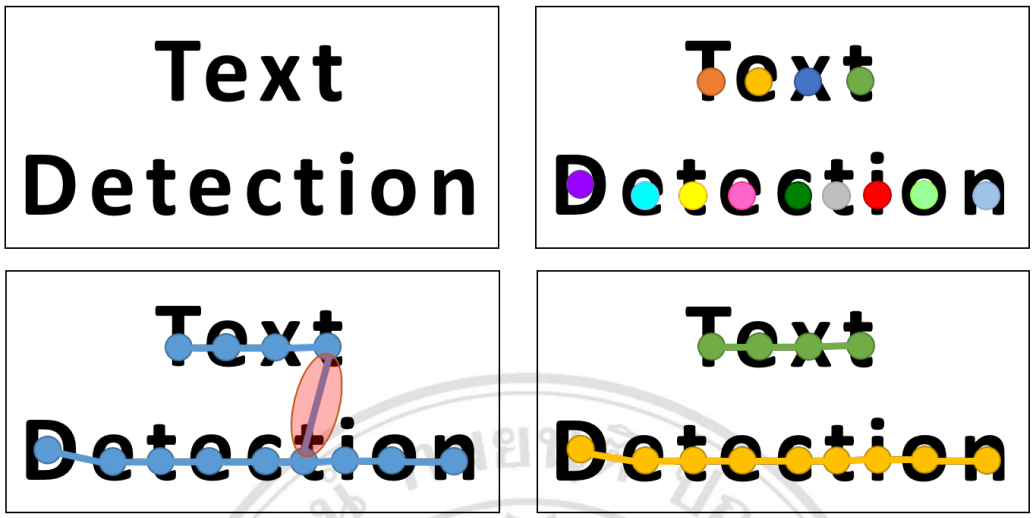
$$TDD_{ij} = \left|TD_i - TD_j\right| \tag{3.10}$$

10) ค่าอันดับความใกล้ระหว่าง Component (Distance Rank) คือ อันดับระยะห่างระหว่างตำแหน่งขององค์ประกอบตัวเลือกสององค์ประกอบตัวเลือก R_i และ R_j ใด ๆ เทียบกับองค์ประกอบตัวเลือกทุกตัว โดยจะแทนด้วยตัวแปร DTR_{ij}

3.5. การรวมกลุ่ม (Clustering)

หลังจากการทำงานของ Component Linker ในขั้นตอนก่อนหน้านี้ ทำให้ได้ความสัมพันธ์ระหว่างองค์ประกอบตัวเลือก โดยในขั้นตอนนี้จะนำความสัมพันธ์ที่ได้นั้นมาจัดกลุ่มขององค์ประกอบ โดยคู่ขององค์ประกอบตัวเลือกใด ๆ ที่อยู่ในคลาส Link หรือคลาสที่อยู่ในคำเดียวกันและอยู่ติดกัน จะถูกรวมเข้าด้วยกันเป็นกลุ่มเดียวกัน จากนั้นจะทำการแบ่งบรรทัดเพื่อแก้ปัญหากรณีที่กลุ่มข้อความที่เชื่อมถึงกันไปเชื่อมกับตัวอักษรบรรทัดอื่น ๆ หรือสิ่งที่ไม่ใช่ข้อความ

จากรูปที่ 3.10 เริ่มต้นหลังจากที่สกัดองค์ประกอบตัวเลือกมาแล้วองค์ประกอบตัวเลือกของแต่ละตัวอักษรจะแสดงด้วยสีที่แตกต่างกัน จากนั้นจึงจะรวมกลุ่มโดยใช้ผลลัพธ์ความสัมพันธ์ที่ได้จาก Component Linker ซึ่งในบางครั้งจะมีการเชื่อมตัวอักษรระหว่างบรรทัดโดยในตัวอย่างนี้คือตัว “t” ของทั้งสองคำได้เชื่อมถึงกันทำให้ทั้งสองบรรทัดเชื่อมกันจนเป็นสีเดียวกัน ดังนั้นการแก้ปัญหาในกรณีนี้คือจะทำการหาทิศทางหลักขององค์ประกอบตัวเลือกภายในกลุ่ม โดยหากเส้นเชื่อมใดที่มีทิศทางต่างจากทิศทางหลัก และมีระยะห่างกับตัวข้าง ๆ มากเกินไปเมื่อเทียบกับตัวอื่น ๆ ในกลุ่มก็จะทำการตัดเส้นเชื่อมนั้นทิ้ง และทำการรวมกลุ่มใหม่อีกครั้ง



