บทที่ 3

ระบบพิสูจน์ลายมือชื่อโดยกระบวนการทางคณิตศาสตร์

แนวคิดของการพัฒนาระบบพิสูจน์ลายมือชื่อ คือ การหาตัวแทนทางคณิตศาสตร์ที่เหมาะสม เพื่ออธิบายแต่ละเอกลักษณ์ที่ประกอบอยู่ในลายมือชื่อ แล้วใช้กระบวนการตัดสินใจทางสถิติเพื่อทดสอบความมีนัยสำคัญ ระหว่างเอกลักษณ์ของลายมือชื่อที่ต้องการพิสูจน์ กับขอบเขตความเชื่อมั่นของเอกลักษณ์นั้นๆจากลายมือชื่อที่เก็บรวบรวมมา โดยมีข้อสมมติฐานของการเขียนลายมือชื่อที่ผู้เชี่ยวชาญใช้ในการพิจารณา คือ

1) ผู้เป็นเจ้าของลายมือชื่อ จะมีลักษณะเฉพาะที่เป็นของตนในการเขียน

2) การเขียนลายมือชื่อแต่ละครั้ง เจ้าของลายมือชื่อจะเขียนโดยยึดถือตามเอกลักษณ์ของตน ถ้ามีการเปลี่ยนแปลงจะมีการเปลี่ยนแปลงที่น้อยมาก

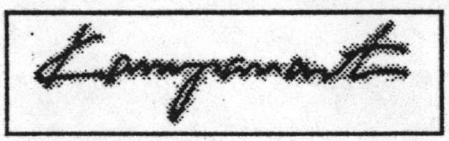
3) ลายมือชื่อที่ผ่านการปลอมแปลง ไม่สามารถรักษาเอกลักษณ์ของลายมือชื่อที่แท้จริงได้อย่างครบถ้วน

การหาตัวแทนทางคณิตศาสตร์ เพื่ออธิบายแต่ละเอกลักษณ์ของลายมือชื่อ จำเป็นต้องกำหนดคำจำกัดความของลายมือชื่อ เพื่อที่จะสามารถใช้กรรมวิธีทางคณิตศาสตร์ ในการหาตัวแทนของแต่ละเอกลักษณ์ที่ประกอบลายมือชื่อ ได้ดังนี้

1) ลายมือชื่อใดๆ คือ ภาพๆหนึ่ง ซึ่งมีขนาดความกว้าง และความยาวตามขอบเขตที่กำหนด

2) แต่ละจุดบนลายมือชื่อ คือ ค่าระดับความเข้ม (Intensity) ในตำแหน่งนั้นๆบนลายมือชื่อ

ดังนั้นจึงสามารถอธิบายลายมือชื่อ ด้วยความสัมพันธ์ทางคณิตศาสตร์ได้ดังนี้



ภาพของลายมือชื่อ กำหนดให้มีความสัมพันธ์ = F (x,y) ซึ่งคือค่าความเข้ม ในตำแหน่ง x และ y ใดๆบนลายมือชื่อ โดยที่ $x_{min} \leq x \leq x_{max}$ และ $y_{min} \leq y \leq y_{max}$

การพิสูจน์ลายมือชื่อ ทำได้โดยการหาตัวแทนทางคณิตศาสตร์ที่สามารถอธิบายเอกลักษณ์ ที่ประกอบอยู่ในลายมือชื่อ โดยที่เอกลักษณ์ของลายมือชื่อดังที่แสดงในบทที่ 2 ประกอบด้วย

1) ภาพโดยรวมของลายมือชื่อ
2) ค่าเฉลี่ยความเข้มของลายเส้น
3) ค่าความยาวโดยเปรียบเทียบของลายมือชื่อ
4) การแบ่งกลุ่มของลายมือชื่อ

การหาตัวแทนทางคณิตศาสตร์ ที่สามารถอธิบายเอกลักษณ์ของลายมือชื่อทำได้ดังนี้

3.1 <u>การหาตัวแทนของภาพโดยรวมของลายมือชื่อ</u>

ภาพโดยรวมของลายมือชื่อ คือ การพิจารณาความสัมพันธ์ระหว่างจุดใดจุดหนึ่งบนลายมือชื่อ กับจุดทั้งหมดของลายมือชื่อ ลายมือชื่อที่แตกต่างกันย่อมที่จะมีความสัมพันธ์ ระหว่างจุดใดจุดหนึ่งกับจุดทั้งหมดภายในลายมือชื่อนั้นๆที่แตกต่างกัน กรรมวิธีทางคณิตศาสตร์ที่สามารถอธิบายความสัมพันธ์ระหว่างข้อมูลตัวใดตัวหนึ่ง กับข้อมูลทั้งหมด คือ การแก้ปัญหาด้วยค่าเจาะจง (Eigensolution)

