บทที่ 2

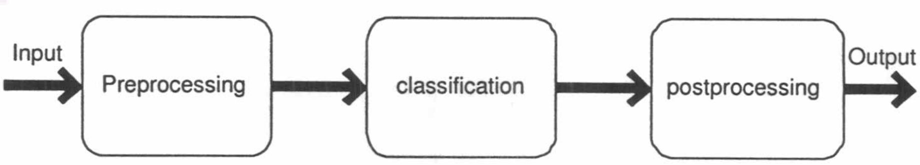
ทฤษฎีและแนวคิด

การรู้จำแบบรูปมีรูปแบบทั่วไปดังแสดงในรูปที่ 2.1 วัตถุที่จะทำการตรวจสอบจะถูกทำการวัดโดยใช้เซ็นเซอร์ซึ่งเป็นการวัดเชิงปริมาณ ข้อมูลที่ได้จะผ่านกระบวนการก่อนหน้า (Preprocessing) เพื่อให้ข้อมูลเหมาะสมสำหรับการใช้งานมากขึ้น จากนั้นนำข้อมูลที่ได้ผ่านเข้าสู่การสกัดค่าคุณลักษณะสำคัญ (Feature extraction) เพื่อหาคุณลักษณะสำคัญ (Feature) ของข้อมูลสำหรับใช้ในการตัดสินใจในการจำแนกแบบรูป และผ่านกระบวนการภายหลัง (Postprocessing) เพื่อปรับผลการตัดสินใจตามเงื่อนไขเพิ่มเติมต่าง ๆ เพื่อให้ได้ความถูกต้องมากขึ้น ในบทนี้จะกล่าวถึงกระบวนการก่อนหน้าที่ใช้กับการรู้จำตัวอักษรแบบออนไลน์ และทฤษฎีของระบบการรู้จำ 2 ชนิด คือ แบบจำลองฮิดเดนมาร์คอฟและฟัซซีโลจิก

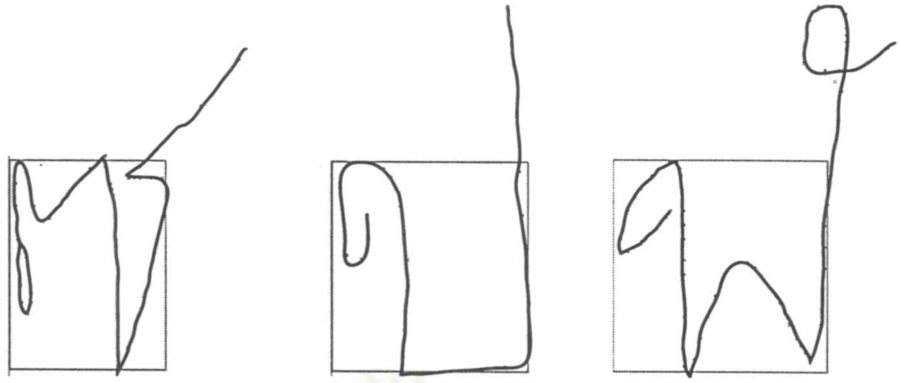
2.1 กระบวนการก่อนหน้า

ข้อมูลของตัวอักษรแบบออนไลน์เป็นจุดคู่ลำดับ (x, y) บนจอภาพ ได้มาจากกระดานอิเล็กทรอนิกส์ (tablet) หลังจากนั้นทำการแยกองค์ประกอบเป็นส่วนย่อยโดยใช้การขึ้นลงของปากกา และระยะเวลาระหว่าง การขึ้น และลงจากนั้นจะทำการลดจุดคู่ลำดับส่วนเกิน (Redundant) โดยใช้ขั้นตอนการหาใน Nair และ Leedham [23] ซึ่งมีขั้นตอนดังต่อไปนี้

1. ลบการยกปากกาโดยบังเอิญ
2. ทำเส้นให้เรียบ (Smoothing)
3. ทำการกรองระยะทางที่สั้น (Minimum Distance Filtering)
4. ลบ ส่วนที่เกิดขึ้นโดยไม่ตั้งใจเวลาลงปากกา หรือยกปากกาขึ้น (Serif Removal)
5. ทำนอร์มอลไลซ์ (Normalization)



