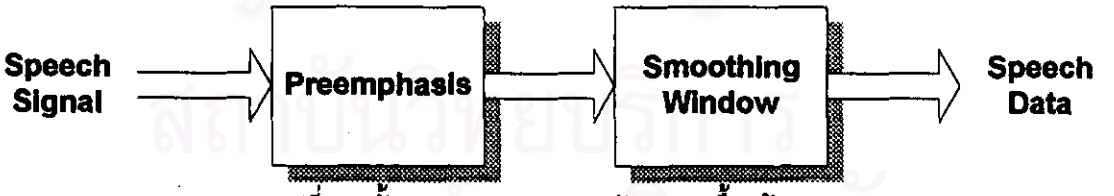
บทที่ 2
หลักการ ทฤษฎีและขั้นตอนวิธีการที่เกี่ยวข้อง

ในบทนี้กล่าวถึงหลักการ ทฤษฎี และขั้นตอนวิธีการสำคัญที่อ้างอิงถึง และนำมาประยุกต์ใช้กับระบบการรู้จำผู้พูดในส่วนของการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูด สามารถแบ่งได้เป็นสี่หัวข้อใหญ่ๆ คือการประมวลผลสัญญาณเบื้องต้น การสกัดลักษณะสำคัญ ขั้นตอนวิธีการฝึกฝนชุดรหัส การควอนไทซ์แบบเวกเตอร์ และแบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต

2.1 การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

การประมวลผลสัญญาณเบื้องต้นเป็นขั้นตอนในการจัดเตรียมสัญญาณเสียงพูด (Speech Signal) ที่ได้จากการบันทึกเสียง นำมาผ่านกรรมวิธีสัญญาณดิจิตอลได้เป็นข้อมูลเสียงพูด (Speech Data) เพื่อนำไปใช้กับการประมวลผลของขั้นตอนต่อไป เนื่องจากสัญญาณเสียงพูดมีค่าทางสถิติเปลี่ยนแปลงตามเวลา (Nonstationary) ทำให้ไม่สามารถจำลองสัญญาณเสียงพูดเป็นค่าทางสถิติได้ ดังนั้นในการประยุกต์ใช้สัญญาณเสียงพูดกับกรรมวิธีสัญญาณดิจิตอล จำเป็นต้องแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อย (Frame) โดยแต่ละส่วนย่อยจะมีความยาวประมาณ 10 ถึง 40 มิลลิวินาที ทำให้สัญญาณเสียงพูดในส่วนย่อยมีคุณสมบัติเปลี่ยนแปลงตามเวลาน้อยมากหรือไม่มีเลย จึงนิยามให้ส่วนย่อยนี้เป็นสัญญาณที่มีค่าทางสถิติไม่แปรเปลี่ยนตามเวลา (Stationary) ทำให้สามารถนำส่วนย่อยนี้มาคำนวณหาค่าทางสถิติเพื่อใช้แทนสัญญาณเสียงพูดได้ (Rabiner and Levinson, 1981; Furui, 1989)

ขั้นตอนในการประมวลผลสัญญาณเบื้องต้นสามารถแบ่งได้เป็น 2 ขั้นตอน คือ ขั้นตอนกรรมวิธีการเน้นล่วงหน้า (Preemphasis) ขั้นตอนกรรมวิธีการวางกรอบขนาดสัญญาณ (Smoothing Window) ดังแสดงในรูปที่ 2.1



รูปที่ 2.1 ขั้นตอนการประมวลผลสัญญาณเบื้องต้น

2.1.1 ขั้นตอนกรรมวิธีการเน้นล่วงหน้า (Preemphasis)

ขั้นตอนกรรมวิธีการเน้นล่วงหน้าเป็นการบีบอัดสัญญาณเสียงพูดในช่วงพิสัยพลวัต (Dynamic Range) มีผลทำให้อัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal to Noise Ratio) มีค่าสูงขึ้น โดยนำสัญญาณเสียงพูดผ่านวงจรกรองดิจิตอลอันดับที่หนึ่ง (First-Order Digital Filter) ที่มีฟังก์ชันถ่ายโอนดังแสดงในสมการที่ (2.1) และสมการผลต่างแสดงอยู่ในสมการที่ (2.2) (Furui, 1985)

$$H(z) = 1 - az^{-1} \tag{2.1}$$

$$\tilde{s}(n) = s(n) - as(n-1) \tag{2.2}$$

เมื่อ a เป็นค่าสัมประสิทธิ์ของวงจรกรองจะนิยมกำหนดให้เท่ากับ 0.95 (Rabiner and Juang, 1993)

$\tilde{s}(n)$ เป็นค่าของสัญญาณเสียงพูดขาออกที่ผ่านกรรมวิธีลำดับที่ n

$s(n)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ n

และ $s(n-1)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ $n-1$

2.1.2 ขั้นตอนกรรมวิธีการวางกรอบขนาดสัญญาณ (Smoothing Window)

ขั้นตอนนี้เป็นการแบ่งสัญญาณเสียงพูดเป็นส่วนย่อยๆ โดยการคูณแต่ละค่าของสัญญาณเสียงพูดที่อยู่ในกรอบสัญญาณเสียงพูดที่กำหนดไว้ด้วยค่าฟังก์ชันกรอบ (Window Function) ในงานวิจัยได้เลือกใช้ฟังก์ชันกรอบชนิด Hamming Window ดังแสดงในรูปที่ 2.2 มีผลทำให้เกิดการลดทอนแอมปลิจูดอย่างช้าๆ ที่บริเวณปลายแต่ละข้างของกรอบสัญญาณเสียงพูด เพื่อป้องกันการเปลี่ยนแปลงที่ไม่ต่อเนื่องที่บริเวณปลายของส่วนย่อยๆ สัญญาณเสียงพูดที่ผ่านขั้นตอนนี้จะได้เป็นข้อมูลเสียงพูดเพื่อนำไปใช้กับกรรมวิธีสัญญาณดิจิตอลต่อไป ซึ่งกรรมวิธีการวางกรอบขนาดสัญญาณจะเป็นไปตามสมการที่ (2.3) และ (2.4)

$$\tilde{x}_l(n) = x_l(n).w(n) \tag{2.3}$$

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \tag{2.4}$$

เมื่อ $l = 0,1,\ldots,L-1$ และ $n = 0,1,\ldots,N-1$

กำหนดให้ $x_l(n)$ คือค่าสัญญาณเสียงพูดของข้อมูลที่ n

$\tilde{x}_l(n)$ คือค่าสัญญาณเสียงพูดที่ผ่านกรรมวิธีการวางกรอบ

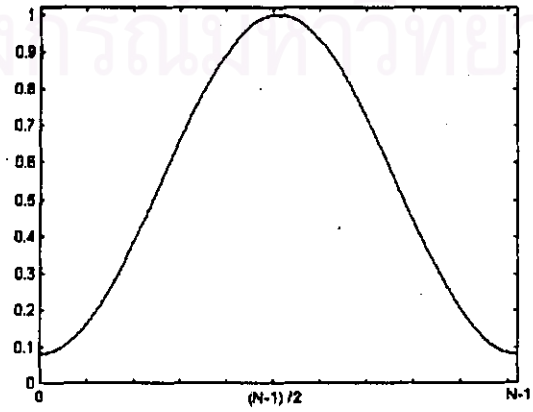
$w(n)$ คือฟังก์ชันกรอบชนิด Hamming Window

N คือจำนวนของข้อมูลในแต่ละกรอบสัญญาณเสียงพูด

n คือลำดับข้อมูลในกรอบสัญญาณเสียงพูดที่ l

L คือจำนวนของกรอบสัญญาณเสียงพูด

l คือลำดับของกรอบสัญญาณเสียงพูด



รูปที่ 2.2 ฟังก์ชันกรอบชนิด Hamming Window

2.2 การสกัดลักษณะสำคัญ (Feature Extraction)

การสกัดลักษณะสำคัญคือการวิเคราะห์หาค่าที่นำมาแทนข้อมูลเสียงพูดเพื่อนำไปใช้ในการสร้างแบบจำลองทางสถิติของสัญญาณเสียงพูดนั้นๆ เพราะฉะนั้นจำเป็นต้องเลือกใช้กรรมวิธีที่สามารถสกัดลักษณะสำคัญของสัญญาณเสียงพูดที่เราบ่งความต่างของผู้พูดแต่ละบุคคลได้เป็นอย่างดี ในงานวิจัยได้นำเสนอการวิเคราะห์ลักษณะสำคัญ 3 ชนิดเพื่อนำมาประยุกต์ใช้กับระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดได้แก่ ค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น (Linear Prediction Coefficients, LPC) ค่าสัมประสิทธิ์เซปสตรอล (Cepstral Coefficient, CEP) และค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล (Mel Frequency Cepstral Coeffieceint, MFCC) ลักษณะสำคัญทั้งสามชนิดนี้เป็นลักษณะสำคัญของช่องทางเดินเสียง (Vocal Tract) (Campbell, 1997) ซึ่งเป็นลักษณะสำคัญประเภทหนึ่งที่นิยมใช้ในกับระบบการรู้จำผู้พูดดังที่กล่าวในบทที่ 1 เนื่องจากช่องทางเดินเสียงเป็นลักษณะทางกายภาพที่สามารถแสดงความแตกต่างของแต่ละบุคคลได้ (วิสิทธิ์ ลีลาศิริวงศ์, 2535) ส่วนรายละเอียดและวิธีการคำนวณของลักษณะสำคัญทั้งสามกล่าวในหัวข้อถัดไป

2.2.1 สัมประสิทธิ์ของการประมาณพันธะเชิงเส้น (Linear Prediction Coefficients, LPC)

การประมาณพันธะเชิงเส้น (Linear Prediction) เป็นเทคนิคหนึ่งที่นิยมใช้ในการวิเคราะห์คุณสมบัติทางกายภาพของสัญญาณ เช่น Spectral Magnitude ของสัญญาณในช่วงเวลาที่เราสนใจได้อย่างแม่นยำ รวดเร็วและมีประสิทธิภาพ โดยใช้แบบจำลองที่ใช้การวิเคราะห์ค่าพารามิเตอร์ (Parametric Model) ที่สันนิษฐานว่าค่าพารามิเตอร์ที่ได้นี้เกิดจากช่องทางเดินเสียงเพื่อจำลองเป็นสัญญาณเสียงพูด เพราะฉะนั้นสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นได้จากแบบจำลองการประมาณพันธะเชิงเส้นของสัญญาณเสียงพูด ซึ่งแบบจำลองการประมาณพันธะเชิงเส้นคือการประมาณข้อมูลแบบ Linear Least Square หรือเรียกว่า Prediction (Sorenson, 1970) สำหรับงานวิจัยทางด้านสัญญาณเสียงพูดได้มีการนำค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นมาประยุกต์ใช้งานในเรื่องของการวิเคราะห์ สังเคราะห์ เข้ารหัสสัญญาณเสียงพูด และสร้างเป็นชุดอ้างอิงของการรู้จำแบบรูป (Pattern Recognition) จำนวนมาก เนื่องจากค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นช่วยในการประมาณการสั่นสะเทือนของช่องทางเดินเสียงขณะออกเสียงพูด รูปร่างของช่องทางเดินเสียง ความถี่และแบนด์วิดท์ของเรโซแนนซ์ที่ช่องทางเดินเสียง (O'Shaughnessy, 1988)

หลักการพื้นฐานของแบบจำลองการประมาณพันธะเชิงเส้น (Linear Prediction Model) คือการคำนวณหาค่าอนาคตของสัญญาณจากการประมาณการผสมเชิงเส้น (Linear Combination) ของค่าของสัญญาณก่อนหน้านั้น โดยสมมติให้สัญญาณที่เราสนใจเกิดจากแหล่งกำเนิดสัญญาณที่ถูกกระตุ้นจากวงจรกรองแบบเชิงเส้น ในงานวิจัยนี้สัญญาณที่สนใจคือสัญญาณเสียงพูดเพราะฉะนั้นจะกล่าวเฉพาะแบบจำลองสัญญาณเสียงพูดเท่านั้น แหล่งกำเนิดของสัญญาณก็คือลมที่ผ่านผ่านช่องว่างระหว่างเส้นเสียง (Vocal Cord) ออกมาเป็นรายคาบหรือสัญญาณรบกวนที่เกิดจากช่องแคบของการหดตัวของช่องทางเดินเสียง และที่มีผลกระทบกับช่องทางเดินเสียงส่วนบน พิจารณาให้สัญญาณ $s(t)$ ถูกสุ่มตัวอย่างทุกๆ เวลา T วินาที ดังนั้น

$$s(n) \stackrel{\Delta}{=} s(nT) \qquad \text{เมื่อ } n \text{ คือจำนวนเต็ม} \tag{2.5}$$

และใช้การแปลงแบบแซดจะได้เป็น $S(z)$ ซึ่งเราสมมติให้แบบจำลองของการเกิดสัญญาณ $s(n)$ ประกอบไปด้วยแหล่งกำเนิดที่กระตุ้น (Excitation Source) $U(z)$ และผ่านวงจรกรองเชิงความถี่ $H(z)$ จะได้

$$S(z) = U(z)H(z) \tag{2.6}$$

การวิเคราะห์ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นของสัญญาณ $s(n)$ คือการประมาณดีคอนโวลูชัน (Deconvolves) ของ $\hat{U}(z)$ และ $\hat{H}(z)$ ซึ่งเป็นการหาค่าที่ใกล้เคียงที่สุดหรือค่าที่เป็นไปได้มากที่สุดในแง่ Mean Square จากสัญญาณ $S(z)$ นั้นเอง กำหนดให้ $\hat{h}(n)$ คือการประมาณผลตอบสนองต่ออิมพัลส์ (Impulse Response) ของระบบที่กำเนิดสัญญาณ $s(n)$ เนื่องจากไม่สามารถหา $h(n)$ ได้โดยตรง ปัญหาพื้นฐานของการจำลองคือ $\hat{U}(z)$ มักจะถูกกำหนดให้เป็น Flat Spectral Envelope ซึ่งเป็นการจำกัดรายละเอียดของสเปกตรัมใน $\hat{H}(z)$ ทำให้การจำลองสัญญาณเสียงอโฆษะ (Unvoiced) ได้ไม่ดี เนื่องเสียงอโฆษะเป็นการกระตุ้นของสัญญาณรบกวนที่เหมือนกับสัญญาณรบกวนสีขาว จึงทำให้กรอบคลื่น (Envelope) ของสเปกตรัมไม่เรียบ ส่วนสัญญาณเสียงพูดในช่วงที่เป็นรายคาบคือเสียงโฆษะ (Voiced) มีฮาร์โมนิกของเส้นสเปกตรัมเป็นพื้นที่สม่ำเสมอ (Uniform-Area) ทำให้มีเส้นสเปกตรัมเรียบ