กำหนดให้ F(x,y) แทนความสัมพันธ์ของจุดใดๆบนลายมือชื่อ จะได้ว่า

$F(x,y).\ F(z) = \lambda.\ F(z)$

โดยที่ F(z) คือ ความสัมพันธ์ที่ไม่เป็นศูนย์ (non Zero function) และเรียก F(z) ว่า ความสัมพันธ์ของตัวถูกกระทำ ในที่นี้ λ คือ ค่าคงตัว และเรียกค่าคงตัวนี้ว่า ตัวกระทำการ หรือ ค่าเจาะจง (Eigen value) โดยที่ $\lambda \varepsilon (\lambda_1, \lambda_2, \ldots, \lambda_n)$

การใช้ค่า Eigen value เพื่อแทนภาพโดยรวมของลายมือชื่อ เนื่องจากค่า Eigen value ที่มากที่สุดของ F(x,y) จะอธิบายถึงเปอร์เซ็นต์ที่มากที่สุดที่แต่ละค่าของ F(x,y) จะสัมพันธ์กับทุกค่าของ F(x,y) (Jennings.A,1977) ในกรณีของลายมือชื่อคือ เปอร์เซ็นต์ที่มากที่สุดที่ความเข้มของจุดใดจุดหนึ่งบนลายมือชื่อ จะสัมพันธ์กับความเข้มทุกๆจุดบนลายมือชื่อ และจากการศึกษาเรื่องการรับรู้รูปแบบตัวพิมพ์อักษรไทย (Kimpan.C,1986) พบว่าการแก้ปัญหาด้วยค่าเจาะจงนี้สามารถใช้เป็นตัวแทนของข้อมูล เพื่อใช้ในกระบวนการจดจำได้เช่นกัน

3.2 <u>การหาตัวแทนของค่าเฉลี่ยความเข้มของลายเส้น</u>

ค่าเฉลี่ยความเข้มของลายเส้น คือ การหาค่าเฉลี่ยของความเข้มทุกๆจุดบนลายมือชื่อที่เป็นส่วนของลายเส้น ทำได้โดยการกำหนดช่วงของระดับความเข้มที่กำหนดเป็นลายเส้น (threshold) โดยที่ช่วงของระดับความเข้มจะอยู่ระหว่าง 0 ถึง 255 ซึ่งความเข้มระดับ 0 คือ สีดำ และ ความเข้มระดับ 255 คือ สีขาว เมื่อทำการตรวจสอบทุกๆจุดบนลายมือชื่อ ถ้าค่าความเข้มของจุดใดอยู่ในช่วงของการกำหนดเป็นลายเส้น จะทำการรวมค่าความเข้มของจุดนี้ เพื่อหาค่าเฉลี่ย เมื่อสิ้นสุดการตรวจสอบทุกๆจุดบนลายมือชื่อ

กำหนดให้ F(x,y) คือ ความสัมพันธ์ของความเข้ม ของจุดใด ๆบนลายมือชื่อ

I คือ ค่าเฉลี่ยความเข้มของลายเส้น

threshold คือ ช่วงของระดับความเข้มที่เป็นลายเส้น

n คือ จำนวนจุดที่เป็นลายเส้น

จะได้

$$I = \frac{\int_{y_{min}}^{y_{max}} \int_{x_{min}}^{x_{max}} F(x,y)\, dx\, dy}{n} \qquad \text{เมื่อ } F(x,y) \leq \text{threshold}$$

ค่าเฉลี่ยความเข้มของลายเส้น (I) สามารถเป็นตัวแทนทางคณิตศาสตร์ของค่าเฉลี่ยความเข้มของลายมือชื่อได้เช่นกัน

3.3 <u>การหาตัวแทนความยาวโดยเปรียบเทียบ</u>

ความยาวโดยเปรียบเทียบ คือ การหาสัดส่วนระหว่าง ความยาวของลายเส้นทั้งหมดกับความกว้างของลายมือชื่อ กรรมวิธีที่ผู้เชี่ยวชาญใช้ในการวัดความยาวของลายเส้น ทำได้โดยการใช้เส้นด้ายทาบทับไปบนลายเส้น แล้ววัดความยาวของเส้นด้ายทั้งหมด แต่วิธีนี้ไม่เป็นที่นิยมเนื่องจากใช้เวลานาน และลายมือชื่อโดยมากจะมีขนาดเล็ก เกิดความผิดพลาดของการทาบทับโดยง่าย บางครั้งผู้เชี่ยวชาญอาจไม่หาเอกลักษณ์นี้ ในการพัฒนาระบบพิสูจน์ลายมือชื่อด้วยคอมพิวเตอร์นี้ สามารถทำการวัดความยาวของลายเส้นได้อย่างรวดเร็ว และแม่นยำ