รูปที่ 3.10 ตัวอย่างการรวมกลุ่มและแบ่งบรรทัด

3.6. การสกัดคุณลักษณะของกลุ่ม (Group Attributes Extraction)

กลุ่มแต่ละกลุ่มที่ได้มาจากขั้นตอนที่แล้วสามารถนำไปหาคุณลักษณะเพื่อนำไปวิเคราะห์และตัดสินใจว่าเป็นข้อความหรือไม่ได้ โดยตัวอย่างคุณลักษณะที่ใช้มีดังนี้

1) จำนวนองค์ประกอบตัวเลือกในกลุ่ม (Number of Components) คือ จำนวนขององค์ประกอบตัวเลือก R_i ใด ๆ ที่เป็นสมาชิกในกลุ่ม G_i นั้น ๆ โดยจะแทนด้วยตัวแปร NC_i

2) ค่าความแปรปรวนของความหนาของเส้นขององค์ประกอบตัวเลือกในกลุ่ม (Average Stroke Width Variance) คือ ค่าเฉลี่ยของค่าความแปรปรวนของความหนาของเส้นแต่ละองค์ประกอบ R_i ใด ๆ ที่เป็นสมาชิกในกลุ่ม G_i นั้น ๆ โดยจะแทนด้วยตัวแปร $GSWV_i$

3) จำนวนพื้นที่โดนปิดล้อมในกลุ่ม (Number of Hole) คือ การนับจำนวนรูที่เกิดขึ้นภายในกลุ่ม G_i ใด ๆ โดยจะแทนด้วยตัวแปร NH_i

4) ค่าขอบเขตเฉลี่ย (Average Extent) คือ ค่าเฉลี่ยของอัตราส่วนระหว่างจำนวนพิกเซลเทียบกับพื้นที่ของกล่องขอบเขตสี่เหลี่ยมที่ครอบคลุมส่วนที่เป็นองค์ประกอบตัวเลือก R_i ใด ๆ ที่เป็นสมาชิกในกลุ่ม G_i นั้น ๆ โดยจะแทนด้วยตัวแปร $GEXT_i$

5) ค่าความทึบเฉลี่ย (Average Solidity) คือ ค่าเฉลี่ยของอัตราส่วนระหว่างจำนวนพิกเซลเทียบกับค่าพื้นที่ในเส้นรอบรูป (Convex Hull) ขององค์ประกอบตัวเลือก R_i ใด ๆ ที่เป็นสมาชิกในกลุ่ม G_i นั้น ๆ โดยจะแทนด้วยตัวแปร $GSLD_i$

6) องค์ประกอบความถี่สูงของระยะห่างจากจุดศูนย์กลางมวลเฉลี่ย (Average High-Frequency Component of Centroid Distance Signature) คือ ค่าที่ได้จากการคำนวณหาระยะห่างจากจุดศูนย์กลางมวล D_i ขององค์ประกอบตัวเลือก R_i ใด ๆ ที่เป็นสมาชิกในกลุ่ม G_i นั้น ๆ และนำไปหากำลังของผลลัพธ์ที่ได้จากการผ่านตัวกรองความถี่สูง โดยจะแทนด้วยตัวแปร $HFCD_i$