สมมติให้วงจรกรอง $H(z)$ มีจำนวน Pole เท่ากับ p และจำนวน Zero เท่ากับ $q+1$ หมายความว่าการจำลองสัญญาณ $\hat{s}(n)$ เป็นการผสมแบบเชิงเส้น (Linear Combination) ของตัวอย่างขาออกก่อนหน้านี้จำนวน p ค่ากับตัวอย่างขาเข้าก่อนหน้านี้จำนวน $q+1$ ค่า เพื่อสังเคราะห์หาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น ดังสมการที่ (2.7)

$$\hat{s}(n) = \sum_{k=1}^{p} a_k \hat{s}(n-k) + G\sum_{l=0}^{q} b_l \hat{u}(n-l) \tag{2.7}$$

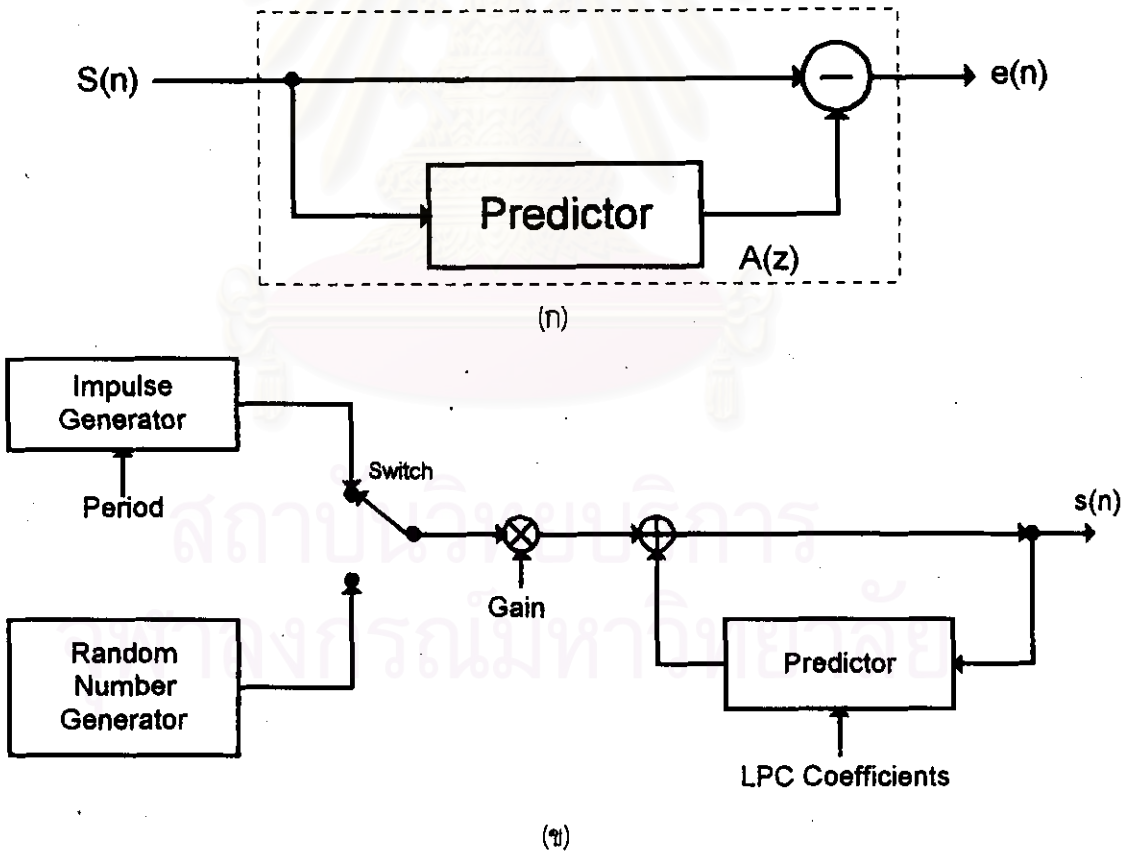
เมื่อ G คืออัตราขยาย (Gain Factor) ของการกระตุ้น โดยสมมติให้ $b_0 = 1$ และ a_k คือค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น จากสมการที่ (2.6) จะได้

$$\hat{H}(z) = \frac{\hat{S}(z)}{\hat{U}(z)} = G\frac{1+\sum_{l=1}^{q} b_l z^{-l}}{1-\sum_{k=1}^{p} a_k z^{-k}} \tag{2.8}$$

แบบจำลองของวงจรกรองที่ใช้ในการหาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นมี 3 แบบจำลองดังนี้ (O'Shaughnessy, 1988)

1. แบบจำลอง All-Pole หรือเรียกว่า Autoregressive Model (AR Model) โดยกำหนดให้ $q = 0$
2. แบบจำลอง All-Zero มีอีกชื่อหนึ่งว่า Moving Average Model (MA Model) โดยกำหนดให้ข้อมูลออกเกิดจากการหาค่าเฉลี่ยของน้ำหนักจากข้อมูลเข้า เมื่อจำนวน Zero เท่ากับ $q+1$ ค่า
3. แบบจำลองที่อยู่ในรูปทั่วไปของทั้งสองแบบคือมีทั้ง Pole และ Zero รู้จักกันดีในชื่อของ Autoregressive Moving Average Model (ARMA Model) โดยให้ $p, q > 0$

การหาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นจากแบบจำลอง AR เป็นการใช้สมการเชิงเส้น (Linear Equation) ส่วนแบบจำลอง MA จะใช้สมการไม่เชิงเส้น (Nonlinear Equation) ซึ่งแบบจำลอง ARMA ให้ผลที่แม่นยำกว่าทั้งสองแบบจำลองที่กล่าวมา แต่แบบจำลองที่อยู่ในรูปอย่างง่ายก็คือแบบจำลอง AR ซึ่งเป็นแบบจำลองที่จำลองสเปกตรัมของสัญญาณได้ดีมากเฉพาะในช่วงเรโซแนนซ์เท่านั้น ค่าพลังงานที่ความถี่ต่างๆ แสดงคุณสมบัติทางกายภาพของสัญญาณมากกว่าช่วงที่ไม่มีความถี่ ซึ่งหมายความว่าสัญญาณที่มียอดของสเปกตรัม (Spectral Peak) แสดงคุณสมบัติได้ดีกว่าสัญญาณที่มีค่า Zero ตัวอย่างเช่น ช่องทางเดินเสียงแสดงลักษณะของหลอดเสียง (Acoustic Tube) จากการเปลี่ยนแปลงพื้นที่ของช่องทางเดินเสียงโดยแสดงเป็นความถี่เรโซแนนซ์หนึ่งค่าต่อแบนด์วิดท์หนึ่งกิโลเฮิร์ต และค่า Zero ในสเปกตรัมของสัญญาณเสียงพูดขึ้นอยู่กับการเปลี่ยนแปลงเส้นทางเดินของเสียง เช่น เสียงที่ผ่านจากปอดทั้งสองข้างหรือโพรงจมูกมีผลต่อการได้ยินน้อยกว่าเสียงที่พ่นออกมาจากปาก เป็นต้น แสดงให้เห็นว่าระบบการได้ยินของมนุษย์ไวต่อยอดของสเปกตรัม (Spectral Peak) มากกว่ารูปคลื่นในช่วงที่ไม่มีความถี่หรือพลังงาน ทำให้สามารถประมาณผลกระทบที่เกิดจากค่า Zero จากแบบจำลอง All-Pole ได้ (Markel and Gray, 1980; O'Shaughnessy, 1988)



รูปที่ 2.3 แบบจำลอง All-Pole ที่ใช้วิเคราะห์หาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น (O'Shaughnessy, 1988)

(ก) แบบจำลองของการวิเคราะห์การประมาณพันธะเชิงเส้นสำหรับเสียงพูด

(ข) แบบจำลองการสังเคราะห์เสียงพูดจากแบบจำลองการประมาณพันธะเชิงเส้น

ให้สัญญาณ $s(n)$ คือสัญญาณที่ถูกกรองโดยวงจร Inverse หรือ Predictor ดังแสดงในรูปที่ 2.3 (ก) เป็นการ Inverse ของสัญญาณที่ได้จากวงจรกรองแบบ All-Pole ที่ใช้ในการสังเคราะห์ $\hat{H}(z)$ ดังสมการที่ (2.9)

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \tag{2.9}$$

และเอาท์พุตที่ได้คือ $e(n)$ เรียกว่าค่าผิดพลาด (Error) หรือสัญญาณ Residual ดังสมการที่ (2.10)

$$e(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k) \tag{2.10}$$

ในสัญญาณเสียงโฆษะค่าผิดพลาดคือสัญญาณที่มีค่า Peak สูงมากและอยู่ห่างกันในทางเวลาของความถี่ Pitch ส่วนค่าผิดพลาดของสัญญาณเสียงอโฆษะสังเกตได้จากการสังเคราะห์ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นและสัญญาณรบกวนสีขาว ซึ่งมีคุณสมบัติทางกายภาพของสัญญาณบางอย่างที่วงจรกรองแบบ All-Pole ไม่สามารถจำลองได้

วิธีการ Least-squares

วิธีการ Least-squares เป็นวิธีการเลือกค่าสัมประสิทธิ์ a_k ให้มีค่าเฉลี่ยพลังงานของ $e(n)$ ในแต่ละเฟรม (Frame) มีค่าน้อยที่สุด สามารถคำนวณหาค่าสัมประสิทธิ์ a_k ได้สองวิธีคือวิธีอัตสหสมพันธ์ (Autocorrelation) กับวิธีแปรปรวนร่วม (Covariance) ในงานวิจัยนี้ได้ใช้วิธีอัตสหสมพันธ์ จึงกล่าวถึงแต่เฉพาะวิธีอัตสหสมพันธ์เท่านั้น ให้สัญญาณเสียงพูด $s(n)$ ที่ผ่านการประมวลผลสัญญาณเบื้องต้นและวางกรอบขนาดสัญญาณแบบ Hamming คือ $x(n)$ ดังสมการที่ (2.3)

กำหนดให้ E คือพลังงานของค่าผิดพลาด เมื่อ $e(n)$ คือค่าผิดพลาดของสัญญาณ $x(n)$ ดังสมการที่ (2.11)

$$E = \sum_{n=-\infty}^{\infty} e^2(n) \tag{2.11}$$

$$= \sum_{n=-\infty}^{\infty} \left[x(n) - \sum_{k=1}^{p} a_k x(n-k) \right]^2 \tag{2.12}$$

$$= \sum_{n=-\infty}^{\infty} x^2(n) - \sum_{n=-\infty}^{\infty} \left[2x(n) \sum_{k=1}^{p} a_k x(n-k) \right] + \sum_{n=-\infty}^{\infty} \left[\sum_{k=1}^{p} a_k x(n-k) \right]^2 \tag{2.13}$$

$$= \sum_{n=-\infty}^{\infty} x^2(n) - 2\sum_{k=1}^{p} a_k \sum_{n=-\infty}^{\infty} x(n)x(n-k) + \sum_{n=-\infty}^{\infty} \left[\sum_{k=1}^{p} a_k x(n-k) \right]^2 \tag{2.14}$$

การหาค่า a_k ที่ทำให้ E มีค่าต่ำที่สุดสามารถกระทำได้โดยการหา $\partial E / \partial a_i = 0$ ในสมการที่ (2.14) ของแต่ละค่า $i = 1,2,3,\ldots,p$ เมื่อ $k = 1,2,3,\ldots,p$ ได้เป็นสมการเชิงเส้น p สมการ และในแต่ละสมการมี k ค่า ซึ่งก็คือ a_k ที่ไม่ทราบค่า p ตัวดังนี้

$$\frac{\partial E}{\partial a_i} = 0 = -2\sum_{n=-\infty}^{\infty} x(n)x(n-i) + 2\sum_{n=-\infty}^{\infty}\left[\sum_{k=1}^{p} a_k x(n-k)\right]x(n-i) \tag{2.15}$$

$$\sum_{n=-\infty}^{\infty} x(n-i)x(n) = \sum_{k=1}^{p} a_k \sum_{n=-\infty}^{\infty} x(n-i)\cdot x(n-k) \tag{2.16}$$

เนื่องจากพจน์แรกเป็นอัตสหสมพันธ์ $R(i)$ ของ $x(n)$ ที่มีความยาวจำกัด สามารถเขียนสมการที่ (2.16) ใหม่ได้เป็นสมการที่ (2.17)

$$\sum_{k=1}^{p} a_k R(i-k) = R(i), \qquad 1 \le i \le p \tag{2.17}$$

เมื่อ $$R(i) = \sum_{n=i}^{N-1} x(n)x(n-i) \tag{2.18}$$

เมื่อแก้สมการที่ (2.14) ซึ่งประกอบไปด้วยสมการเชิงเส้น p สมการได้เป็นค่าสัมประสิทธิ์ a_k ที่ต้องการออกมา โดยเขียนสมการที่ (2.14) ให้อยู่ในรูปเมตริกซ์ $Ra = r$

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ r_1 & r_0 & & \vdots \\ \vdots & & \ddots & r_1 \\ r_{p-1} & \cdots & r_1 & r_0 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} \tag{2.19}$$

จากสมการที่ (2.16) เห็นได้ว่าเมตริกซ์ R เป็นเมตริกซ์ชนิด Toepliz ซึ่งเป็นเมตริกซ์ที่สมมาตรคือมีค่าในแนวทแยงมุมเท่ากันทั้งหมดทุกแนว ดังนั้นการคำนวณหาค่าสัมประสิทธิ์ a_k จะอาศัยวิธีการวนซ้ำของ Levinson-Durbin (Picone, 1996; O'Shaughnessy, 1988)

ขั้นตอนวิธีการวนซ้ำของ Levinson-Durbin

ขั้นตอนวิธีการวนซ้ำของ Levinson-Durbin เป็นเทคนิคที่ใช้ในการคำนวณหาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น a_i ในเมตริกซ์ a เมื่อ $i = 1,2,3,\ldots,p$ และ p เป็นอันดับ (Order) ของการวิเคราะห์หาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นจากวิธีอัตสหสมพันธ์ มีขั้นตอนวิธีการดังต่อไปนี้