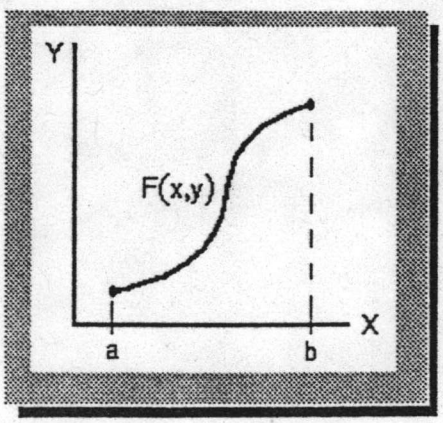
กำหนดให้ F(x,y) คือ ความสัมพันธ์ของความเข้ม ของจุดใด ๆบนลายมือชื่อ

L คือ ความยาวของลายเส้น

W คือ ความกว้างของลายมือชื่อ

threshold คือ ช่วงของระดับความเข้มที่เป็นลายเส้น

สามารถแสดงได้ดังภาพ 3.1



ภาพที่ 3.1 แสดงการหาความยาวของลายเส้น

ความยาวของลายเส้น(L) สามารถแสดงได้ดังนี้

$$L = \int_{y_{min}}^{y_{max}} \int_{x_{min}}^{x_{max}} F(x,y)\, dx\, dy \qquad \text{โดยที่ } F(x,y) \leq \text{threshold}$$

และความกว้างของลายมือชื่อ(W) สามารถเขียนได้คือ

$$W = (b - a) \qquad \text{โดยที่ } a = \text{Min } (x) \text{ เมื่อ } F(x,y) \neq F(x,y+1)$$

และ $b = \text{Max } (x)$ เมื่อ $F(x,y) \neq F(x,y+1)$

ความยาวโดยเปรียบเทียบ คือ สัดส่วนระหว่างความยาวของลายเส้น กับความกว้างของลายมือชื่อ จึงสามารถ แสดงได้ดังนี้

ความยาวโดยเปรียบเทียบ = L / W

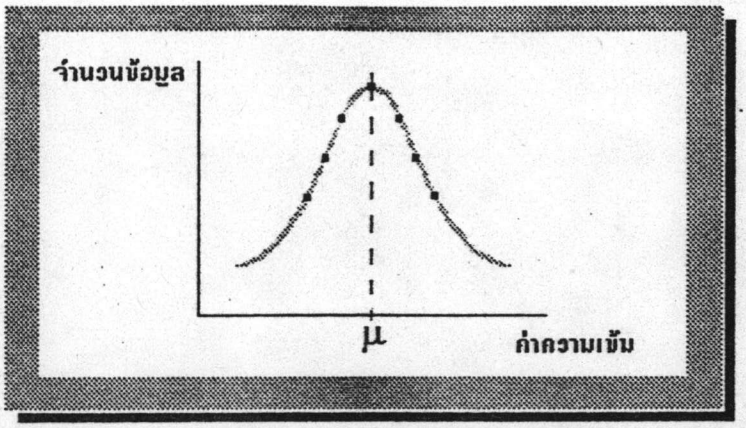
ดังนั้นจึงสามารถใช้ค่าทางคณิตศาสตร์นี้ เป็นตัวแทนของความยาวโดยเปรียบเทียบของลายมือชื่อได้เช่นกัน

3.4 <u>การหาตัวแทนการแบ่งกลุ่มของลายมือชื่อ</u>

จากความหมายของการแบ่งกลุ่มของลายมือชื่อดังแสดงในบทที่ 2 เมื่อพิจารณาการขาดหายของเส้น โดยทำการตรวจสอบลายมือชื่อ ตามแนวตั้งจากขอบซ้ายสุดของภาพไปยังขอบขวาสุดของภาพ ถ้าเกิดการขาดตอนของลายเส้น จะทำการนับการขาดของลายเส้นนี้ เพื่อแบ่งแยกจำนวนกลุ่มของลายมือชื่อ ตัวแทนนี้จะแสดงจำนวนกลุ่มของลายมือชื่อที่สามารถแบ่งได้