7) คะแนนจากการรู้จำเชิงพื้นผิวด้วยคุณลักษณะเด่นของฮาราลิกค์ (Haralick's Feature Texture Classification Score) เป็นคุณลักษณะเด่นเชิงพื้นผิว โดยจะหาค่าลักษณะเด่นของฮาราลิกค์ที่ประกอบไปด้วย Contrast, Energy และ Homogeneity จากค่าสถิติอันดับสองของค่าอันดับเทา (Gray level co-occurrence matrix) โดยจะแทนด้วยตัวแปร $HFCS_i$

8) ค่าคะแนนองค์ประกอบตัวเลือก (Component Grader Score) คือ ค่าเฉลี่ยของค่าคะแนนองค์ประกอบตัวเลือกที่เป็นตัวอักษรที่ได้จาก Component Grader ขององค์ประกอบตัวเลือก R_i ใด ๆ ที่เป็นสมาชิกในกลุ่ม G_i นั้น ๆ โดยจะแทนด้วยตัวแปร GRD_i เนื่องจากกลุ่มที่ส่วนใหญ่ประกอบไปด้วยองค์ประกอบตัวเลือกที่มีค่าคะแนนจาก Component Grader สูงจะเป็นกลุ่มที่เป็นกลุ่มของตัวอักษร

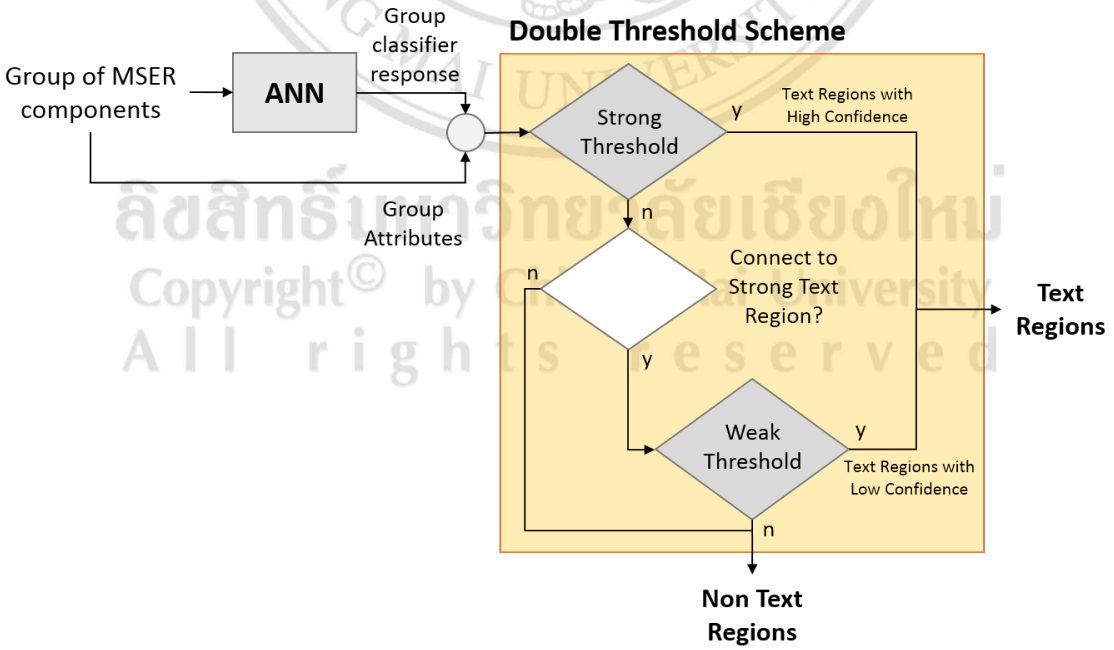
9) จำนวนความสัมพันธ์ของคู่องค์ประกอบตัวเลือก (Component Linker Score) คือ ค่าเฉลี่ยของจำนวนความสัมพันธ์ที่ได้จาก Component Linker ขององค์ประกอบตัวเลือก R_i ใด ๆ ที่เป็นสมาชิกในกลุ่ม G_i นั้น ๆ โดยจะแทนด้วยตัวแปร LNK_i เนื่องจากกลุ่มที่ส่วนใหญ่ประกอบไปด้วยองค์ประกอบตัวเลือกที่มีจำนวนความสัมพันธ์ของคลาส Link และคลาส Group จาก Component Linker มากจะเป็นกลุ่มที่เป็นกลุ่มของตัวอักษร