รูปที่ 2.1 แผนภาพการทำงานของระบบรู้จำแบบรูป



รูปที่ 2.2 ตัวอย่างตัวอักษรที่มีหาง และกรอบตัวอักษรที่ไม่รวมหาง

สำหรับการทำนอร์มอลไลซ์ในภาษาไทยนั้นจะพบตัวอักษรทีมีหางเกินเส้นบรรทัด การขยายขนาดตัวอักษรให้มีความสูงเท่ากันนั้นควรจะสนใจความสูงของตัวอักษรที่ไม่รวมส่วนหาง ดังนั้นเราจึงหากรอบของตัวอักษรเดี่ยวโดยที่หางจะไม่อยู่ในกรอบ เมื่อเราพบว่าตัวอักษรมีโอกาสเป็นหางหรือจุดปลายของตัวอักษรอยู่ด้านบนขวา เราจะพิจารณากรอบด้านบนให้สูงเท่ากับจุดที่สูงที่สุดในส่วนครึ่งด้านซ้ายของตัวอักษรแทน ตัวอย่างการหากรอบตัวอักษรแสดงในรูป 2.2

2.2 แบบจำลองฮิดเดนมาร์คอฟ

เมื่อพิจารณาถึงแบบจำลองมาร์คอฟระบบหนึ่ง ซึ่งไม่สามารถระบุได้ชัดเจนว่า ในขณะหนึ่งระบบอยู่ในสถานะใด หากแต่เราสามารถสังเกตปรากฏการณ์อื่นๆ เพื่อนำมาทำนายความน่าจะเป็นของสถานะในแบบจำลองมาร์คอฟที่เวลาใดๆ ได้ เรียกแบบจำลองดังกล่าวเป็นแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model) [4] โดยเราจะพิจารณาถึง แบบจำลองฮิดเดนมาร์คอฟแบบไม่ต่อเนื่อง (Discrete HMM) ก่อนเพื่อทำเข้าใจในหลักการทำงาน แล้วจะกล่าวถึงแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง (Continuous HMM) ในภายหลัง

2.2.1 องค์ประกอบของแบบจำลองฮิดเดนมาร์คอฟแบบไม่ต่อเนื่อง

1. จำนวนสถานะในแบบจำลอง (N) และกำหนดตัวแปรแทนสถานะในแบบจำลองเป็น $S = \{S_1, S_2, \ldots, S_N\}$

2. จำนวนของสัญลักษณ์จากการสังเกตต่อสถานะ (M) และกำหนดตัวแปรแทนสัญลักษณ์เป็น $V = \{v_1, v_2, \ldots, v_M\}$

3. การกระจายความน่าจะเป็นของการเปลี่ยนสถานะ $A = \{a_{ij}\}$ โดย

$$a_{ij} = P[q_1 = S_j | q_{t-1} = S_i], \qquad 1 \le i, j \le N \tag{2.1}$$

4. การกระจายความน่าจะเป็นของสัญลักษณ์ที่สังเกตเมื่อทราบสถานะ j, $B = \{b_j(k)\}$ โดย

$$b_j(k) = P[v_k \text{ at } t | q_t = S_j], \qquad 1 \le j \le N, 1 \le k \le M \tag{2.2}$$

5. การกระจายความน่าจะเป็นของสถานะเริ่มต้น $\pi = \{\pi_i\}$ เมื่อ

$$\pi_i = P[q_1 = S_j], \qquad 1 \le i \le N \tag{2.3}$$

การอธิบายแบบจำลองฮิดเดนมาร์คอฟต้องระบุจำนวนสถานะ N และจำนวนสัญลักษณ์ M ของแบบจำลอง สามารถเขียนองค์ประกอบของความน่าจะเป็นรวมกันแทนแบบจำลองได้ดังนี้

$$\lambda = (A, B, \pi) \tag{2.4}$$

2.2.2 การใช้แบบจำลองฮิดเดนมาร์คอฟในการรู้จำ

แบบจำลองฮิดเดนมาร์คอฟนิยมนำมาประยุกต์ใช้ในการรู้จำแบบรูป [24] โดยในการรู้จำแบบรูปที่ต้องการแยกประเภทของสิ่งที่เราสังเกตออกเป็น n ชนิด แบ่งเป็น λ_1, λ_2, λ_3, …, λ_n จากลำดับของค่าสังเกต O สิ่งที่เราต้องการจะรู้คือความน่าจะเป็นของสิ่งที่เราสังเกตจะเป็นชนิด i, $P(\lambda_i|O)$ โดยตัดสินใจว่าสิ่งที่เราสนใจเป็นชนิดที่ i เมื่อ