<u>กรรมวิธีพิสูจน์ลายมือชื่อ</u>

ลายมือชื่อแต่ละภาพ สามารถหาค่าตัวแทนทางคณิตศาสตร์ของเอกลักษณ์ต่างๆ ที่ประกอบอยู่ในลายมือชื่อได้ เมื่อเก็บรวบรวมลายมือชื่อที่แท้จริงจำนวนมากกว่า 30 ลายมือชื่อ แล้วกำหนดจุดของแต่ละเอกลักษณ์ลงบนแผนภาพ เพื่อดูลักษณะการกระจายของเอกลักษณ์นั้นๆ จะได้การแจกแจงของแต่ละเอกลักษณ์ในรูปแบบปรกติ (Normal distribution) สามารถแสดงได้ดังภาพที่ 3.2

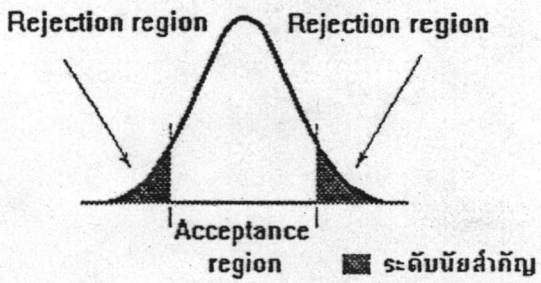


ภาพที่ 3.2 แสดการกระจายของตัวแทนของเอกลักษณ์ที่ประกอบอยู่ในลายมือชื่อที่รวบรวมมา

กำหนดให้ข้อมูลที่ศึกษามีการแจกแจงแบบปรกติ โดยมีค่าเฉลี่ย (μ) และความแปรปรวน (σ^2) ตามลำดับ สามารถแสดงได้คือ $X \sim (\mu,\sigma^2)$ โดยมีรูปแบบของการแจกแจงคือ

$$F(x) = (e^{-(x-\mu)^2/2\sigma^2}) / (\sqrt{2\pi})$$

โดยที่ข้อมูลส่วนใหญ่จะเกาะกลุ่มใกล้กับค่าเฉลี่ยของข้อมูล การพิสูจน์ลายมือชื่อ คือ การทดสอบความมีนัยสำคัญของแต่ละเอกลักษณ์ของลายมือชื่อที่ต้องการพิสูจน์ กับขอบเขตความเชื่อมั่นของเอกลักษณ์นี้จากลายมือชื่อที่รวบรวมมา เราอาศัยคุณสมบัติของการแจกแจงแบบปรกติ และความน่าจะเป็นมาช่วยในการตัดสินใจ สามารถแสดงได้ดังภาพที่ 3.3



ภาพที่ 3.3 แสดงขอบเขตของความเชื่อมั่นเพื่อใช้ในการตัดสินใจ

จากภาพที่ 3.3 หากค่าของเอกลักษณ์จากลายมือชื่อที่ต้องการพิสูจน์ ตกอยู่ในขอบเขตของความเชื่อมั่น (Acceptance region) แสดงว่า ความแตกต่างที่เกิดขึ้นไม่มีนัยสำคัญ เราจึงไม่อาจปฏิเสธได้ว่า เอกลักษณ์จากลายมือชื่อที่ต้องการพิสูจน์ ไม่แตกต่างกับเอกลักษณ์นี้จากลายมือชื่อที่รวบรวมมา แต่ถ้าเอกลักษณ์ของลายมือชื่อที่ต้องการพิสูจน์ ตกอยู่นอกขอบเขตของความเชื่อมั่น หรือตกอยู่ในช่วงของการปฏิเสธ (Rejection region)แสดงว่าเอกลักษณ์ของลายมือชื่อที่ต้องการพิสูจน์มีความแตกต่างกับเอกลักษณ์จากลายมือชื่อที่รวบรวมมาทั้งหมด

ระดับความมีนัยสำคัญ (Level of significance) คือ ค่าของความน่าจะเป็นหรือโอกาสที่เราจะไม่ยอมรับสมมติฐานที่ตั้งไว้ ซึ่งจะใช้สัญญลักษณ์ α แทน โดยทั่วไประดับความมีนัยสำคัญที่นิยมใช้ คือ 0.01 , 0.05 หรือ 0.1 การประมาณค่าขอบเขตแห่งความเชื่อมั่น ของพารามิเตอร์