ซึ่งคุณลักษณะเด่นต่าง ๆ นั้นจะมีคุณสมบัติที่สามารถแยกแยะสิ่งที่เป็นข้อความและไม่ใช่ข้อความได้ต่างกันออกไป เช่น จำนวนองค์ประกอบนั้นข้อความปกติจะไม่ได้มีแค่ตัวเดียว หรือตัวอักษรที่อยู่ในข้อความเดียวกันจะมีความหนาของเส้นที่ใกล้เคียงกันซึ่งก็สามารถวิเคราะห์ได้จากค่าความแปรปรวนความหนาของเส้นภายในแต่ละกลุ่ม นอกจากนี้ค่าความทึบ ค่าขอบเขตเฉลี่ย และจำนวนรูซึ่งเป็นส่วนหนึ่งของการวิเคราะห์ความหนาแน่นของพิกเซล และโดยปกติแล้วตัวอักษรจะมีค่าความหนาแน่นเหล่านี้มากระดับหนึ่งเมื่อเทียบกับพื้นหลังที่มีความทึบสูงจำพวกอิฐบล็อกที่มีลักษณะขององค์ประกอบเป็นทรงตัน เป็นต้น ดังนั้นจึงเลือกใช้คุณลักษณะเด่นเหล่านี้เพื่อใช้จำแนกกลุ่มของข้อความในขั้นตอนต่อไป

3.7. การจำแนกกลุ่มข้อความ (Group Classification)

ขั้นตอนสุดท้ายจะทำการจำแนกประเภทของกลุ่มที่ได้มาว่าเป็นกลุ่มที่เป็นตัวอักษร (Text Regions) หรือกลุ่มที่ไม่ใช่ตัวอักษร (Non-text Regions) โดยได้นำการทำงานจากงานวิจัยของ Wiwatcharakoses และคณะ [8] ที่ได้นำเสนอเกณฑ์การเปรียบเทียบสองระดับ (Double Threshold Scheme) ซึ่งคล้ายกับการหาเส้นขอบของวิธีการแคนนี่ (Canny edge detector) มาประยุกต์ใช้ ดังรูปที่ 3.11 โดยจะแบ่งผลลัพธ์เป็นสามคลาสคือ กลุ่มที่มีความมั่นใจสูงว่าจะเป็นข้อความ กลุ่มที่มีความมั่นใจต่ำว่าจะเป็นข้อความ และกลุ่มที่ไม่ใช่ข้อความ ซึ่งจะนำคุณลักษณะเด่นของกลุ่มจากขั้นตอนก่อนหน้า ร่วมกับค่าคะแนนที่ได้จากการคัดกรองกลุ่มด้วย ANN ที่มีการเรียนรู้แบบแพร่ย้อนกลับโดยใช้คุณลักษณะเด่นของกลุ่ม นำเข้าไปจำแนกโดยใช้เกณฑ์เปรียบเทียบ 2 ระดับ ที่ประกอบไปด้วยเกณฑ์เปรียบเทียบที่มีเงื่อนไขเข้มงวด และเกณฑ์เปรียบเทียบที่มีเงื่อนไขผ่อนผัน