$$P(\lambda_i|O) \ge P(\lambda_j|O), \qquad 1 \le i, j \le n \tag{2.5}$$

จากกฎของเบย์ (Baye's Rule) จะได้

$$P(\lambda_i|O) = \frac{P(O|\lambda_i)P(\lambda_i)}{P(O)} \tag{2.6}$$

หากตั้งสมมติฐานว่าทุกประเภทที่จะจำแนกมีโอกาสเกิดขึ้นเท่ากัน จะพบว่า

$$P(\lambda_i|O) \propto P(O|\lambda_i) \tag{2.7}$$

ดังนั้นสิ่งที่อยากทราบอันดับแรกคือ $P(O|\lambda_i)$ เมื่อทราบแบบจำลองฮิดเดนมาร์คอฟแล้ว สิ่งที่น่าสนใจอันดับต่อไปคือลำดับของสถานะที่แท้จริงที่เกิดขึ้นตั้งแต่เวลา 1 ถึง เวลา T ($Q = q_1 q_2 \ldots q_T$) เมื่อเราทราบลำดับของค่าสังเกตในช่วงเวลานั้น ($O = O_1 O_2 \ldots O_T$) เนื่องจากเป็นส่วนที่ซ่อนอยู่ในแบบจำลอง อันดับสุดท้ายคือจะสร้างแบบจำลองอย่างไรเพื่อทำให้ $P(O|\lambda_i)$ มีค่ามากที่สุด ในส่วนต่อไปจะเสนอการแก้ปัญหาทั้ง 3 ประการที่ได้กล่าวมาข้างต้น

2.2.2.1 การคำนวณหาความน่าจะเป็นของลำดับของค่าสังเกตเมื่อทราบแบบจำลองแล้ว $P(O|\lambda)$

เมื่อทราบลำดับของค่าสังเกตและแบบจำลอง $O = O_1O_2 \ldots O_T$ และ λ ตามลำดับ และกำหนดให้มีลำดับของสถานะดังนี้

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(O_t|q_t,\lambda) \tag{2.8}$$

หากเราพิจารณาในกรณีที่ชุดสังเกตแต่ละชุดเป็นอิสระต่อกันทางสถิติจะได้

$$P(O|Q,\lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdots b_{q_T}(O_T) \tag{2.9}$$

และ

$$P(Q|\lambda) = \pi_{q_1} \cdot a_{q_1q_2} \cdot a_{q_2q_3} \cdots a_{q_{T-1}q_T} \tag{2.10}$$

จาก

$$P(O,Q|\lambda) = P(O|Q,\lambda)P(Q|\lambda) \tag{2.11}$$

จะพบว่า

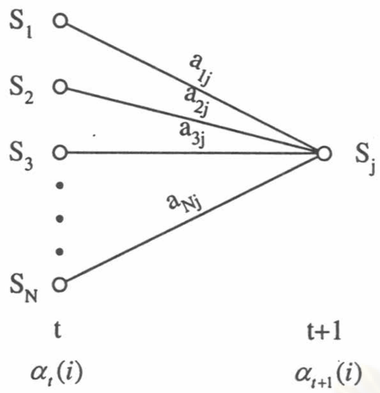
$$\begin{aligned} P(O|\lambda) &= \sum_{\text{all } Q} P(O|Q,\lambda)P(Q|\lambda) \\ &= \sum_{q_1,q_2,\ldots,q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1q_2} b_{q_2}(O_2) \cdots a_{q_{T-1}q_T} b_{q_T}(O_T) \end{aligned} \tag{2.12}$$

เมื่อพิจารณาสมการที่ (2.12) จะพบว่าจะต้องการคูณกันถึง $(2T-1)N^T$ ครั้ง ดังนั้นจึงไม่นิยมนำสมการ (2.12) ไปใช้กันในทางปฏิบัติโดยตรง เนื่องจากมีกระบวนการที่ทำงานได้เร็วกว่าคือ กระบวนการไปข้างหน้าและย้อนกลับ (Forward-Backward Procedure)