วนซ้ำขั้นตอนที่ 1 ถึง 4 ของค่า i แต่ละค่า เมื่อ $i = 1,2,3,\ldots,p$

ขั้นตอนที่ 1 กำหนดค่าเริ่มต้น

$$E_0 = R(0) \quad \text{และ} \quad \alpha_0 = 0 \tag{2.20}$$

ขั้นตอนที่ 2 คำนวณหาค่าสัมประสิทธิ์การสะท้อน (Reflection Coefficient)

$$k_i = \frac{R(i) - \sum_{j=1}^{i-1} \alpha_{i-1}(j)R(i-j)}{E_{i-1}} \tag{2.21}$$

เมื่อ $R(i)$ และ $R(i-j)$ คำนวณได้จากสมการที่ (2.18)

ขั้นตอนที่ 3 คำนวณหาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น

$$\alpha_i(i) = k_i \tag{2.22}$$

$$\alpha_i(j) = \alpha_{i-1}(j) - k_i\alpha_{i-1}(i-j) \tag{2.23}$$

วนซ้ำสมการที่ (2.22) และ (2.23) เมื่อ $j = 1,2,3,\ldots,i-1$

ขั้นตอนที่ 4 คำนวณค่าผิดพลาดใหม่

$$E_i = (1 - k_i^2)E_{i-1} \tag{2.24}$$

เมื่อสิ้นสุดการวนซ้ำขั้นตอนทั้งสี่แล้ว เราจะได้ค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นที่รอบสุดท้าย ($i = p$) คือ $\alpha_p(j)$ เมื่อ $j = 1,2,3,\ldots,p$ นั้นคือค่าสัมประสิทธิ์ที่มีค่าความผิดพลาดของการประมาณพันธะเชิงเส้นน้อยที่สุดนั้นเอง ดังในสมการที่ (2.25)

$$a_k = \alpha_p(k) \tag{2.25}$$

เมื่อ k คืออันดับของค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น

2.2.2 สัมประสิทธิ์เซปสตรอล (Cepstral Coefficients, CEP)

สัมประสิทธิ์เซปสตอลสามารถคำนวณหาได้หลายวิธี เช่น คำนวณจากการแปลงฟูริเยร์อย่างเร็ว (Fast Fourier Transform) และคำนวณจากแบบจำลองการประมาณพันธะเชิงเส้น เป็นต้น ในงานวิจัยนี้ได้เลือกใช้ค่าสัมประสิทธิ์เซปสตรอล (c_m) ที่คำนวณจากแบบจำลองการประมาณพันธะเชิงเส้นเนื่องจากว่าง่ายต่อการคำนวณและใช้เวลาน้อยมาก วิธีการคำนวณหาสัมประสิทธิ์เซปสตรอลแสดงในสมการที่ (2.26) และ (2.27) (Rabiner and Juang, 1993)

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \qquad 1 \le m \le p \tag{2.26}$$

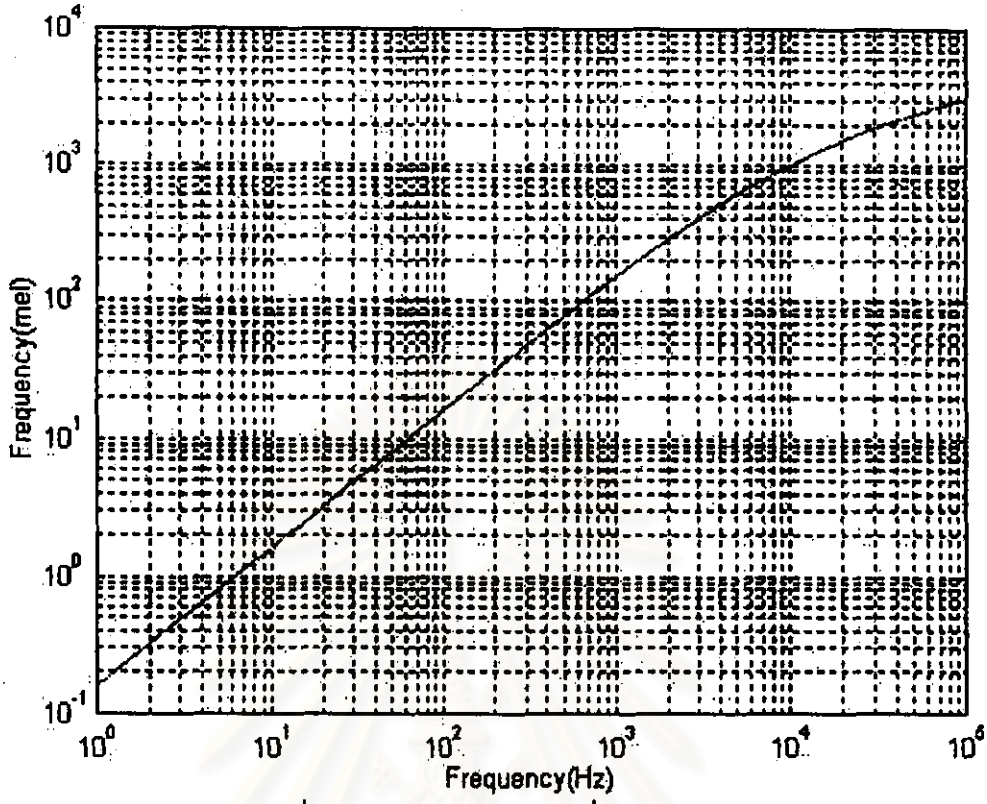
$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \qquad m > p \tag{2.27}$$

เมื่อ p คืออันดับของค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น

2.2.3 สัมประสิทธิ์เซปสตรอลบนความถี่เมล (Mel Frequency Cepstral Coefficients, MFCC)

สัมประสิทธิ์เซปสตรอลบนความถี่เมลคำนวณได้จากการวิเคราะห์โดยไม่ใช้พารามิเตอร์ (Non-Parametric Analysis) คือไม่มีการสร้างแบบจำลองสัญญาณเพื่อปรับค่าพารามิเตอร์ในแบบจำลองให้ได้เป็นสัญญาณที่เราต้องการ แต่เป็นการนำสัญญาณผ่านวงจรกรองแบบผ่านแถบความถี่ (Band Pass Filter) หลายวงจร โดยแต่ละวงจรกรองมีช่วงความถี่ที่ผ่านได้แตกต่างกันดังแสดงในรูปที่ 2.4 และวงจรกรองที่ใช้เป็นชุดวงจรกรองแบบดิจิตอล (Digital Filter Bank) ที่ใช้วิเคราะห์สเปกตรัมของสัญญาณเสียงพูดโดยเลียนแบบตามการได้ยินของมนุษย์สามารถทำได้หลายวิธีขึ้นอยู่กับกรรมวิธีที่นำมาใช้กับแกนความถี่ได้แก่ (Tolba and O'Shaughnessy, 1998; Tuzun, Demirekler and Nakiboglu, 1994)

- Uniform Spacing (FFT)
- Exponential Spacing (Wavelet Transform)
- Perceptually-Derived Spacing (Mel Scale หรือ Bark Scale)



รูปที่ 2.4 Mel Scale ของความถี่สัญญาณเสียง

ในงานวิจัยนี้ได้เลือกใช้กรรมวิธี Perceptually-derived Spacing หรือเรียกอีกชื่อหนึ่งว่า Mel scale เป็นการเปลี่ยนแกนความถี่ของสัญญาณเสียง (Acoustic Frequency) ไปเป็นแกนความถี่ของการได้ยิน (Perceptual Frequency) จึงกล่าวถึงแต่กรรมวิธีนี้เท่านั้น Mel scale ใช้หลักการ Mapping จากการประมาณความถี่ช่วงที่เป็นเชิงเส้น (Linear Frequency Scale) ให้เป็นความถี่ที่ไม่เชิงเส้น (Nonlinear Frequency Scale) ตามลักษณะการได้ยินของมนุษย์ดังแสดงสมการที่ (2.28) (Tolba and O'Shaughnessy, 1998)

$$mel(f) = 2595\log_{10}(1+\frac{f}{700}) \tag{2.28}$$

เมื่อ f คือช่วงความถี่ที่ยังเป็นเชิงเส้น จากรูปที่ 2.4 เห็นได้ว่าลักษณะการได้ยินของมนุษย์ในช่วงที่เป็นเชิงเส้น (Linear scale) คือประมาณช่วงความถี่ตั้งแต่ 0 ถึง 1000 Hz และช่วงที่ไม่เป็นเชิงเส้น (Logarithmic scale) คือช่วงความถี่ตั้งแต่ 1000 Hz ขึ้นไป แบนด์วิดท์ของวงจรกรองสามารถคำนวณได้จากสมการที่ (2.29) ดังนี้

$$BW_{critical} = 25 + 75\left[1+1.4(f/1000)^2\right]^{0.69} \tag{2.29}$$

วิธีการคำนวณแบบจำลองของชุดวงจรกรอง (Filter Bank Model) ที่ง่ายที่สุดและให้ผลที่มีประสิทธิภาพดีคือการแปลงแบบฟูริเยร์ (Fourier Transform) ของสัญญาณและหาข้อมูลออกของชุดวงจรกรองของแต่ละวงจร (X_N) จากความถี่ที่ใส่เข้าไป การคำนวณหาข้อมูลออกของชุดวงจรกรองสามารถคำนวณได้ตามสมการที่ (2.30)

$$X_N = \frac{1}{N_{FB}} \sum_{k=1}^{N_{FB}} \omega_{FB}(k) X(f + \delta f(f,k)) \qquad (2.30)$$

เมื่อ N_{FB} คือจำนวนของตัวอย่างที่ใช้ในการหาค่าเฉลี่ย

ω_{FB} คือฟังก์ชันการถ่วงน้ำหนัก

$\delta f(f,k)$ คือฟังก์ชันที่อธิบายถึงความถี่ที่อยู่ใกล้เคียงกับความถี่ f เพื่อนำมาใช้คำนวณหาค่าเฉลี่ย

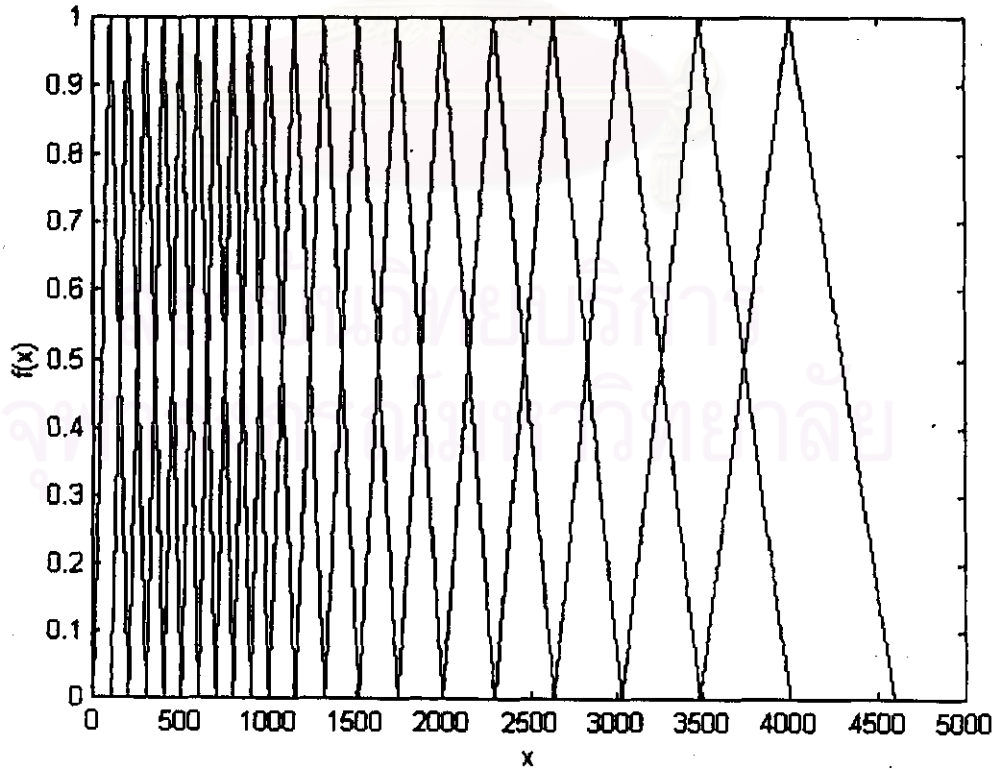
ในการคำนวณหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมลใช้ค่าข้อมูลออกที่ได้จากวงจรกรองแถบผ่าน N วงจรเมื่อใส่ค่าลอการิทึมของขนาดสเปกตรัมของสัญญาณเสียงเข้าไป เพื่อประมาณความถี่ตอบสนองของ Basilar Membrane ใน Cochlea ที่อยู่ในหูชั้นใน วงจรกรองที่นำมาใช้นี้เป็นวงจรกรองแถบผ่านแบบสามเหลี่ยมดังแสดงในรูปที่ 2.5 เพราะฉะนั้นการคำนวณหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล (C_n) สามารถหาได้จากสมการดังต่อไปนี้ (Tolba and O'Shaughnessy, 1998)

$$C_n = \sum_{k=1}^{N} X_k \cos(\frac{\pi n}{N}(k - 0.5)), \quad n = 1,2,\ldots,M \qquad (2.31)$$

เมื่อ M คือจำนวนของสัมประสิทธิ์เซปสตอล

N คืออันดับในการวิเคราะห์

X_k คือค่าลอการิทึมของพลังงานที่ค่าข้อมูลออกของวงจรกรอง k วงจร เมื่อ $k = 1,2,\ldots,N$