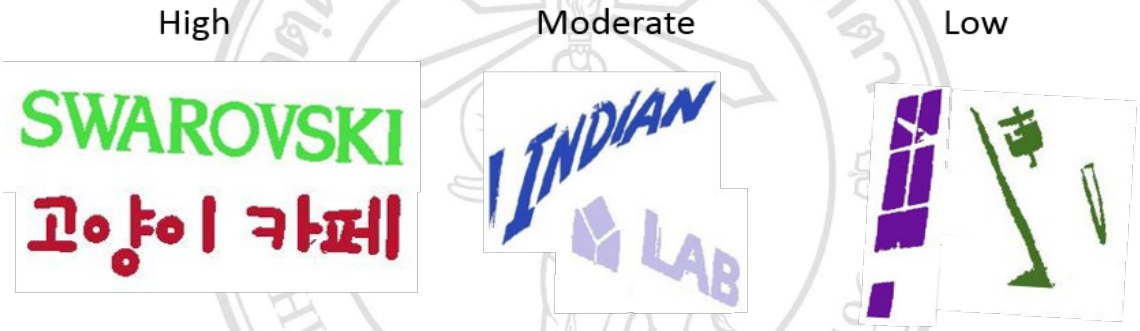
กลุ่มที่มีความมั่นใจสูงว่าจะเป็นข้อความคือกลุ่มที่สามารถผ่านเกณฑ์เปรียบเทียบที่มีเงื่อนไขเข้มงาดได้ แต่กลับกันหากไม่สามารถผ่านเงื่อนไขเข้มงวดแต่สามารถผ่านเงื่อนไขผ่อนผัน และมีความสัมพันธ์กับกลุ่มที่มีความมั่นใจสูงจะยังกลายเป็นกลุ่มที่มีความมั่นใจต่ำได้ แต่ท้ายที่สุดหากกลุ่มใด ๆ ไม่สามารถผ่านเกณฑ์เปรียบเทียบทั้งสองได้จะกลายเป็นกลุ่มที่ไม่ใช่ข้อความนั่นเอง โดยความสัมพันธ์ที่นำมาพิจารณานั้นมีหลายความสัมพันธ์เช่น ระยะห่าง ความต่างของแกนหลัก ความต่างของสี ความต่างของความหนาของเส้น ความต่างขององศาภายในกลุ่ม เป็นต้น



รูปที่ 3.11 แผนภาพการทำงานของการจำแนกกลุ่มข้อความ

3.8. การปรับปรุงการทำงานของตัวจำแนก (Classifier Update)

ในส่วนการทำงานนี้จะทำการคัดเลือกตัวแทนขององค์ประกอบจากรูปต่าง ๆ หลังจากที่โปรแกรมทำการจัดกลุ่มข้อความแล้ว เพื่อนำไปใช้ฝึกสอนกับส่วนจำแนกทั้งหมดคือ Component Grader, Component Linker และส่วนจำแนกกลุ่มข้อความ (Group Classification) โดยจะเลือกตัวแทนของทั้งองค์ประกอบ และความสัมพันธ์ของคู่ขององค์ประกอบ เพื่อเพิ่มความสามารถในการคัดเลือกองค์ประกอบที่เป็นตัวอักษร และปรับปรุงการทำงานในส่วนการจัดกลุ่มองค์ประกอบให้ดีขึ้น ซึ่งจะพิจารณาด้วยค่าคะแนนจาก Component Grader, Component Linker และ Group Classifier ซึ่งสามารถดูตัวอย่างกลุ่มที่มีค่าคะแนนแบบต่าง ๆ ได้จากรูปที่ 3.12 โดยแต่ละส่วนมีรายละเอียดวิธีการเลือกตัวแทนดังต่อนี้