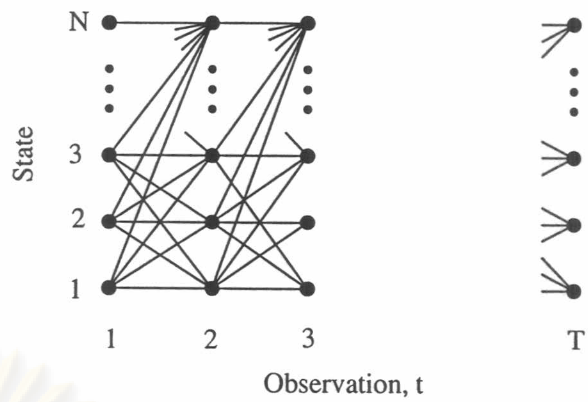
กำหนดตัวแปรไปข้างหน้า (Forward Variable: $\alpha_t(i)$) ดังนี้

$$\alpha_t(i) = P(O_1O_2 \ldots O_t, q_t = S_i|\lambda) \tag{2.13}$$

กระบวนการไปข้างหน้าทำได้ดังนี้



(ก)



(ข)

รูปที่ 2.3 (ก) แผนภูมิลำดับการคำนวณตัวแปรไปข้างหน้า (ข) แผนภูมิในการสร้างการคำนวณ

1. เริ่มต้นด้วย

$$\alpha_1(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N \tag{2.14}$$

2. คำนวณหาค่าถัดไปตามสมการ

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}), \qquad 1 \le t \le T-1, 1 \le j \le N \tag{2.15}$$

3. จะได้ผลลัพธ์สุดท้ายคือ

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{2.16}$$

จะพบว่าการคำนวณตามกระบวนการข้างต้นใช้การคำนวณทั้งหมดเพียง $N^2(T-1)$ ครั้ง โดยการคำนวณสามารถแสดงได้เป็นโครงสร้างผลึกตามรูปที่ 2.3 และในทำนองเดียวกันกับกระบวนการไปข้างหน้า กระบวนการย้อนกลับจะกำหนดตัวแปรย้อนกลับ (Backward Variable: $\beta_t(i)$) ดังนี้

$$\beta_t(i) = P(O_{t+1}O_{t+2}\ldots O_T|q_t = S_i, \lambda) \tag{2.17}$$

1. ตั้งค่าเริ่มต้น

$$\beta_T(i) = 1 \tag{2.18}$$

2. หาค่าตัวถัดไปได้จาก

$$\beta_t(i) = \sum_{i=1}^{N} a_{ij} b_i(O_{t+1}) \beta_{t+1}(j), \qquad t = T+1, T-2, \ldots, 1; 1 \le i \le N \tag{2.19}$$

2.2.2.2 การหาลำดับของสถานะที่เหมาะสม

การเลือกสถานะที่เหมาะสมควรเลือกสถานะที่น่าจะเป็นไปได้ในทางสถิติมากที่สุด โดยจะตั้งตัวแปรขึ้นหนึ่งตัวซึ่งแสดงถึงความน่าจะเป็นที่จะเกิดสถานะ i ที่เวลา t เมื่อข้อมูลที่สังเกตได้คือ O ดังนี้

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \tag{2.20}$$

หรือจะเขียนได้ในอีกรูปหนึ่งเป็น

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{2.21}$$

และมีคุณสมบัติ

$$\sum_{i=1}^{N} \gamma_t(i) = 1 \tag{2.22}$$

เราจะตัดสินว่าที่เวลา t ระบบจะมีสถานะใดโดยเราจะใช้กฎในการตัดสินใจดังนี้

$$q_t = \arg\max[\gamma_t(i)], \qquad 1 \leq t \leq T \tag{2.23}$$

การเลือกสถานะโดยใช้กฎตามสมการที่ (2.23) พบว่าในบางกรณีที่การเปลี่ยนจากสถานะหนึ่งไปอีกสถานะหนึ่งอาจเป็นไปไม่ได้ที่จะเกิดขึ้น ($a_{ij} = 0$ สำหรับบาง i หรือj) ดังนั้นสถานะที่เลือกอาจไม่ใช่สถานะที่ถูกต้องเสมอไป ซึ่งแก้ปัญหานี้โดยใช้อัลกอริทึมของ Viterbi

ต้องการหาลำดับของสถานะที่ดีที่สุด $Q = \{q_1 q_2 \ldots q_T\}$ โดยให้

$$\delta_t(i) = \max_{q_1 q_2 \ldots q_{t-1}} P[q_1 q_2 \ldots q_t = i, O_1 O_2 \ldots O_t | \lambda] \tag{2.24}$$