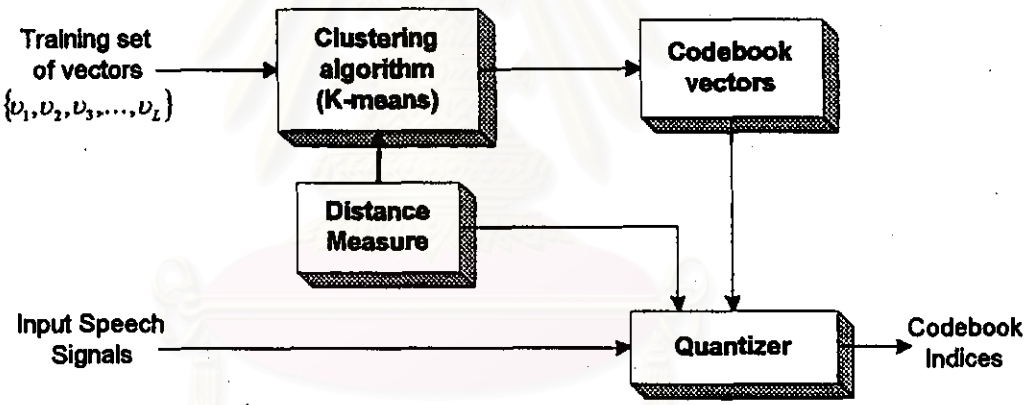


รูปที่ 2.5 วงจรกรองแบบผ่านแถบความถี่ที่ให้ค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล

2.3 การควอนไทซ์แบบเวกเตอร์ (Vector Quantization, VQ)

การควอนไทซ์ (Quantization) คือกระบวนการประมาณค่าแอมปลิจูดของสัญญาณต่อเนื่องโดยใช้สัญญาณไม่ต่อเนื่อง ซึ่งเป็นกระบวนการสำคัญในการบีบอัดข้อมูล (Data Compression) หรือการเข้ารหัส (Coding) เพื่อเป็นการลดจำนวนบิตที่จำเป็นในการส่งหรือเก็บข้อมูล การควอนไทซ์สามารถแบ่งได้เป็น 2 ประเภทคือการควอนไทซ์แบบสเกลาร์ (Scalar Quantization) และการควอนไทซ์แบบเวกเตอร์ (Vector Quantization) การควอนไทซ์ที่อิสระกับแต่ละค่าสัญญาณหรือพารามิเตอร์เรียกว่าการควอนไทซ์แบบสเกลาร์ ส่วนการควอนไทซ์แบบเวกเตอร์จะควอนไทซ์ชุดของค่าพารามิเตอร์เป็นเวกเตอร์เดียว

การควอนไทซ์แบบเวกเตอร์ได้รับความนิยมใช้ในการเข้ารหัสสัญญาณเสียงพูด (Speech Coding) การเข้ารหัสข้อมูลภาพ (Image Coding) และการรู้จำเสียงพูด (Speech Recognition) ในระบบการรู้จำเสียงพูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Models, HMM) การควอนไทซ์แบบเวกเตอร์มีความสำคัญในการอธิบายต้นแบบสัญญาณเสียงดิสครีต (Discrete Acoustic Prototype) ของสัญญาณเสียงพูดของแบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต (Discrete Hidden Markov Models, DHMM)



รูปที่ 2.6 การควอนไทซ์แบบเวกเตอร์ (Zhang, 1995)

หลักการพื้นฐานของการควอนไทซ์แบบเวกเตอร์โดยทั่วๆ ไปแบ่งได้เป็น 2 ช่วงคือช่วงการออกแบบชุดรหัสและช่วงการควอนไทซ์ ดังในรูปที่ 2.6 ช่วงการออกแบบชุดรหัสสร้างชุดรหัสและฝึกฝนชุดรหัสจากข้อมูลสัญญาณเสียงในชุดฝึกฝน เพื่อนำไปสร้างเป็นต้นแบบอ้างอิงชุดรหัส (Codebook Reference Template) ส่วนในช่วงของการควอนไทซ์คือการนำสัญญาณเสียงพูดมาควอนไทซ์กับต้นแบบอ้างอิงชุดรหัสและได้เป็นค่ารหัสของสัญญาณเสียงพูดนั้นๆ สามารถสรุปได้ว่าการควอนไทซ์แบบเวกเตอร์คือการแบ่งข้อมูลของปริภูมิสัญญาณ (Signal Space) ไปตามจำนวนของพื้นที่หรือเรียกว่าเซลล์ (Cell) และใช้เวกเตอร์จุดศูนย์ถ่วง (Centroid Vector) ของแต่ละพื้นที่แทนจุดของข้อมูลทั้งหมดที่อยู่ในพื้นที่นั้นๆ สามารถอธิบายตามหลักการและทฤษฎีได้ดังนี้ (Huang, 1989; Zhang, 1995)

สมมติให้ $x = [x_1, x_2, x_3, \ldots, x_d]^t \in R^d$ คือเวกเตอร์ที่มีมิติเท่ากับ d และประกอบไปด้วย $x_k, 1 \le k \le d$ และเป็นค่าจำนวนจริงที่ได้จากตัวแปรสุ่มที่มีแอมปลิจูดแบบต่อเนื่อง (Continuous-

amplitude Random Variable) ในการควอนไทซ์แบบเวกเตอร์จะเปลี่ยนเวกเตอร์ x ให้เป็นค่าจำนวนจริงที่ได้จากแอมปลิจูดไม่ต่อเนื่องที่ขนาดเวกเตอร์เท่ากับ d มิติ (y) เราเรียกว่าเวกเตอร์ x ถูกควอนไทซ์ไปเป็นเวกเตอร์ y

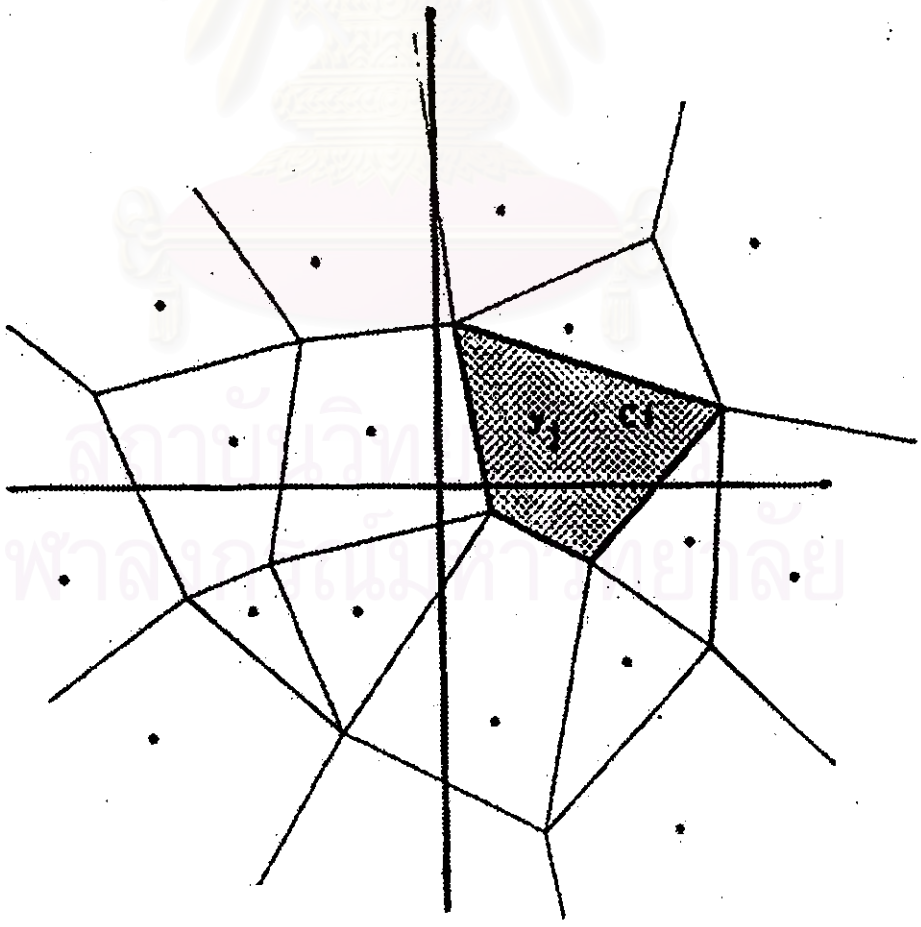
$$y = q(x) \tag{2.32}$$

จากสมการที่ (2.32) $q(\ )$ คือโอเปอเรเตอร์ของการควอนไทซ์ (Quantisation Operator) y คือเวกเตอร์ที่ประกอบไปด้วยเซตจำกัด $Y = \{y_i, 1 \le i \le L$ เมื่อ $y_i = [y_1, y_2, y_3, \ldots, y_d]$ เซต Y นี้คือชุดรหัส (Codebook) L คือขนาดของชุดรหัสและ $\{y_i\}$ คือเซตของเวกเตอร์รหัส (Codeword) ขนาด L ของชุดรหัสเรียกว่า "จำนวนขั้นในชุดรหัส"

การออกแบบชุดรหัสคือการแบ่งปริภูมิที่มีขนาด d มิติของเวกเตอร์ x ที่ถูกสุ่มมาจากสัญญาณตั้งต้นออกเป็น L เซลล์ $C_i, 1 \le i \le L$ และแทนแต่ละเซลล์ C_i ด้วยเวกเตอร์ y_i ตัวควอนไทซ์ถูกแทนด้วยเวกเตอร์รหัส y_i ถ้า x อยู่ใน C_i

$$q(x) = y_i, \ if \ x \in C_i \tag{2.33}$$

กระบวนการออกแบบในขั้นตอนนี้เรียกว่า "การฝึกฝนชุดรหัส" ตัวอย่างการแบ่งข้อมูลในปริภูมิที่มีขนาด 2 มิติ ($d = 2$) เพื่อนำไปใช้ในการควอนไทซ์แสดงดังในรูปที่ 2.7



รูปที่ 2.7 การแบ่งปริภูมิที่มีขนาด 2 มิติ (Huang, 1989)

เมื่อ x ถูกควอนไทซ์เป็น y ผลของความผิดพลาดที่เกิดจากการควอนไทซ์และค่าความเพี้ยนวัดเป็นระยะห่างระหว่าง x กับ y คือ $d(x,y)$ ซึ่งเป็นการวัดคุณภาพของการควอนไทซ์ การวัดนี้ต้องบอกถึงความแตกต่างของค่าความเพี้ยนที่ใช้บอกความแตกต่างของสัญญาณเสียงพูดได้ การวัดที่ธรรมดาที่สุดวิธีหนึ่งคือการวัดระยะห่างของยูคริเดียน (Euclidean Distance) ดังสมการที่ (2.34) โดยสมมติให้ความแตกต่างของค่าความเพี้ยนของการควอนไทซ์พารามิเตอร์ต่างๆ มีค่าเท่ากัน

$$d(x,y)=(x-y)^T(x-y) \tag{2.34}$$

แต่ถ้ามีการถ่วงน้ำหนักไม่เท่ากันจะใช้การวัดระยะห่างของมาฮาลาโนบิส (Mahalanobis Distance) โดยมีการใช้ความแปรปรวนผกผันของ y มาคิดด้วย ดังสมการที่ (2.35)

$$d(x,y)=(x-y)^T\Sigma^{-1}(x-y) \tag{2.35}$$

การออกแบบชุดรหัส L ระดับคือการแบ่งข้อมูลในปริภูมิ (Space) d มิติออกเป็น L เซลล์และแทนแต่ละเซลล์ด้วยเวกเตอร์ที่ถูกควอนไทซ์แล้ว การหาเวกเตอร์ของตัวควอนไทซ์ที่เหมาะสมสามารถหาได้จากค่าความเพี้ยนรวมทั้งหมดของแต่ละระดับของชุดรหัสที่มีค่าต่ำที่สุด ค่าความเพี้ยนรวมทั้งหมดสามารถหาได้จาก

$$D=E[d(x,y_i)] \tag{2.36}$$

$$=\sum_{i=1}^{L}\Pr(x\in C_i)E[d(x,y_i)/x\in C_i] \tag{2.37}$$

$$=\sum_{i=1}^{L}\Pr(x\in C_i)\int_{x\in C_i}d(x,y_i)f(x)dx \tag{2.38}$$

เมื่อ $\Pr(x\in C_i)$ คือค่าความน่าจะเป็นแบบดิสครีต (Discrete Probability) ของ x ที่อยู่ในเซลล์ C_i $f(x)$ คือฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function) ของ x และ $E[\ ]$ คือค่าคาดหวัง

2.3.1 การออกแบบชุดรหัส (Codebook Design)

การออกแบบชุดรหัสอาจเรียกได้ว่าเป็นการจัดกลุ่มของข้อมูลแบบหนึ่ง เพื่อนำไปสร้างเป็นต้นแบบอ้างอิงชุดรหัส (Codebook Reference Template) ใช้ในการทดสอบความคล้ายคลึงกันของรูปแบบของชุดข้อมูลที่เข้ามาทดสอบกับต้นแบบอ้างอิงชุดรหัสก่อนควอนไทซ์แบบเวกเตอร์และเข้าสู่แบบจำลองทางสถิติต่อไป การจัดกลุ่มคือการควอนไทซ์ที่เหมาะสมเมื่อสมการที่ (2.38) มีค่าน้อยที่สุดจากทุกๆ L ระดับของตัวควอนไทซ์ ซึ่งหมายความว่าต้องเป็นไปตามเงื่อนไขของความเหมาะสมที่จำเป็นทั้ง 2 ประการ ประการแรกคือ ตัวควอนไทซ์ที่เหมาะสมที่สุดต้องเป็นไปตามกฎความเพี้ยนน้อยที่สุด (Minimum Distortion) หรือกฎการเลือกบริเวณที่ใกล้เคียงที่สุด (Nearest Neighbour Rule) ดังนี้

$$q(x)=y_i \text{ if and only if } d(x,y_i)\le d(x,y_j),\ j\neq i,\ 1\le j\le L \tag{2.39}$$

จากสมการที่ (2.39) หมายความว่าตัวควอนไทซ์ต้องเลือกเวกเตอร์รหัส (Codeword) ที่มีความเพี้ยนน้อยที่สุดเมื่อเทียบกับ x