รูปที่ 3.12 ตัวอย่างกลุ่มที่มีค่าคะแนนจาก Group Classifier ที่ต่างกัน

1) Component Grader

- ตัวแทนองค์ประกอบที่เป็นข้อความ จะพิจารณาตัวแทนองค์ประกอบจากกลุ่มที่มีค่าคะแนนจาก Group Classifier สูง และมีค่าเบี่ยงเบนมาตรฐานขององศาในการเรียงตัวภายในกลุ่มน้อย จากนั้นจึงจะพิจารณาองค์ประกอบจากกลุ่มที่ได้มา โดยเลือกองค์ประกอบที่มีจำนวนความสัมพันธ์แบบคลาส L เชื่อมต่อกับองค์ประกอบอื่น ๆ น้อย และค่าความน่าจะเป็นของความสัมพันธ์นั้น ๆ มีค่ามาก เนื่องจากว่าโดยปกติแล้วองค์ประกอบที่เป็นตัวอักษรจะเชื่อมต่อกับองค์ประกอบอื่น ๆ รอบข้างไม่มาก จากธรรมชาติการเรียงตัวของข้อความ เช่นในภาษาอังกฤษที่มีการเรียงตัวแค่ในแนวเดียว ดังนั้นจึงมีการเชื่อมต่อระหว่างตัวอักษรไม่เกินสองเส้นเชื่อมเท่านั้น เป็นต้น และเส้นที่เชื่อมระหว่างองค์ประกอบนั้นจะต้องมีค่าความมั่นใจสูง จึงจะเลือกมาเป็นตัวแทนของคลาสที่เป็นตัวอักษรหรือ Text

- ตัวแทนองค์ประกอบที่ไม่เป็นข้อความ จะพิจารณาตัวแทนองค์ประกอบจากกลุ่มที่มีค่าคะแนนจาก Group Classifier น้อย จากนั้นจึงจะพิจารณาองค์ประกอบจากกลุ่มที่ได้มา โดยเลือกองค์ประกอบที่มีจำนวนความสัมพันธ์แบบคลาส L เชื่อมต่อกับองค์ประกอบอื่น ๆ มาก และเป็นองค์ประกอบที่ถูกวิเคราะห์จาก Component Grader ว่าเป็นองค์ประกอบที่ไม่ใช่ข้อความ และเนื่องจากเงื่อนไขที่พิจารณาทำให้จำนวนขององค์ประกอบที่ได้ในส่วนนี้มีปริมาณมาก ดังนั้นจะเลือกองค์ประกอบบางส่วนมาเป็นตัวแทนของคลาส Non-Text โดยใช้ค่าความน่าจะเป็นที่ไม่ใช่ข้อความจาก Component Grader เป็นปัจจัยในการพิจารณา