พบว่า

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] b_j(O_{t+1}) \tag{2.25}$$

การหาลำดับของสถานะที่ดีที่สุดตามขั้นตอนดังต่อไปนี้

เริ่มต้นโดย

$$\delta_1(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N \tag{2.26}$$

ทำซ้ำตามสมการด้านล่างไปทุกๆ ค่า t

$$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_i(O_t), \qquad 2 \le t \le T, 1 \le j \le N \tag{2.27}$$

$$\beta_t(j) = \arg \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}], \qquad 2 \le t \le T, 1 \le j \le N \tag{2.28}$$

จะได้ว่า

$$P^* = \max_{1 \le i \le N} [\delta_T(i)] \tag{2.29}$$

$$Q^* = \arg \max_{1 \le i \le N} [\delta_T(i)] \tag{2.30}$$

2.2.2.3 การหาพารามิเตอร์ของแบบจำลองที่เหมาะสม

การจะหาค่าพารามิเตอร์ที่เหมาะสมนั้นมีหลายกระบวนการ โดยในวิทยานิพนธ์นี้จะเสนอกระบวนการที่เข้าใจง่าย และเป็นที่นิยมใช้กระบวนการหนึ่งคือ การประมาณค่าซ้ำของ Baum และ Welch เนื่องจากมั่นใจได้ว่าสามารถสร้างแบบจำลองที่เหมาะสมได้เมื่อใช้การประมาณค่าซ้ำไปจนกระทั่งได้แบบจำลองที่ดีที่สุด

เพื่ออธิบายกระบวนการนี้เราจะกำหนด $\xi_t(i,j)$ แสดงถึงความน่าจะเป็นที่จะเกิดสถานะ S_i ที่เวลา t และเกิดสถานะ S_j ที่เวลา $t+1$ เมื่อทราบแบบจำลองและลำดับของค่าสังเกต

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \tag{2.31}$$

จะพบว่า $\xi_t(i,j)$ มีความสัมพันธ์กับตัวแปรไปข้างหน้า ($\alpha_t(i)$) และตัวแปรย้อนกลับ ($\beta_t(j)$) ดังนี้

$$\begin{aligned} \xi_t(i,j) &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \end{aligned} \tag{2.32}$$

จากในหัวข้อที่แล้ว $\gamma_t(i)$ ซึ่งแสดงถึงความน่าจะเป็นที่จะเกิดสถานะ i ที่เวลา t มีความสัมพันธ์กับ $\xi_t(i,j)$ ดังนี้

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \tag{2.33}$$

เมื่อรวม $\gamma_t(i)$ ทุกๆ เวลา t จะได้ปริมาณที่แสดงถึงค่าคาดหวัง (Expected value) ของจำนวนครั้งที่เกิดสถานะ S_i และการรวม $\xi_t(i,j)$ ทุกๆ เวลา t หมายถึงค่าคาดหวังของจำนวนครั้งที่เปลี่ยนสถานะจาก S_i ไปยังสถานะ S_j โดยจะสรุปได้ดังนี้

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{expected number of transitions from } S_i \tag{2.34ก}$$

$$\sum_{t=1}^{T-1} \xi_t(i,j) = \text{expected number of transitions from } S_i \text{ to } S_j \tag{2.34ข}$$

จากสมการดังกล่าวจะสามารถประมาณค่าของพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟได้ดังนี้

$$\overline{\pi}_i = \text{expected frequency in state } S_i \text{ at time } (t=1) \; = \gamma_1(i) \tag{2.35ก}$$

$$\overline{a}_{ij} = \frac{\text{expected number of transitions from } S_i \text{ to } S_j}{\text{expected number of transitions from } S_i} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{2.35ข}$$

$$\overline{b}_j(k) = \frac{\text{expected number of times in state } S_j \text{ and observing symbol } v_k}{\text{expected number of times in state } S_j} = \frac{\sum_{\substack{t=1 \\ \text{s.t.} O_k = v_k}}^{T-1} \gamma_t(j)}{\sum_{t=1}^{T-1} \gamma_t(j)} \tag{2.35ค}$$