เงื่อนไขที่สองคือ การเลือกเวกเตอร์รหัส y_i ต้องให้ความเพี้ยนเฉลี่ยในเซลล์ C_i มีค่าน้อยที่สุด นั้นคือ y_i เป็นเวกเตอร์ y ที่ทำให้สมการที่ (2.40) มีค่าน้อยที่สุด

$$D_i = E[d(x,y)/x \in C_i] \tag{2.40}$$

$$= \int_{x \in C_i} d(x,y) f(x) dx \tag{2.41}$$

เวกเตอร์ดังกล่าวนี้เรียกว่า "จุดศูนย์ถ่วง" (Centroid) ของเซลล์ C_i ซึ่งเขียนได้เป็น

$$y_i = cent(C_i) \tag{2.42}$$

การคำนวณหาค่าจุดศูนย์ถ่วงของแต่ละบริเวณนั้นขึ้นอยู่กับนิยามของการวัดค่าความเพี้ยน ในทางปฏิบัติต้องกำหนดให้ชุดของเวกเตอร์ฝึกฝน $\{x_k, 1 \le k \le T\}$ โดยที่ชุดย่อย K_i ของเวกเตอร์นี้อยู่ในเซลล์ C_i เพราะฉะนั้นค่าความเพี้ยนเฉลี่ย (D_i) ในเซลล์ C_i เป็นไปตามสมการที่ (2.43)

$$D_i = \frac{1}{K_i} \sum_{x \in C_i} d(x, y_i) \tag{2.43}$$

ค่าความเพี้ยนเฉลี่ยโดยรวมสามารถคำนวณได้จาก

$$D = \sum_{i=1}^{L} D_i / T \tag{2.44}$$

ให้ค่าความเพี้ยนที่วัดได้ในกลุ่มคือ C_i และ $d(x,y)$ คือค่า D_i ที่น้อยที่สุด จากสมการที่ (2.43) เขียนใหม่ได้เป็น

$$y_i = \frac{1}{K_i} \sum_{x \in C_i} x \tag{2.45}$$

เมื่อ y_i คือค่าเฉลี่ยของตัวอย่างที่ได้จากชุดเวกเตอร์ฝึกฝนทั้งหมด (x) ที่อยู่ในกลุ่ม C_i

ขั้นตอนวิธีการแบ่งกลุ่มที่ใช้การวนซ้ำคือขั้นตอนวิธีการหนึ่งที่ใช้ในการออกแบบชุดรหัส เพื่อให้ได้ความเพี้ยนเฉลี่ยน้อยที่สุด (Iteratively Minimise The Average Distortion) เช่น ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน (K-means Algorithm) ขั้นตอนวิธีการ Linde, Buzo และ Gray (LBG Algorithm) และขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุด (Expectation Maximization Algorithm) เป็นต้น (Hunag, 1989) ขั้นตอนในการออกแบบชุดรหัสสามารถแบ่งได้เป็นสองขั้นตอนคือ การสร้างชุดรหัสและการฝึกฝนชุดรหัส เนื่องจากงานวิจัยนี้ได้ใช้ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน และขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุด จึงกล่าวถึงเฉพาะรายละเอียดและทฤษฎีของขั้นตอนวิธีการทั้งสองวิธีนี้เท่านั้น

ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน (K-means Algorithm)

ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนเป็นขั้นตอนวิธีการที่ได้รับความนิยมมากที่สุดในการแบ่งกลุ่มแบบวนซ้ำ (Iterative Clustering Algorithm) (Huang, 1989) ที่ใช้พื้นฐานของการแบ่งชุดข้อมูลฝึกฝนเป็น L กลุ่ม $\{C_i, 1 \le i \le L\}$ ที่ต้องใช้เงื่อนไขที่จำเป็นดังที่กล่าวไว้ข้างต้นในการหาค่าที่เหมาะสม เพื่อให้ได้ค่าเพี้ยนเฉลี่ย (Average Distortion) น้อยที่สุด โดยใช้พื้นฐานการแบ่งชุดของเวกเตอร์ฝึกฝน $\{x(n)\}$ ออกเป็น L กลุ่ม เมื่อกำหนดให้ $L = K$ มีขั้นตอนดังต่อไปนี้

ขั้นตอนการสร้างชุดรหัสเริ่มต้น (Initial Codebook)

ขั้นตอนในการสร้างชุดรหัสมีหลายวิธี (Makhoul, Roucos, and Gish, 1985) ได้แก่ การค้นหาแบบทวิภาค (Binary Search) การควอนไทซ์แบบต่อเรียงกัน (Cascade Quantization) รหัสผลคูณ (Product Codes) และชุดรหัสแบบสุ่ม (Random Codebooks) สำหรับงานวิจัยนี้ใช้วิธีการชุดรหัสแบบสุ่มในการสร้างชุดรหัสเริ่มต้นเพื่อนำไปใช้ในการฝึกฝนชุดรหัสต่อไป เนื่องจากการสร้างชุดรหัสแบบสุ่มมีความรวดเร็วและทำได้ง่ายกว่าวิธีอื่น วิธีการสร้างชุดรหัสแบบนี้คือการสุ่มเลือกเวกเตอร์รหัสเริ่มต้นจากชุดข้อมูลฝึกฝนตามขนาดของชุดรหัส (L)

ขั้นตอนการฝึกฝนชุดรหัส (Codebook Training)

ขั้นตอนนี้เป็นขั้นตอนในการหาค่าเหมาะสมที่สุดจากชุดเวกเตอร์ฝึกฝนมี 4 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การสร้างชุดรหัสเริ่มต้น (Initialisation)

สร้างชุดรหัสแบบสุ่มจากชุดข้อมูลฝึกฝนที่มีชุดเวกเตอร์ของตัวควอนไทซ์เท่ากับ $y_i, 1 \le i \le L$

ขั้นตอนที่ 2 การจำแนก (Classification)

จำแนกข้อมูลในชุดเวกเตอร์ฝึกฝน (Training Vectors) x_k ทีละตัวว่าอยู่ในกลุ่มของ C_i กลุ่มใด โดยใช้กฎการเลือกเพื่อนบ้านข้างเคียงใกล้ที่สุด (Nearest Neighbor Rule) ดังนี้

$$x \in C_i, iff\ d(x, y_i) \le d(x, y_j)\ \ for\ all\ j \ne i \tag{2.46}$$

ขั้นตอนที่ 3 การปรับชุดรหัส (Codebook Updating)

เป็นการปรับค่าเวกเตอร์รหัสในทุกๆ กลุ่มโดยใช้การคำนวณจุดศูนย์ถ่วงของชุดเวกเตอร์ฝึกฝนในแต่ละกลุ่มใหม่ดังนี้

$$y_i = cent(C_i),\ 1 \le i \le L \tag{2.47}$$

ขั้นตอนที่ 4 การสิ้นสุด (Termination)

ถ้าค่าความเพี้ยนโดยรวม D ในรอบปัจจุบันเทียบกับค่าความเพี้ยนโดยรวมในรอบก่อนหน้านี้มีค่าต่ำกว่าจุดเริ่มเปลี่ยน (Threshold) ก็จะสิ้นสุดขั้นตอนวิธีการแต่ถ้าไม่ก็จะวนซ้ำในขั้นตอนที่ 2 ต่อไป

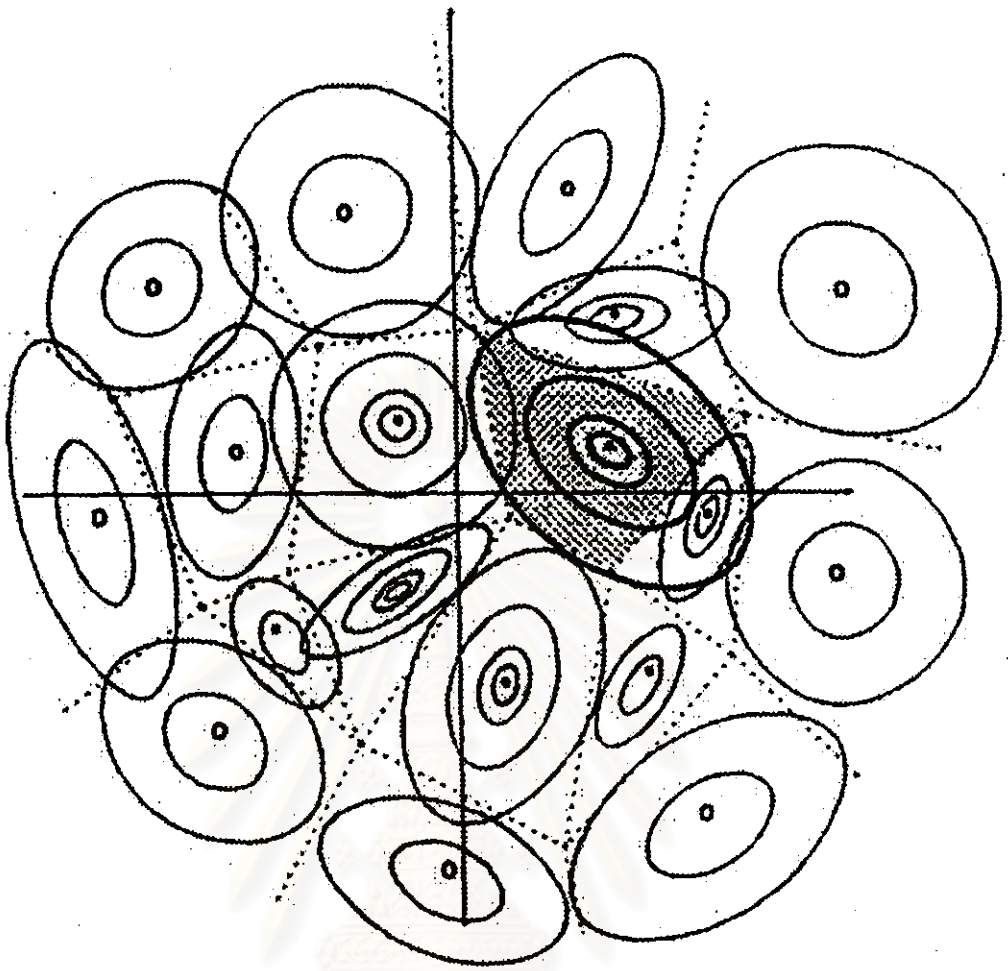
ในกระบวนการหาค่าความเพี้ยนเฉลี่ยต่ำที่สุดของขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนจะสิ้นสุดใน 2 กรณี คือเมื่อได้ค่าศูนย์ถ่วง y_i หรือค่าเฉลี่ยของแต่ละกลุ่มข้อมูล C_i แล้วใช้กระบวนการหาค่าต่ำที่สุด (Minimization Process) เพื่อแบ่งกลุ่มของเวกเตอร์ทั้งหมดที่สอดคล้องกับค่าศูนย์ถ่วงที่อยู่ใกล้ที่สุดโดยใช้การวัดค่าความเพี้ยนเฉลี่ย D_i เป็นตัวตัดสิน หรือในทางตรงกันข้ามถ้าแบ่งกลุ่มของข้อมูลได้ทั้งหมดแล้วใช้กระบวนการหาค่าต่ำที่สุด เพื่อหาจุดศูนย์ถ่วงใหม่ในแต่ละกลุ่มข้อมูลที่ให้ค่าความเพี้ยนเฉลี่ยต่ำสุด

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

D_i จากการวนซ้ำของทั้งสองกรณีทำให้ค่าความเพี้ยนเฉลี่ย D_i ใหม่ที่มีค่าน้อยกว่าค่าในรอบก่อนหน้านี้ อย่างไรก็ตามขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนลู่เข้าสู่ค่าเหมาะสมที่สุดเฉพาะแห่ง (Local Optimum) ทำให้ผลลัพธ์ที่ได้ไม่เป็นหนึ่งเดียว ดังนั้นการหาค่าเหมาะสมที่สุดที่ครอบคลุมโดยรวม (Global Optimum) กระทำได้โดยการใช้ค่าเริ่มต้นของเวกเตอร์รหัสหลายค่าที่แตกต่างกันแล้วทำตามขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน และซ้ำขั้นตอนนี้กับค่าเริ่มต้นที่แตกต่างกันหลายๆ ครั้ง เพื่อให้ชุดรหัสที่มีค่าความเพี้ยนโดยรวมทั้งหมดน้อยที่สุด

ขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุด (Expectation Maximization Algorithm, EM)

ขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุดเป็นขั้นตอนวิธีการแบ่งกลุ่มข้อมูลที่อยู่ในปริภูมิ d มิติ โดยใช้แบบจำลองแบบเกาส์ (Guassian) ในการจำลองการแจกแจงข้อมูลแต่ละกลุ่ม ดังแสดงในรูปที่ 2.8 โดยใช้หลักการพื้นฐานจากการฝึกฝนแบบ Unsupervise ถือว่าข้อมูลที่ได้จากการสุ่มตัวอย่าง ($\boldsymbol{x}$) เป็นข้อมูลที่สังเกตได้ (Observable Data) และถูกเรียกว่า "ข้อมูลไม่บริบูรณ์" (Incomplete Data) เพราะว่ายังขาดข้อมูลที่สังเกตไม่ได้ (Unobservable Data) เพราะฉะนั้นข้อมูลที่มีทั้งข้อมูลที่สังเกตได้กับข้อมูลที่สังเกตไม่ได้เรียกว่า "ข้อมูลบริบูรณ์" (Complete Data) จุดประสงค์ของขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุดคือการหาค่าลอการิทึมของความน่าเป็นจริง (Log-likelihood) ที่มากที่สุดจากข้อมูลไม่สมบูรณ์ โดยใช้การวนซ้ำเพื่อให้การคาดหวังค่าลอการิทึมของความน่าเป็นจริงมากที่สุดจากข้อมูลสมบูรณ์ ในแต่ละรอบของการวนซ้ำของขั้นตอนวิธีการนี้จะประกอบไปด้วย 2 ขั้นตอนคือ ขั้นตอนการคาดหวัง (Expectation Step) กับขั้นตอนการหาค่ามากที่สุด (Maximization) จึงเรียกขั้นตอนวิธีการนี้ว่า "ขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุด" (Expectation Maximization, EM) อาจกล่าวได้ว่าขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุดเป็นวิธีการประมาณค่าความน่าเป็นจริงที่มากที่สุด (Maximum Likelihood Estimation Method) ที่มีการคำนวณความซับซ้อนน้อยและมีความน่าเชื่อถือว่าลู่เข้าค่าเหมาะสมที่สุดครอบคลุมโดยรวม (Huang, 1989; Zhang, Alder & Togneri, 1994; Zhang, 1995)



รูปที่ 2.8 การแบ่งปริภูมิที่มีขนาด 2 มิติโดยใช้แบบจำลองการแจกแจงแบบเกาส์ (Huang, 1989)

ทฤษฎีและหลักการสำคัญที่เกี่ยวข้องกับขั้นตอนวิธีการมีดังต่อไปนี้ การหาฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function) แบบเกาส์หาได้ตามสมการต่อไปนี้

$$f(x)=\frac{|C|^{-1/2}}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right\} \tag{2.48}$$