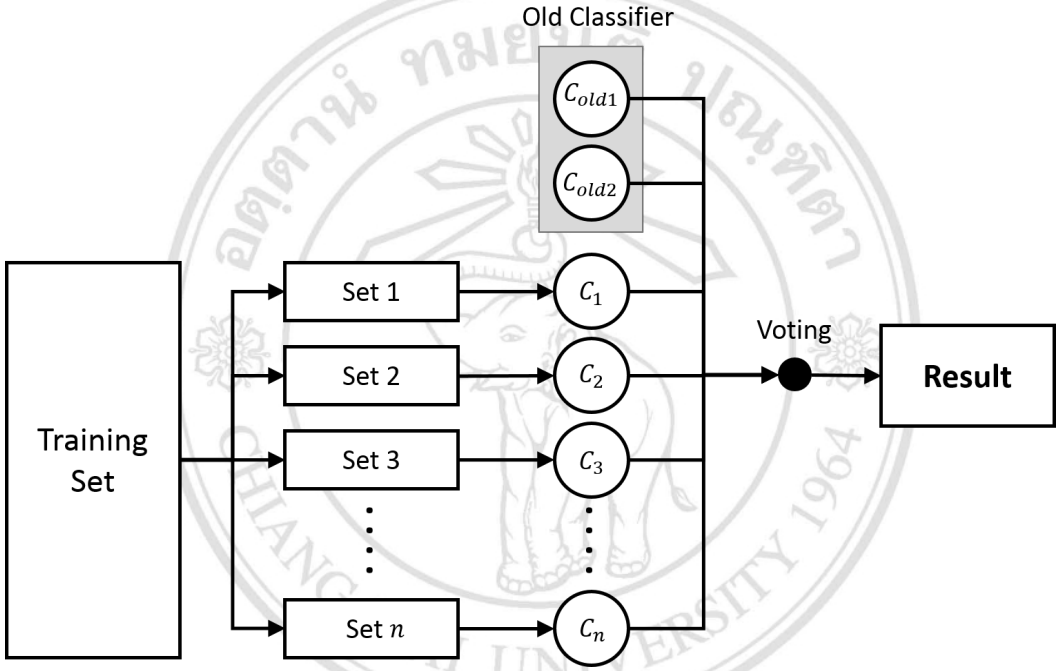
2) Component Linker

- ตัวแทนคู่ขององค์ประกอบคลาส L จะพิจารณาจากกลุ่มที่มีค่าคะแนนจาก Group Classifier ปานกลางถึงสูง และมีค่าเบี่ยงเบนมาตรฐานขององศาในการเรียงตัวภายในกลุ่ม $GSDL_i$ น้อย จากนั้นจึงจะหาคู่ขององค์ประกอบที่อยู่ติดกันจากค่าอันดับความใกล้เป็นอันดับแรก และทั้งสององค์ประกอบนั้นเป็นองค์ประกอบที่ถูกวิเคราะห์จาก Component Grader ว่าเป็นองค์ประกอบที่เป็นข้อความ และจะเลือกจำนวนตัวแทน โดยกำหนดเป็นตัวแปร *NLink* เพื่อเป็นตัวแทนของคลาส L
- ตัวแทนคู่ขององค์ประกอบคลาส G จะพิจารณาจากตัวแทนของคลาส L โดยจะเลือกคู่ขององค์ประกอบที่อยู่ในกลุ่มเดียวกันแต่ไม่ได้อยู่ติดกัน และจะเลือกจำนวนตัวแทน โดยกำหนดเป็นตัวแปร *NGroup* เพื่อเป็นตัวแทนของคลาส G
- ตัวแทนคู่ขององค์ประกอบคลาส N จะพิจารณาจากกลุ่มที่มีค่าคะแนนจาก Group Classifier น้อย และหาองค์ประกอบจากกลุ่มนี้โดยองค์ประกอบที่ถูกวิเคราะห์จาก Component Grader ว่าเป็นองค์ประกอบที่ไม่เป็นข้อความ จากนั้นจึงจะนำมาจับคู่กับตัวแทนของคลาส L ที่ได้มาก่อนหน้านี้ ซึ่งเป็นคู่ขององค์ประกอบตัวเลือกที่ ถูกวิเคราะห์จาก Component Linker ว่าเป็นคู่ที่ไม่ได้อยู่ในคำเดียวกัน และจะเลือกจำนวนตัวแทน โดยกำหนดเป็นตัวแปร *NNon* เพื่อเป็นตัวแทนของคลาส N

ซึ่งในการทำงานของ Component Linker นั้นจะแบ่งวิธีการเลือกออกเป็น 2 แบบ คือแบบเข้มงวดและแบบผ่อนผัน เพื่อให้ได้ตัวแทนที่เหมาะสมกับข้อมูลที่แตกต่างกันซึ่งจะกล่าวถึงในบทที่ 4

และในส่วนของ Group Classification นั้น โดยหลังจากที่ Component Grader และ Component Linker ได้ทำการอัพเดตแล้ว ทำให้ค่าคะแนนที่เป็นคุณลักษณะเด่นเพื่อใช้ในการจำแนกข้อความได้เปลี่ยนไปจากเดิม ดังนั้นจึงจำเป็นต้องมีการฝึกสอนตัวจำแนกด้วยค่าของคุณลักษณะใหม่ โดยในงานวิจัยนี้การเรียนรู้ในส่วนนี้จะไม่ได้มีการเปลี่ยนแปลงจำนวนข้อมูลฝึกสอนดังเช่น Component Grader และ Component Linker