เมื่อกำหนดให้สถานะปัจจุบันเป็น $\lambda = (A, B, \pi)$ และคำนวณตามสมการ (2.35ก-2.35ค) จะได้แบบจำลองที่เกิดจากการประมาณค่าเป็น $\overline{\lambda} = (\overline{A}, \overline{B}, \overline{\pi})$ และหากนำแบบจำลองที่ได้มาใหม่ $\overline{\lambda}$ ไปแทนที่ λ และนำไปคำนวณตามสมการ (2.35ก-2.35ค) ซ้ำไปเรื่อยๆ จากการพิสูจน์ของ Baum และทีมงานพบว่าแบบจำลองที่ประมาณค่าได้ยิ่งคำนวณซ้ำหลายรอบ จะมีค่ามากขึ้นจนเข้าใกล้จุดลิมิตค่าหนึ่งซึ่งเป็นจุดที่ $P(O|\lambda)$ จะมีค่ามากที่สุด

2.2.3 แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง

จากแบบจำลองฮิดเดนมาร์คอฟแบบไม่ต่อเนื่องที่ได้กล่าวมาข้างต้นพิจารณาเพียงสัญลักษณ์แบบไม่ต่อเนื่องที่เลือกมาจากชุดสัญลักษณ์ที่มีจำนวนจำกัด ในกรณีที่ข้อมูลสังเกตเป็นสัญญาณแบบต่อเนื่องจะนำข้อมูลนั้นมาควอนไทซ์ เพื่อให้ได้รูปแบบของสัญลักษณ์ที่จำกัด ในบางครั้งการควอนไทซ์ นั้นทำให้

เกิดความผิดพลาดในการแทนสัญลักษณ์ของข้อมูลสังเกตได้ ดังนั้นการใช้แบบจำลองฮิดเดนมาร์คอฟที่มีความหนาแน่นของข้อมูลสังเกตแบบต่อเนื่องจึงให้ผลลัพธ์ที่ดีกว่า

การนำความหนาแน่นของชุดข้อมูลสังเกตแบบต่อเนื่องมาใช้งาน โดยใช้รูปแบบของฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function, pdf) จะต้องมีพารามิเตอร์ที่สามารถประมาณค่าซ้ำได้ตามวิธีการที่ได้กล่าวไปแล้ว รูปแบบของฟังก์ชันความหนาแน่นความน่าจะเป็นที่นิยมใช้คือฟังก์ชันความหนาแน่นผสมที่ประกอบไปด้วยฟังก์ชันความหนาแน่นย่อยชนิดเดียวกันจำนวนจำกัด ดังแสดงได้ตามสมการต่อไปนี้

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{jm}\mathfrak{N}[\mathbf{O}, \boldsymbol{\mu}_{jm}, \mathbf{U}_{jm}], \qquad 1 \leq j \leq N \tag{2.36}$$

เมื่อ $\mathbf{O}$ คือเวกเตอร์สังเกตที่จะนำมาจำลองแบบ c_{jm} คือ สัมประสิทธิ์ความหนาแน่นย่อยของฟังก์ชันความหนาแน่นผสมตัวที่ m ในสถานะ j และ $\mathfrak{N}$ คือฟังก์ชันความหนาแน่นย่อย ซึ่งนิยมใช้ความหนาแน่นแบบเกาส์ โดยจะต้องมี c_{jm} ที่ตรงตามข้อกำหนดของกระบวนการเฟ้นสุ่ม (stochastic) ดังนี้

$$\sum_{m=1}^{M} c_{jm} = 1, \qquad 1 \leq j \leq N \tag{2.37ก}$$

$$c_{jm} \geq 0, \qquad 1 \leq j \leq N, 1 \leq m \leq M \tag{2.37ข}$$

และฟังก์ชันความหนาแน่นความน่าจะเป็นจะถูกนอร์แมลไลซ์ ให้เป็นดังนี้

$$\int_{-\infty}^{\infty} b_j(\mathbf{x})d\mathbf{x} = 1, \qquad 1 \leq j \leq N \tag{2.38}$$

สำหรับการประมาณค่าซ้ำเพื่อหาพารามิเตอร์ของฟังก์ชันความหนาแน่นผสม c_{jm} $\boldsymbol{\mu}_{jk}$ และ $\mathbf{U}_{jk}$ เป็นไปตามสมการเหล่านี้

$$\bar{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T}\sum_{k=1}^{M} \gamma_t(j,k)} \tag{2.39}$$

$$\bar{\boldsymbol{\mu}}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot \mathbf{O}_t}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{2.40}$$

$$\overline{\mathbf{U}}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot (\mathbf{O}_t - \boldsymbol{\mu}_{jk})(\mathbf{O}_t - \boldsymbol{\mu}_{jk})^{\mathbf{T}}}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{2.41}$$