เมื่อ C คือเมตริกซ์ของความแปรปรวนแบบเกาส์

μ คือเวกเตอร์ค่าเฉลี่ยของข้อมูลที่อยู่แบบจำลองแบบเกาส์

x คือเวกเตอร์ข้อมูลที่เราจะหาความหนาแน่นความน่าจะเป็น

n คือมิติของเวกเตอร์ของข้อมูล x

Multivariate Gaussian Mixture ($g(x|\theta)$) ที่ประกอบไปด้วยการแจกแจงแบบเกาส์ k กลุ่ม สามารถคำนวณหาได้จาก

$$g(x|\theta)=\sum_{i=1}^{k} w_i f_i(x|\mu_i,\Sigma_i) \tag{2.49}$$

เมื่อ $f_i(x|\mu_i,\Sigma_i)$ คือฟังก์ชันความหนาแน่นความน่าจะเป็นแบบเกาส์ที่ i ซึ่งแสดงการแจกแจงข้อมูลที่มีค่าเฉลี่ย (Mean, μ_i) เป็นเวกเตอร์ขนาด d มิติ ความแปรปรวนเป็นเมตริกซ์ขนาด

$d \times d$ มิติ w_i คือค่าถ่วงน้ำหนักการแจกแจงแบบเกาส์ที่ i และ $\sum_{i=1}^{k} w_i = 1$ จากสมการที่ (2.49) θ คือเวกเตอร์ที่ประกอบไปด้วยพารามิเตอร์ w_i, μ_i และ $\sum_i$ เมื่อ $i = 1,2,3,\ldots,k$

$g(x|\theta)$ คือความน่าเป็นจริง (Likelihood) ของ x และใช้บอกระยะระหว่างจุดของข้อมูลกับแบบจำลองแบบเกาส์แต่ละกลุ่ม กำหนดให้ N คือจำนวนของข้อมูลที่มีขนาด d มิติ $X = x_1, x_2, x_3, \ldots, x_N$ สามารถจำลองได้จาก $g(x|\theta)$ ที่มีการแจกแจงแบบเกาส์จำนวน k เกาส์ การวัดว่าการจำลองข้อมูลด้วยการแจกแจงแบบเกาส์เข้าใกล้ข้อมูลมากหรือน้อยวัดได้จากค่าความน่าเป็นจริงโดยรวม (Total Likelihood) ค่าความน่าเป็นจริงของแต่ละกลุ่มคำนวณได้จาก

$$L(X/\theta) = \prod_{j=1}^{N} g(x_j|\theta) \tag{2.50}$$

เพราะฉะนั้นค่าลอการิทึมความน่าเป็นจริงโดยรวม (Log Total Likelihood) คำนวณได้ดังนี้

$$\ell(X/\theta) = \log L(X/\theta) = \sum_{j=1}^{N} \log\ g(x_j|\theta) \tag{2.51}$$

สังเกตได้ว่าปัญหาของการจำลองข้อมูลด้วยการแจกแจงแบบเกาส์คือการหา $\theta = \hat{\theta}$ และค่าลอการิทึมของความน่าเป็นจริงโดยรวมที่มีค่ามากที่สุด ขั้นตอนวิธีการที่ใช้ในการหาค่าประมาณลอการิทึมของความน่าเป็นจริงที่มากที่สุดและสอดคล้องกับการประมาณ $\theta = \hat{\theta}$ ก็คือขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุด เป็นขั้นตอนวิธีการที่ใช้กระบวนการวนซ้ำเพื่อสร้างลำดับ (Sequence) ของการประมาณ $\{\theta^m\}$ กระบวนการนี้สามารถลู่เข้าสู่ค่าลอการิทึมของความน่าเป็นจริงโดยรวมที่เหมาะสมที่สุดเฉพาะแห่ง แต่ถ้าปรับเปลี่ยนเงื่อนไขเริ่มต้น (θ^0) ทำให้การสร้างลำดับลู่เข้าค่า $\hat{\theta}$ และมีค่าลอการิทึมของความน่าเป็นจริงโดยรวมที่แตกต่างกันตามเงื่อนไขเริ่มต้น การแก้ปัญหานี้คือการหาวิธีการเลือกเงื่อนไขเริ่มต้นที่ไม่เปลี่ยนแปลงการประมาณค่า $\hat{\theta}$ และได้ค่าลอการิทึมของความน่าเป็นจริงที่เหมาะสม

วิธีการสร้างชุดรหัสเริ่มต้นให้กับขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุดจะใช้กลยุทธแบบสุนัขกับกระต่าย (Dog-Rabbit Strategy) เพื่อหาจุดศูนย์กลางของกลุ่มข้อมูล

ขั้นตอนการสร้างชุดรหัสเริ่มต้น (Initial Codebook)

การสร้างชุดรหัสเริ่มต้นให้กับขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุดในงานวิจัยนี้ไม่ได้ใช้การสุ่มแต่ได้เลือกใช้ขั้นตอนวิธีการแบบสุนัขกับกระต่าย (Dog Rabbit Algorithm) เนื่องจากวิธีแบบสุนัขกับกระต่ายเป็นวิธีการแบ่งกลุ่มได้ดีกว่าวิธีการแบ่งเฉลี่ย K ส่วน (McKenzie and Alder, 1995) และช่วยลดเวลาที่ใช้ในขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุดให้น้อยลง ขั้นตอนวิธีการแบบสุนัขกับกระต่ายเป็นกระบวนการวนซ้ำแบบพลวัต (Dynamic Process) เพื่อเคลื่อนย้ายจุด k หรือจุดศูนย์กลางของกลุ่มไปใกล้จุดศูนย์กลางของชุดข้อมูล

ขั้นตอนวิธีการแบบสุนัขกับกระต่ายมีทั้งหมด 6 ขั้นตอนดังนี้

ขั้นตอนที่ 1 กำหนดสุนัขเริ่มต้น k ตัว ที่อยู่ในตำแหน่ง c_j เมื่อ $j = 1,2,3,\ldots,k$ ที่ได้จากการสุ่มตำแหน่งจากขอบเขตของข้อมูล และกำหนดค่าความอ่อนล้า (Fatigue) ของสุนัข $f_j = 1$

ขั้นตอนที่ 2 เลือกกระต่าย x แบบสุ่มจากชุดข้อมูล

ขั้นตอนที่ 3 คำนวณหาระยะห่างของสุนัขแต่ละตัวกับกระต่าย x และหาสุนัขตัวที่อยู่ใกล้ที่สุด c_c

ขั้นตอนที่ 4 เคลื่อนย้ายตำแหน่งของสุนัขเข้าหากระต่ายแบบพลวัตดังนี้

$$c_c' = c_c + \frac{2D_c}{(1+D_c)^{f_c}}(x - c_c) \tag{2.52}$$

$$c_j' = c_j + \frac{D_j}{\Lambda + D_j}\frac{2D_j}{(1+D_j)^{f_j}}(x - c_j) \quad , \; j \neq c \tag{2.53}$$

เมื่อ D_c และ D_j คือระยะห่างแบบยูคลิเดียน ($D_j = \|c_j - x\|$)

ขั้นตอนที่ 5 ถ้า $D_c < 1$ ให้เพิ่มค่าความอ่อนล้า f_c ของ c_c ซึ่งคือสุนัขที่ใกล้กระต่ายมากที่สุด

ขั้นตอนที่ 6 วนซ้ำขั้นตอนที่ 2 ใหม่จนกว่าสุนัขจะเคลื่อนย้ายตำแหน่งน้อยกว่าค่าจุดเริ่มเปลี่ยนที่กำหนดไว้

ขั้นตอนการฝึกฝนชุดรหัส (Codebook Training)

ขั้นตอนการฝึกฝนชุดรหัสเป็นขั้นตอนการหาค่าพารามิเตอร์ของแบบจำลองการแจกแจงแบบเกาส์ โดยใช้กระบวนการวนซ้ำเพื่อหาค่าประมาณความน่าเป็นจริงมากที่สุดจากชุดข้อมูล มีขั้นตอนโดยทั่วๆ ไปดังนี้

ขั้นตอนที่ 1 เลือกพารามิเตอร์ θ เริ่มต้นได้แก่ ค่าจุดศูนย์กลางของกลุ่ม เมตริกซ์ความแปรปรวนของกลุ่มข้อมูล และค่าถ่วงน้ำหนักของแต่ละกลุ่มข้อมูลที่ต้องการแบ่งกลุ่มจากวิธีแบบสุนัขกับกระต่าย

ขั้นตอนที่ 2 คำนวณหาค่าคาดหวังจาก

$$Q(\theta, \theta^m) = E\left[\log\ g(Y/\theta)/X, \theta^m\right] \tag{2.54}$$

ขั้นตอนที่ 3 เลือก θ^{m+1} ที่ทำให้ $Q(\theta, \theta^m)$ มีค่ามากที่สุด

ขั้นตอนที่ 4 กำหนดให้ $\theta = \theta^{m+1}$ และวนซ้ำในขั้นตอนที่ 2 จนกว่าจะลู่เข้า

เมื่อ Y คือชุดของข้อมูลบริบูรณ์และมีชุดข้อมูลไม่บริบูรณ์ (X) รวมอยู่ด้วย m คือรอบของการวนซ้ำ และคำนวณค่าพารามิเตอร์ θ ในแต่ละรอบของการวนซ้ำได้จากสมการต่อไปนี้

$$w_i^{m+1} = \frac{1}{N}\sum_{j=1}^{N}\frac{w_i^m f_i(x_j \mid \mu_i^m, \Sigma_i^m)}{g(x_j \mid \theta^m)} \tag{2.55}$$

$$\mu_i^{m+1} = \frac{1}{n_i^m}\sum_{j=1}^{N}\frac{w_i^m f_i(x_j \mid \mu_i^m, \Sigma_i^m)}{g(x_j \mid \theta^m)}\, x_j \tag{2.56}$$

$$\Sigma_i^{m+1} = \frac{1}{n_i^m}\sum_{j=1}^{N}\frac{w_i^m f_i(x_j \mid \mu_i^m, \Sigma_i^m)}{g(x_j \mid \theta^m)}\,(x_j - \mu_i^m)(x_j - \mu_i^m)^T \tag{2.57}$$

โดยที่ $$n_i^m = \sum_{j=1}^{N}\frac{w_i^m f_i(x_j \mid \mu_i^m, \Sigma_i^m)}{g(x_j \mid \theta^m)} \tag{2.58}$$

เมื่อ w_i^m คือค่าถ่วงน้ำหนักของการแจกแจงแบบเกาส์ที่ i ในรอบที่ m

μ_i^m คือค่าเฉลี่ยของการแจกแจงแบบเกาส์ที่ i ในรอบที่ m

Σ_i^m คือค่าความแปรปรวนของการแจกแจงแบบเกาส์ที่ i ในรอบที่ m

n_i^m คือจำนวนของข้อมูลที่อยู่ในการแจกแจงแบบเกาส์ที่ i ในรอบที่ m

x_j คือข้อมูลในชุดฝึกฝน

N คือจำนวนของข้อมูลในชุดฝึกฝน

สมการที่ (2.55) ถึง (2.58) เป็นสมการที่ใช้การคำนวณค่าความน่าเป็นจริงของแต่ละจุดข้อมูลกับค่าพารามิเตอร์ในการจำลองการแจกแจงแบบเกาส์ของแต่ละกลุ่ม $w_i^m f_i(x_j \mid \mu_i^m, \Sigma_i^m)$ สรุปได้ว่าขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุดคำนวณค่าพารามิเตอร์ต่างๆ ของการแจกแจงแบบเกาส์ใหม่จนกว่าจะได้ค่าลอการิทึมความน่าเป็นจริงโดยรวมลู่เข้า ซึ่งหมายถึงความแตกต่างของค่าลอการิทึมความน่าเป็นจริงโดยรวมในรอบการวนซ้ำก่อนหน้านั้นกับค่าลอการิทึมความน่าเป็นจริงโดยรวมในรอบปัจจุบันมีค่าน้อยกว่าจุดเริ่มเปลี่ยนที่กำหนดไว้

2.4 แบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต (Discrete Hidden Markov Models)

แบบจำลองฮิดเดนมาร์คอฟเป็นแบบจำลองทางสถิติ (Statistical Model) คือการใช้วิธีการทางสถิติและความน่าจะเป็นมาใช้ในจำแนกรูปแบบ โดยอาศัยคุณสมบัติทางสถิติของสัญญาณในการแยกแยะสัญญาณ มีงานวิจัยจำนวนมากที่นิยมใช้แบบจำลองฮิดเดนมาร์คอฟแบบดิสครีตในการรู้จำเสียงพูดเนื่องจากเป็นแบบจำลองที่ใช้เทคนิคความน่าจะเป็นจากการศึกษาอนุกรมของเวลาแบบไม่ต่อเนื่องซึ่งตรงกับลักษณะของสัญญาณเสียงพูดที่มีคุณสมบัติเปลี่ยนแปลงตามเวลา (Rabiner, 1989; Huang, 1989; Zhang, 1995)

ค่าความน่าจะเป็นที่ออกจากแบบจำลองนี้เป็นค่าความน่าจะเป็นระหว่างลำดับของสถานะ (State Sequence) ของแบบจำลองกับค่าสังเกตของอนุกรมเวลา ทำให้ผลของการจำลองนี้เป็นกระบวนการเฟ้นสุ่ม (Stochastic Process) ซ้อนกับกระบวนการเฟ้นสุ่มที่ไม่สามารถสังเกตได้ เนื่องจากว่าลำดับของสถานะไม่สามารถสังเกตได้จึงเป็นที่มาของคำว่า "ฮิดเดน" (Hidden)