3.9. การเรียนรู้แบบเพิ่มพูน (Incremental Learning)



รูปที่ 3.13 ตัวอย่างการทำงานแบบเพิ่มพูน โดยใช้วิธีการ Voting

ในงานวิจัยนี้ได้มีเสนอการทำงานแบบเพิ่มพูนของส่วนการงานของตัวจำแนกทั้งสองคือ Component Liner และ Component Grader ซึ่งในขั้นตอนการทำงานนั้นสามารถดูได้จากรูปที่ 3.13 กล่าวคือได้มีการแบ่งชุดฝึกสอนออกเป็นหลาย ๆ ส่วน เพื่อนำมาสร้างตัวจำแนก และนำตัวจำแนกเหล่านั้นมาช่วยกันตัดสินใจเพื่อให้ได้ผลลัพธ์สุดท้าย โดยในงานวิจัยนี้จะเรียกการทำงานในส่วนนี้ว่า “Voting”

วิธีการแบ่งชุดข้อมูลฝึกสอนนั้นเริ่มต้นจากสมการที่ 3.11 เพื่อหาค่า α หรือจำนวนข้อมูลของคลาสที่มีจำนวนน้อยที่สุดมา โดยคลาสที่มีจำนวนข้อมูลน้อยที่สุดจะแสดงด้วย A เช่นใน Component Grader จะเทียบจำนวนระหว่างคลาส Text และคลาส Non-Text เป็นต้น จากนั้นจะกำหนดค่า n หรือจำนวนเซตของข้อมูลฝึกสอนที่ต้องการจะแบ่ง และค่า x หรือจำนวนของข้อมูล

ฝึกสอนตั้งต้นที่ต้องการ โดยค่านี้จะมีค่าไม่เกิน α เพื่อใช้เป็นค่าตั้งต้นสำหรับกำหนดจำนวนข้อมูลของคลาสใด ๆ ในแต่ละเซต จากนั้นจะคำนวณจากสมการที่ 3.12 เพื่อหาค่า r หรือค่าที่กำหนดอัตราส่วนระหว่างแต่ละคลาส และสุดท้ายหลังจากที่ได้ครบทุกค่าก็จะสามารถคำนวณหา Y หรือจำนวนข้อมูลฝึกสอนสำหรับคลาสใด ๆ ในแต่ละเซตได้จากสมการที่ 3.13 (เว้นแต่คลาส A จะมีจำนวนข้อมูลเท่ากับ x)

$$[\alpha, \mathrm{A}] = \min(N_1, N_2, ..., N_m) \tag{3.11}$$

$$n \times r_m \times x \geq N_m \tag{3.12}$$

$$Y = r_m \times x \tag{3.13}$$

โดยกำหนดให้

- α คือจำนวนข้อมูลของคลาสที่มีค่าน้อยที่สุด
- A คือคลาสที่มีจำนวนข้อมูลน้อยที่สุด
- N คือจำนวนข้อมูลของคลาสตั้งแต่ 1 ถึง m
- n คือจำนวนเซตของข้อมูลฝึกสอนที่ต้องการจะแบ่ง
- r คือค่าอัตราส่วนระหว่างแต่ละคลาส
- x คือจำนวนของข้อมูลฝึกสอนตั้งต้น
- Y คือจำนวนข้อมูลฝึกสอนของคลาส m ในแต่ละเซต

โดยหลังจากที่ได้จำนวนฝึกสอนของแต่ละคลาสแล้วนั้นจะทำการเลือกข้อมูลสำหรับแต่ละคลาสโดยจะเลือกจากชุดฝึกสอนที่ต้องการแบ่งการทำงาน ซึ่งจะเลือกให้มีการกระจายข้อมูลมากที่สุด (Maximum Distribution) และมีการทับซ้อนกันน้อยที่สุด (Minimum Overlap) สำหรับข้อมูลระหว่างเซตใด ๆ