เมื่อ $\gamma_t(j,k)$ คือ ความน่าจะเป็นที่จะอยู่ในสถานะ j ที่เวลา t ด้วย ฟังก์ชันความหนาแน่นผสมที่อธิบาย $\mathbf{O}_t$

$$\gamma_t(j,k) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)}\right]\left[\frac{c_{jk}\mathfrak{N}(\mathbf{O}_t, \boldsymbol{\mu}_{jk}, \mathbf{U}_{jk})}{\sum_{m=1}^{M} c_{jm}\mathfrak{N}(\mathbf{O}_t, \boldsymbol{\mu}_{jm}, \mathbf{U}_{jm})}\right] \tag{2.42}$$

สำหรับการประมาณค่าซ้ำของ a_{ij} นั้นเหมือนกับในกรณีไม่ต่อเนื่อง

2.3 ฟัซซีโลจิก

ทฤษฎีทางฟัซซีโลจิกได้ถูกเสนอขึ้นโดย Lofti Zadeh ในปี 1965 เป็นการใช้ทฤษฎีของเซตมาประยุกต์ เพื่อแสดงลักษณะที่คลุมเครือของสิ่งต่างๆ ที่มีความไม่แน่นอน หรือมีความซับซ้อน ซึ่งยากต่อการสร้างแบบจำลองทางคณิตศาสตร์ ฟัซซีโลจิกถูกนำมาเพื่อสร้างแบบจำลองที่มีความใกล้เคียงความเป็นจริง โดยมีการพิจารณาลักษณะของปัญหาที่สนใจ แล้วหาค่าความเป็นสมาชิกเพื่อเป็นตัวแปรที่นำมาตัดสินใจ [25][26] ซึ่งการประยุกต์ใช้ฟัซซีโลจิกทำให้ระบบสามารถบ่งบอกถึงความไม่แน่นอน และเมื่อมีการนำมาใช้ร่วมกับการตัดสินใจทำให้มีความยืดหยุ่นในการตัดสินใจมากยิ่งขึ้น นอกจากนี้ได้มีการนำเอาวิธีทางฟัซซีโลจิกมาประยุกต์ใช้กับงานหลาย ๆ ด้านเช่น ระบบควบคุม ระบบทางชีววิทยา และระบบรู้จำ

2.3.1 นิยามฟัซซีเซต

กำหนดให้ X เป็นกลุ่มของวัตถุซึ่งแทนด้วยเซต ซึ่งเป็นได้ทั้งแบบต่อเนื่องและไม่ต่อเนื่อง ดังนั้นฟัซซีเซต F ใน X ถูกแทนด้วยเซตของคู่ลำดับ

$$F = \{(x, \mu_x(x)) | x \in X\} \tag{2.43}$$

โดย $\mu_F(x)$ เป็นฟังก์ชันความเป็นสมาชิก หรือ ระดับความเป็นสมาชิก ของสมาชิก x ในฟัซซีเซต F และ ในฟัซซีเซต F ในช่วงค่าจำนวนจริงระหว่าง $[0, 1]$

2.3.2 ฟังก์ชันความเป็นสมาชิก

ฟังก์ชันความเป็นสมาชิกในฟัซซีเซต คือฟังก์ชันที่บ่งบอกถึงระดับของคุณลักษณะ โดยมีค่าเป็นค่าจริงอยู่ระหว่าง 0 กับ 1 ฟังก์ชันความเป็นสมาชิกเป็นได้ทั้งฟังก์ชันต่อเนื่อง และไม่ต่อเนื่องฟังก์ชันความเป็นสมาชิกที่ใช้กันทั่วไปมีหลายรูปทรงไม่ว่าจะเป็น สามเหลี่ยม (Triangular shape) รูปแบบเอส และพาย (S- and π- shape)

2.3.3 ตัวดำเนินการฟัซซี

ตัวดำเนินการพื้นฐานของฟัซซีเซทมีอยู่ 3 อย่างคือ ยูเนียน (union) อินเตอร์เซคชัน (intersection) และ คอมพลีเมนท์ (complement)

2.3.3.1 ตัวดำเนินการยูเนียน (Union Operator)