แบบจำลองฮิดเดนมาร์คอฟได้ใช้พื้นฐานของลูกโซ่มาร์คอฟ (Markov Chain) ในการจำลองการเปลี่ยนแปลงค่าทางสถิติที่ได้จากค่าสังเกตของสัญญาณเสียงพูด แบบจำลองฮิดเดนมาร์คอฟจึงเป็นแบบจำลองที่ใช้พารามิเตอร์ในการจำลอง (Parametric Model) ขั้นตอนกระบวนการในแบบจำลองฮิดเดนมาร์คอฟเรียกว่า "กระบวนการมาร์คอฟ" (Markov Process) แบบจำลองฮิดเดนมาร์คอฟสามารถแบ่งตามลักษณะการนิยามการแจกแจงความน่าจะเป็นของสัญลักษณ์ของค่าสังเกตได้เป็น 3 ชนิด (Huang, 1992) คือแบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต (Discrete Hidden Markov Models, DHMM) แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง (Continuous Hidden Markov Models, CHMM) และแบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่อง (Semi-Continuous Hidden Markov Models, SCHMM) ซึ่งมีลักษณะเด่นและด้อยแตกต่างกันไปตามแต่ลักษณะการใช้งาน เนื่องจากงานวิจัยนี้ได้เลือกใช้แบบจำลองฮิดเดนมาร์คอฟแบบดิสครีตจึงขอกล่าวถึงเฉพาะนิยามและขั้นตอนวิธีการที่เกี่ยวข้องกับแบบจำลองนี้เท่านั้น

2.4.1 กระบวนการมาร์คอฟ

กระบวนการมาร์คอฟคือกระบวนการเฟ้นสุ่มซ้อนกระบวนการเฟ้นสุ่ม ที่ไม่สามารถสังเกตลูกโซ่มาร์คอฟได้ . โดยนิยามให้เป็นเมตริกซ์ของการเปลี่ยนแปลงสถานะ และแต่ละสถานะของลูกโซ่มาร์คอฟจะถูกแทนด้วยการแจกแจงความน่าจะเป็นขาออกแบบดิสครีต (Discrete Output Probability Distribution) ดังนั้นกระบวนการเฟ้นสุ่มถูกอธิบายโดยกระบวนการมาร์คอฟลำดับที่ j เพื่อแสดงลักษณะบางอย่างที่มีความน่าจะเป็นขนาดใหญ่ที่มักเกิดขึ้นในภาษาที่ใช้พูด ซึ่งถูกรวบรวมโดยคุณสมบัติของมาร์คอฟ (Markov Property) คือทุกๆ ลำดับของเหตุการณ์ในโดเมนทางเวลาขึ้นอยู่กับเงื่อนไขความหนาแน่นของความน่าจะเป็นของเหตุการณ์ในปัจจุบันกับเหตุการณ์ในอดีตทั้งหมด และเหตุการณ์ในปัจจุบันขึ้นอยู่กับเหตุการณ์ที่เกิดขึ้นบ่อยที่สุด j ครั้งเท่านั้น กระบวนการที่เป็นไปตามคุณสมบัติมาร์คอฟถูกเรียกว่า "กระบวนการมาร์คอฟ"

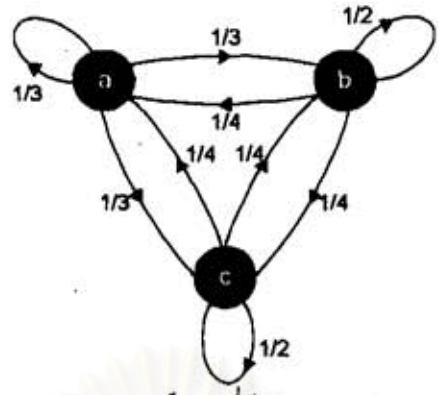
ยกตัวอย่างของกระบวนการมาร์คอฟลำดับที่หนึ่งอย่างง่าย กำหนดให้สัญลักษณ์ 3 ตัวอักษร คือ a, b, c และให้ค่าความน่าจะเป็นของสัญลักษณ์ a ที่ตามด้วยทุกตัวอักษรเป็น 1/3 ให้ค่าความน่าจะเป็นของสัญลักษณ์ b ที่ตามด้วย b เป็น 1/2 และตามด้วยสัญลักษณ์อื่นๆ เป็น 1/4 และให้ค่าความน่าจะเป็นของสัญลักษณ์ c ที่ตามด้วย c เป็น 1/2 และตามด้วย a, b เป็น 1/4 เพราะฉะนั้นสรุปได้ว่า

$$\Pr(a \mid a) = \frac{1}{3},\ \Pr(b \mid a) = \frac{1}{3},\ \Pr(c \mid a) = \frac{1}{3}$$

$$\Pr(a \mid b) = \frac{1}{4},\ \Pr(b \mid b) = \frac{1}{2},\ \Pr(c \mid b) = \frac{1}{4}$$

$$\Pr(a \mid c) = \frac{1}{4},\ \Pr(b \mid c) = \frac{1}{4},\ \Pr(c \mid c) = \frac{1}{2}$$

และเขียนเป็นกราฟการเปลี่ยน (Transition Graph) เพื่อแสดงให้เห็นกระบวนการมาร์คอฟดังในรูปที่ 2.9 ในกรณีนี้เห็นได้ว่ามีสถานะของแบบจำลองเท่ากับ 3 คือ a, b, c ที่เป็นรูปวงกลม



รูปที่ 2.9 กระบวนการมาร์คอฟที่มี 3 สถานะ (Zhang, 1995)

จากรูปที่ 2.9 เส้นแต่ละเส้นคือการเปลี่ยนจากสถานะหนึ่งไปสถานะอื่นที่มีค่าความน่าจะเป็นของการเปลี่ยนคือตัวเลขที่อยู่ข้างๆ เส้น ตัวอย่างเช่น $\Pr(a|b)$ คือเส้นที่ลากจากสถานะ b ไปยังสถานะ a และมีความน่าจะเป็นของการเปลี่ยนเท่ากับ 1/4 จากรูปเห็นได้ว่าแต่ละสถานะมีเส้นลากออกและลากเข้าอย่างละ 3 เส้น ถ้าใช้แบบจำลองมาร์คอฟนี้ในการทำนายสภาพอากาศ โดยกำหนดให้สถานะ a มีข้อมูลออกเป็นแดดออก สถานะ b มีข้อมูลออกเป็นเมฆมากและสถานะ c มีข้อมูลออกเป็นฝนตก ถ้าให้วันนี้เป็นวันที่แดดออกหมายความว่าโอกาสที่พรุ่งนี้สภาพอากาศจะมีแดดออก เมฆมากหรือฝนตกมีโอกาสเท่ากันๆ แต่ถ้าวันนี้สภาพอากาศเป็นเมฆมากหรือฝนตกโอกาสที่พรุ่งนี้จะมีสภาพอากาศเหมือนเดิมคือร้อยละ 50 และมีโอกาสเพียง 1 ใน 4 ที่วันพรุ่งนี้จะมีสภาพอากาศเป็น 2 อย่างที่เหลือ ดังนั้นกระบวนการเฟ้นสุ่มของการทำนายสภาพอากาศสามารถอธิบายได้อย่างคร่าวๆ ด้วยกระบวนการมาร์คอฟ

สมมติให้สถานะ a, b และ c ถูกแทนด้วยหมายเลข 1 2 และ 3 ฉะนั้นกราฟการเปลี่ยนในรูปที่ 2.9 สามารถเขียนให้อยู่ในรูปเมตริกซ์ เมื่อสมาชิกของเมตริกซ์ a_{ij} คือความน่าจะเป็นของการเปลี่ยนจากสถานะปัจจุบัน i ไปยังสถานะต่อไป j เมตริกซ์การเปลี่ยน (A) คือ

$$A = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \tag{2.59}$$

จากสมการที่ (2.59) แสดงให้เห็นว่าการบวกสมาชิกในแต่ละแถวภายในเมตริกซ์การเปลี่ยนจะได้เท่ากับ 1 เสมอ

สรุปได้ว่าแบบจำลองมาร์คอฟใช้การศึกษาค่าสังเกตที่อยู่ในรูปของอนุกรมของเวลาแบบไม่ต่อเนื่อง ลำดับของสถานะถูกสังเกตได้จากแบบจำลองมาร์คอฟ เช่น สภาพอากาศ (แดดออก เมฆมาก หรือฝนตก) เป็นต้น อย่างไรก็ตามแบบจำลองถูกจำกัดให้ใช้กับเฉพาะปัญหาที่สนใจเท่านั้น ในแบบจำลองฮิดเดนมาร์คอฟมีข้อมูลออกของแต่ละสถานะสอดคล้องกับการแจกแจงความน่าจะเป็นที่ข้อมูลออก ซึ่งถูกแทนที่ด้วยเหตุการณ์ที่จะเกิดขึ้น ความน่าจะเป็นของข้อมูลออกได้มาจากลำดับสถานะของแบบจำลองกับอนุกรมทางเวลาของค่าสังเกต โดยลำดับของสถานะไม่สามารถสังเกต

2.4.2 นิยามของแบบจำลองฮิดเดนมาร์คอฟ

เพื่อให้อธิบายแบบจำลองฮิดเดนมาร์คอฟและให้คำจำกัดความของค่าพารามิเตอร์ต่างๆ ดังต่อไปนี้

(1) N คือจำนวนของสถานะที่อยู่ในแบบจำลอง

(2) L คือจำนวนสัญลักษณ์ของค่าสังเกต

(3) S คือเซตของสถานะ $\{s\}$ เช่นสถานะ i ที่เวลา t สามารถเขียนได้ดังนี้ $s_t = i$

(4) $\upsilon = \{\upsilon_1, \upsilon_2, \upsilon_3, \ldots, \upsilon_L\}$ คือเซตแบบไม่ต่อเนื่องของสัญลักษณ์ของค่าสังเกตที่เป็นไปได้ ซึ่ง O_t คือสัญลักษณ์ของค่าสังเกต

(5) A คือการแจกแจงความน่าจะเป็นในการเปลี่ยนแปลงสถานะ (State Transition Probability Distribution)

$$A = \{a_{ij} \mid a_{ij} = \Pr(s_{t+1} = j \mid s_t = i)\}, \quad 1 \le i, j \le N \tag{2.60}$$

เมื่อ a_{ij} คือความน่าจะเป็นในการเปลี่ยนสถานะ i ไปสถานะ j

(6) B คือการแจกแจงความน่าจะเป็นของสัญลักษณ์ของค่าสังเกต (Observation Symbol Probability Distribution)

$$\mathrm{B} = \{\mathrm{b_j(O_t)} \mid \mathrm{b_j(O_t) = Pr(O_t \mid s_t = j)}\} \tag{2.61}$$

แต่ละสถานะ B จะสอดคล้องกับความน่าจะเป็นของข้อมูลออก ในแบบจำลองฮิดเดนมาร์คอฟแบบดิสครีตหมายถึงความน่าจะเป็นที่ได้จากสัญลักษณ์ที่เป็นแบบไม่ต่อเนื่อง υ_k ในสถานะ j ซึ่งเขียนใหม่ได้เป็น

$$b_j(k) = \Pr(\upsilon_k \mid s_t = j) \qquad \text{เมื่อ } 1 \le j \le N \text{ และ } 1 \le k \le L \tag{2.62}$$

(7) π คือการแจกแจงสภาวะเริ่มต้น (Initial State Distribution)

$$\pi = \{\pi_i \mid \pi_i = \Pr(s_1 = i)\}, \ 1 \le i \le N \tag{2.63}$$

โดยการกำหนดค่าที่เหมาะสมให้กับพารามิเตอร์ N, L, A, B, π ของแบบจำลองฮิดเดนมาร์คอฟที่ใช้ในการกำเนิดลำดับค่าสังเกต

$$O = O_1, O_2, O_3, \ldots, O_T \tag{2.64}$$

เมื่อแต่ละค่าสังเกต O_t เป็นสัญลักษณ์ที่ได้จาก υ และ T เป็นจำนวนค่าสังเกตทั้งหมดที่มีในลำดับ โดยมีขั้นตอนวิธีการกำเนิดค่าสังเกตดังนี้

ขั้นตอนที่ 1 เลือกสถานะเริ่มต้น $s_1 = i$ ที่สัมพันธ์กับการแจกแจงสภาวะเริ่มต้น π

ขั้นตอนที่ 2 กำหนดให้ $t = 1$

ขั้นตอนที่ 3 เลือก $O_t = \upsilon_k$ ที่สัมพันธ์กับการแจกแจงความน่าจะเป็นของสัญลักษณ์ในสถานะ i เช่น $b_i(k)$

ขั้นตอนที่ 4 เปลี่ยนไปยังสถานะใหม่ $s_{t+1} = j$ ที่สัมพันธ์กับการแจกแจงความน่าจะเป็นในการเปลี่ยนแปลงสถานะสำหรับสถานะ i

ขั้นตอนที่ 5 กำหนดให้ $t = t+1$ แล้วทวนซ้ำขั้นตอนที่ 3 ใหม่ถ้า $t < T$ นอกเหนือจากนี้ให้ยุติกระบวนการ

ขั้นตอนดังกล่าวนี้สามารถใช้เพื่อกำเนิดค่าสังเกตและเป็นแบบจำลองเพื่อบอกถึงความเหมาะสมในการกำเนิดลำดับค่าสังเกตด้วยแบบจำลองฮิดเดนมาร์คอฟ ดังนั้นการกำหนดคุณสมบัติเฉพาะของแบบจำลองฮิดเดนมาร์คอฟต้องการคุณสมบัติเฉพาะของพารามิเตอร์ของแบบจำลองสองค่า (N และ L) คือคุณสมบัติเฉพาะของสัญลักษณ์ของค่าสังเกต และคุณสมบัติเฉพาะของการวัดค่าความน่าจะเป็นได้แก่ A, B และ π โดยเขียนอยู่ในรูปแบบอย่างย่อเพื่อบ่งบอกชุดของพารามิเตอร์ที่สมบูรณ์ของแบบจำลองดังนี้

$$\lambda = (A, B, \pi) \tag{2.65}$$

2.4.3 ขั้นตอนวิธีการพื้นฐานของแบบจำลองฮิดเดนมาร์คอฟ

จากนิยามของแบบจำลองฮิดเดนมาร์คอฟในหัวข้อที่แล้ว เห็นได้ว่าการประยุกต์แบบจำลองเพื่อใช้งานนั้นต้องแก้ปัญหาพื้นฐานสามข้อที่เกิดขึ้น โดยมีรายละเอียดของปัญหาดังนี้

(1) ปัญหาการประเมินค่า (Evaluation Problem) โดยกำหนดให้ลำดับค่าสังเกต $O = O_1, O_2, O_3, \ldots, O_T$ และพารามิเตอร์ของแบบจำลอง $\lambda = (A, B, \pi)$ เพื่อคำนวณหาค่าความน่าจะเป็นของลำดับค่าสังเกต $\Pr(O \mid \lambda)$ ตามแบบจำลองที่กำหนดให้ได้อย่างไร