สามารถทำได้ดังนี้

กำหนดให้ $\chi_1, \ldots, \chi_n$ คือ ตัวอย่างที่รวบรวมมาซึ่งมีการแจกแจงแบบปรกติ ซึ่งไม่ทราบค่าเฉลี่ย μ และความแปรปรวน σ^2 ของประชากร โดยที่สามารถประมาณค่าความแปรปรวนของตัวอย่างได้ คือ

กำหนดให้ n คือ จำนวนตัวอย่างที่รวบรวมมา

$\bar{\chi}$ คือ ค่าเฉลี่ยของตัวอย่าง

S^2 คือ ค่าความแปรปรวนของตัวอย่าง ซึ่งหาได้จาก

$$S^2 = \sum_{i=1}^{n} (\chi_i - \bar{\chi})/(n-1)$$

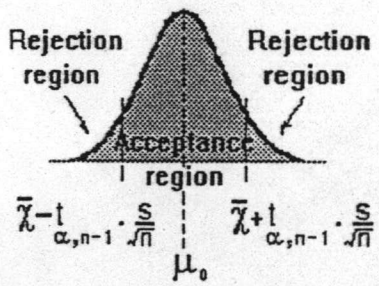
การประมาณค่าพารามิเตอร์ ซึ่งมีการแจกแจงปรกติ และไม่ทราบค่า μ และ σ^2 ด้วยระดับความเชื่อมั่น $100(1-\alpha)$ เปอร์เซ็นต์ ตัวสถิติที่ใช้ในการประมาณค่าคือ t-statistic โดยที่

$$t_{\alpha,n-1} = \frac{\bar{\chi} - \mu_0}{\frac{S}{\sqrt{n}}}$$

ค่า (n-1) คือ ค่าขั้นของความเป็นอิสระ (degree of freedom) ขอบเขตของการประมาณค่าจะอยู่ในช่วง

$$\bar{\chi} \pm t_{\alpha,n-1} \cdot \frac{S}{\sqrt{n}}$$

ซึ่งสามารถแสดงขอบเขตของการประมาณค่า ได้ดังภาพที่ 3.4



ภาพที่ 3.4 แสดงขอบเขตของการประมาณค่า พารามิเตอร์ด้วยระดับนัยสำคัญ α

สรุปขั้นตอนของการพิสูจน์ลายมือชื่อ ได้ดังต่อไปนี้

1) เก็บรวบรวมลายมือชื่อที่แท้จริง

2) หาตัวแทนทางคณิตศาสตร์เพื่ออธิบายแต่ละเอกลักษณ์ของลายมือชื่อที่รวบรวมมา โดยแบ่งได้ดังต่อไปนี้

(1) ค่าเจาะจง แทนภาพรวมของลายมือชื่อ

(2) ค่าเฉลี่ยความเข้มของลายเส้น

(3) ค่าความยาวโดยเปรียบเทียบของลายมือชื่อ

(4) ค่าการแบ่งกลุ่มของลายมือชื่อ

3) คำนวนค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของแต่ละตัวแทน จากลายมือชื่อที่เก็บรวบรวมมาทั้งหมด

4) กำหนดระดับความมีนัยสำคัญ หรือโอกาสที่จะปฏิเสธว่าเกิดความแตกต่างระหว่างตัวแทนของเอกลักษณ์จากลายมือชื่อที่ต้องการพิสูจน์ กับเอกลักษณ์นี้จากกลุ่มของลายมือชื่อที่รวบรวมมา

5) คำนวนขอบเขตความเชื่อมั่นของแต่ละตัวแทนของเอกลักษณ์ จากลายมือชื่อที่รวบรวมมา

6) หาตัวแทนทางคณิตศาสตร์ เพื่ออธิบายแต่ละเอกลักษณ์ของลายมือชื่อที่ต้องการพิสูจน์ ดังข้อ 2

7) เปรียบเทียบตัวแทนของแต่ละเอกลักษณ์ของลายลายมือชื่อที่ต้องกาพิสูจน์กับขอบเขตความเชื่อมั่นที่คำนวนได้ในข้อ 5

8) ถ้าตัวแทนของเอกลักษณ์ ตกอยู่ในขอบเขตความเชื่อมั่น แสดงว่า ความแตกต่างที่เกิดขึ้นไม่มีนัยสำคัญ จึงไม่อาจปฏิเสธได้ว่าเอกลักษณ์ของลายมือชื่อที่ต้องการพิสูจน์ แตกต่างกับเอกลักษณ์นี้จากลายมือชื่อที่รวบรวมมาทั้งหมด