ฟังก์ชันความเป็นสมาชิก $\mu_{A\cup B}(x)$ ถูกกำหนดสำหรับทุก ๆ x ที่เป็นสมาชิกของ X โดย

$$\mu_{A\cup B}(x) = \max\{\mu_A(x), \mu_B(x)\} \tag{2.44}$$

2.3.3.2 ตัวดำเนินการอินเตอร์เซคชัน (Intersection Operator)

ฟังก์ชันความเป็นสมาชิก $\mu_{A\cap B}(x)$ ถูกกำหนดสำหรับทุก ๆ x ที่เป็นสมาชิกของ X โดย

$$\mu_{A\cap B}(x) = \min\{\mu_A(x), \mu_B(x)\} \tag{2.45}$$

2.3.3.3 ตัวดำเนินการคอมพลีเมนท์ (Complement Operator)

ฟังก์ชันความเป็นสมาชิก $\mu_{\overline{A}}(x)$ ถูกกำหนดสำหรับทุก ๆ x ที่เป็นสมาชิกของ X โดย

$$\mu_{\overline{A}}(x) = 1 - \mu_A(x) \tag{2.46}$$

2.3.4 ตัวแปรฟัซซีเชิงภาษา

ส่วนที่สำคัญในแบบจำลองฟัซซี คือการกำหนดตัวแปรเชิงภาษา ในวิธีเชิงภาษาจะใช้คำแทนค่าของตัวเลขเพื่ออธิบายเหตุการณ์ที่ซับซ้อนเกินไป หรือยากที่จะกำหนดคุณลักษณะในเชิงปริมาณ ในการเปรียบเทียบหน่วย ทำได้โดยใช้ฟัซซีเซตมากกว่าหนึ่งตัว เช่น “มาก” “ค่อนข้างมาก” “ค่อนข้างน้อย” “น้อย” คุณสมบัติตัวอย่าง เช่น ความกลม สามารถวัดในลักษณะ “กลมเล็กน้อย” “กลมมาก” “กลมโดยสมบูรณ์”

2.3.5 การประมาณเหตุผลด้วยฟัซซี

ฟังก์ชันความเป็นสมาชิกจะเชื่อมโยงตัวแปรค่าจริงไปยังตัวแปรเชิงภาษา ดังนั้นข้อมูลสามารถประมาณสรุปได้ ในระบบฟัซซีความไม่แน่นอนนี้ถูกเรียกว่าความเป็นฟัซซี เพราะความไม่แน่นอนทำให้ไม่สามารถกำหนดกลุ่มสมาชิกอย่างเด็ดขาด ในการตัดสินใจด้วยกฎทางฟัซซี ทำได้โดยใช้การแสดงเงื่อนไข “ถ้า . . . แล้ว . . .” “ถ้า x คือ A แล้ว y คือ B” สามารถเขียนเป็นสัญลักษณ์ได้เป็น “$A \rightarrow B$” โดยที่ A และ B เป็นฟัซซีเซตในเอกภพสัมพัทธ์ (Universe) X ด้วยฟังก์ชันความเป็นสมาชิก μ_A และ μ_B

2.3.6 การสร้างชุดกฎทางฟัซซี

การสร้างชุดกฎทางฟัซซีทำได้โดยพิจารณาจากโครงสร้างของตัวอักษร ผลลัพธ์ที่ได้จากการหาคุณลักษณะ และคำนวณค่าความเป็นสมาชิกของคุณลักษณะเหล่านั้นในตัวอักษรแต่ละตัวจะมีลักษณะที่เป็นโครงสร้างเฉพาะที่ไม่ซ้ำกัน เช่น ‘ด’ จะมีหัวอยู่ตรงกลางตัวในทิศทางตามเข็มนาฬิกา มีจุดปลายอยู่ทางด้านขวาล่างชี้ลงด้านล่าง และด้านบนไม่มีรอยหยัก หรือ ‘ม’ มีหัวทางด้านมุมซ้ายบนในทิศทางตามเข็มนาฬิกา มีจุดตัดกันของเส้นอยู่ทางด้านซ้ายล่าง และมีจุดปลายทางมุมขวาด้านบนชี้ขึ้นด้านบน จากนั้นนำค่าความเป็นสมาชิกของคุณลักษณะมาพิจารณาความเบี่ยงเบนและความหลากหลายของรูปแบบของลายมือเขียน ซึ่งจะแสดงรายละเอียดการสร้างชุดกฎทางฟัซซีในบทต่อไป