(2) ปัญหาการประมาณค่า (Estimation Problem) กำหนดให้ลำดับค่าสังเกต O จะปรับค่าพารามิเตอร์ของแบบจำลอง $\lambda = (A, B, \pi)$ อย่างไรเพื่อให้ค่าความน่าจะเป็นของลำดับค่าสังเกต $\Pr(O \mid \lambda)$ มีค่ามากที่สุด

(3) ปัญหาการถอดรหัส (Decoding Problem) กำหนดให้ลำดับค่าสังเกต O และพารามิเตอร์ของแบบจำลอง $\lambda = (A, B, \pi)$ แล้วเลือกลำดับสถานะ $S = S_1, S_2, S_3, \ldots, S_T$ ให้สอดคล้องและเหมาะสมที่สุดกับแบบจำลองที่กำหนดให้ได้อย่างไร

ปัญหาข้อที่ 1 เป็นการหาค่าความน่าจะเป็นของลำดับค่าสังเกตของแบบจำลอง โดยกำหนดให้ลำดับค่าสังเกตและพารามิเตอร์ของแบบจำลองมาให้ ซึ่งเป็นการแสดงว่าแบบจำลองที่กำหนดให้เข้าคู่กับลำดับค่าสังเกตที่ให้มาได้ดีแค่ไหน ตัวอย่างเช่น ในกรณีที่พิจารณาเลือกระหว่างแบบจำลองหลายแบบกับลำดับค่าสังเกตที่ให้มา การแก้ปัญหาข้อที่ 1 จะทำให้เราได้แบบจำลองที่เข้าคู่กับลำดับค่าสังเกตได้ดีที่สุด

ปัญหาข้อที่ 2 คือการทำให้พารามิเตอร์ของแบบจำลองเหมาะสมที่สุด เพื่ออธิบายลำดับค่าสังเกตได้ดีที่สุด ลำดับค่าสังเกตที่ใช้ในการปรับพารามิเตอร์ของแบบจำลองเรียกว่า "ลำดับฝึกฝน" (Training Sequence) เนื่องจากถูกใช้ในการฝึกฝนแบบจำลองฮิดเดนมาร์คอฟ การแก้ปัญหาในการฝึกฝนนี้จะช่วยให้ปรับแต่งพารามิเตอร์ของแบบจำลองให้เหมาะสมมากที่สุดกับข้อมูลฝึกฝนที่สังเกตได้

ปัญหาข้อที่ 3 เป็นความพยายามที่จะเปิดเผยส่วนที่แบบจำลองปิดบังไว้ เพื่อหาลำดับสถานะที่ถูกต้อง ในกรณีของแบบจำลองที่ด้อยประสิทธิภาพทำให้ไม่สามารถหาลำดับสถานะที่ถูกต้องได้ ดังนั้นในทางปฏิบัติจึงต้องใช้กฎเกณฑ์ที่เหมาะสมที่สุดในการแก้ไขปัญหานี้เท่าที่จะทำได้

ขั้นตอนวิธีการไปหน้าและย้อนกลับ (Forward-Backward Algorithm)

ขั้นตอนวิธีการไปหน้าและย้อนกลับเป็นวิธีการคำนวณหาความน่าจะเป็นของค่าสังเกตอย่างตรงไปตรงมาที่สุด โดยหาค่าทุกลำดับสถานะที่มีความยาว T ที่เป็นไปได้ทั้งหมดตามลำดับ เมื่อ T คือจำนวนของค่าสังเกต สมมติให้ลำดับสถานะมีจำกัดดังนี้

$$S = s_1, s_2, s_3, \ldots, s_T \tag{2.66}$$

ความน่าจะเป็นของลำดับค่าสังเกต O คือ

$$\Pr(O \mid S, \lambda) = b_{s_1}(O_1) b_{s_2}(O_2) \ldots b_{s_T}(O_T) \tag{2.67}$$

ความน่าจะเป็นของแต่ละสถานะ S คือ

$$\begin{aligned} \Pr(S \mid \lambda) &= \pi_{s_1} a_{s_1 s_2} a_{s_2 s_3} \cdots a_{s_{T-1} s_T} \\ &= a_{s_0 s_1} a_{s_1 s_2} a_{s_2 s_3} \cdots a_{s_{T-1} s_T} \end{aligned} \tag{2.68}$$

เมื่อให้ π_{s_1} แทนด้วย $a_{s_0 s_1}$

ความน่าจะเป็นร่วมระหว่าง O และ S เป็นผลคูณของสมการ 2 สมการข้างต้นดังนี้

$$\Pr(O, S \mid \lambda) = \Pr(O \mid S, \lambda) \Pr(S \mid \lambda) \tag{2.69}$$

ความน่าจะเป็น $\Pr(O \mid \lambda)$ คือผลรวมของสมการที่ (2.69) ของลำดับสถานะที่เป็นไปได้ทั้งหมดดังนี้

$$\begin{aligned} \Pr(O \mid \lambda) &= \sum_{all\ S} \Pr(O \mid S, \lambda) \Pr(S \mid \lambda) \\ &= \sum_{all\ S} \prod_{t=1}^{T} a_{s_{t-1} s_t} b_{s_t}(O_t) \end{aligned} \tag{2.70}$$

จากสมการที่ (2.70) เป็นการเริ่มการเปลี่ยนจากสถานะเริ่มต้นที่เวลา $t = 1$ มีความน่าจะเป็น $a_{s_0 s_1}(\pi_{s_1})$ กำเนิดสัญลักษณ์ O_1 และมีความน่าจะเป็นขาออกคือ $b_{s_1}(O_1)$ โดยสอดคล้องกับสถานะ s_1 ไปยังสถานะ s_2 ที่มีความน่าจะเป็นการเปลี่ยนคือ $a_{s_1 s_2}$ กำเนิดสัญลักษณ์ O_2 และมีความน่าจะเป็นขาออกคือ $b_{s_2}(O_2)$ ที่สอดคล้องกับสถานะ s_2 จะกระทำต่อไปเรื่อยๆ จนถึงการเปลี่ยนสุดท้ายจากสถานะ s_{T-1} ไปยังสถานะ s_T ที่มีความน่าจะเป็นการเปลี่ยนคือ $a_{s_{T-1} s_T}$ และความน่าจะเป็นขาออกคือ $b_{s_T}(O_T)$ กำเนิดสัญลักษณ์ O_T

ภาระในกระบวนการคำนวณที่กล่าวมานั้นมีประมาณ $O(N^T)$ ครั้ง ที่ทุกๆ เวลา $t = 1,2,\ldots,T$ และ N คือสถานะที่เป็นไปได้ที่จะเข้าถึง ขั้นตอนวิธีการที่มีประสิทธิภาพในการคำนวณจากคุณลักษณะดังกล่าวเรียกว่า "ขั้นตอนวิธีการไปหน้าและย้อนกลับ" แบ่งได้เป็นสองขั้นตอนคือขั้นตอนวิธีการไปหน้าและขั้นตอนวิธีการย้อนกลับมีรายละเอียดดังต่อไปนี้

กำหนดให้ตัวแปรไปหน้า $\alpha_t(i)$ คำนวณหาได้ตามสมการที่ (2.71)

$$\alpha_t(i) = \Pr(O_1, O_2, \ldots, O_t, s_t = i \mid \lambda) \tag{2.71}$$

$\alpha_t(i)$ คือความน่าจะเป็นของลำดับค่าสังเกตบางส่วนและสถานะ i ที่เวลา t เมื่อกำหนดพารามิเตอร์ของแบบจำลอง λ มาให้ สามารถคำนวณหา $\alpha_t(i)$ ได้ดังต่อไปนี้

ขั้นตอนวิธีการไปหน้า

ขั้นตอนที่ 1 กระบวนการเริ่มต้น

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \le i \le N \tag{2.72}$$

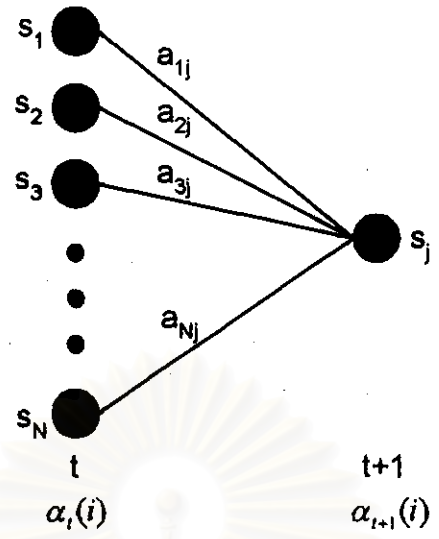
ขั้นตอนที่ 2 กระบวนการอุปนัย

$$\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i)\, a_{ij} \right] b_j(O_t), \quad 2 \le t \le T, \quad 1 \le i \le N \tag{2.73}$$

ขั้นตอนที่ 3 กระบวนการสิ้นสุด

$$\Pr(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{2.74}$$

จากกระบวนการข้างต้นนี้ ขั้นตอนแรกเป็นการกำหนดความน่าจะเป็นเริ่มต้นให้กับทุกๆ สถานะของแบบจำลอง และสมการที่ (2.73) แสดงถึงสถานะ j ที่สามารถเข้าถึงได้ที่เวลา t จากสถานะที่เป็นไปได้ i สถานะที่เวลา $t-1$ ดังแสดงในรูปที่ 2.10 เมื่อ $\alpha_{t-1}(i)$ คือความน่าจะเป็นร่วมของค่าสังเกต $O_1, O_2, \ldots, O_{t-1}$ ที่สถานะ t ดังนั้นผลคูณของ $\alpha_{t-1}(i)a_{ij}$ คือความน่าจะเป็นร่วมของค่าสังเกต $O_1, O_2, \ldots, O_{t-1}$ ที่เข้าถึงสถานะ j ที่เวลา t และผ่านสถานะ i ที่เวลา $t-1$ ผลรวมของผลคูณของสถานะที่เป็นไปได้ทั้งหมด (i) ที่เวลา $t-1$ ได้ผลลัพธ์เป็นความน่าจะเป็นในการเข้าถึงสถานะ j ที่เวลา t โดยผ่านค่าสังเกตบางส่วนก่อนหน้านั้นทั้งหมด และคูณด้วยความน่าจะเป็นของค่าสังเกต $(b_j(O_t))$ ที่เข้าถึงสถานะ j ที่กำเนิด O_t ผลของ $\alpha_t(i)$ คือความน่าจะเป็นของลำดับค่าสังเกตใหม่ $O_1, O_2, \ldots, O_{t-1}$ ที่เวลา t และสถานะ j



รูปที่ 2.10 รายละเอียดของตัวแปรในการคำนวณค่าตัวแปรไปหน้า $\alpha_t(i)$ (Rabiner, 1989)

ในขั้นตอนที่ 3 เป็นการคำนวณหา $\Pr(O \mid \lambda)$ ซึ่งคือผลบวกตัวแปรไปหน้าสุดท้ายที่สถานะสุดท้าย เพราะว่า $\alpha_T(i) = \Pr(O_1 O_2 \ldots O_T, s_T = i \mid \lambda)$ และการเปลี่ยนสถานะจะสิ้นสุดที่ S_F การคำนวณหา $\alpha_t(i)$ ด้วยขั้นวิธีการไปหน้าใช้คำนวณเพียง $O(N^2T)$ ครั้งเท่านั้น

ในลักษณะเดียวกันกำหนดให้ตัวแปรย้อนกลับ $\beta_t(i)$ คำนวณหาได้ตามสมการที่ (2.75)

$$\beta_t(i) = \Pr(O_{t+1}, O_{t+2}, \ldots, O_T, s_t = i \mid \lambda) \tag{2.75}$$

ความน่าจะเป็นของลำดับค่าสังเกตบางส่วนจาก $t+1$ ไปสิ้นสุดการสังเกตที่ T เมื่อกำหนดให้สถานะ i ที่เวลา t และพารามิเตอร์ของแบบจำลอง λ จะสามารถหา $\beta_t(i)$ โดยอุปนัยคล้ายๆ กับการหาตัวแปรไปหน้า $\alpha(\ )$ ได้ดังต่อไปนี้

ขั้นตอนวิธีการย้อนกลับ

ขั้นตอนที่ 1 กระบวนการเริ่มต้น

$$\beta_T(i) = 1, \quad 1 \le i \le N \tag{2.76}$$

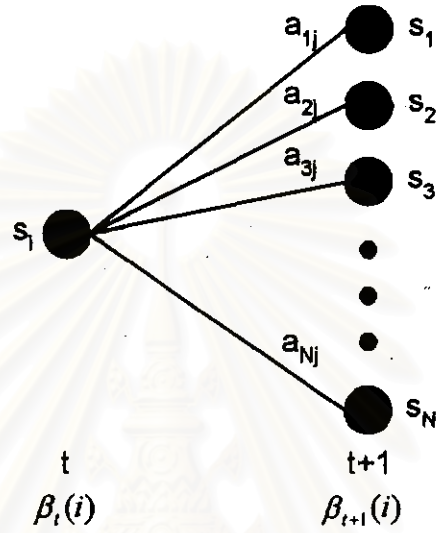
ขั้นตอนที่ 2 กระบวนการอุปนัย

$$\beta_t(j) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \ldots, 1\ , \quad 1 \le i \le N \tag{2.77}$$

ขั้นตอนที่ 3 กระบวนการสิ้นสุด

$$\Pr(O \mid \lambda) = \sum_{i=1}^{N} \pi_i b_i(O_i) \beta_1(i) \tag{2.78}$$

ในขั้นตอนที่ 1 เป็นการกำหนดค่าเริ่มต้นให้กับ $\beta_T(i)$ มีค่าเท่ากับ 1 ทุกสถานะ i ในขั้นตอนที่ 2 เป็นการหา $\beta_t(i)$ ในสถานะ j ที่เวลา t และเปลี่ยนจากสถานะ j ไปทุกๆ สถานะที่เป็นไปได้ที่เวลา $t+1$ มีสัญลักษณ์ค่าสังเกต O_{t+1} และลำดับค่าสังเกตที่สอดคล้องกับสถานะ ความซับซ้อนในการคำนวณหา $\beta_t(i)$ เหมือนกับการคำนวณหา $\alpha_t(i)$



รูปที่ 2.11 รายละเอียดของตัวแปรในการคำนวณค่าตัวแปรย้อนกลับ $\beta_t(i)$ (Rabiner, 1989)

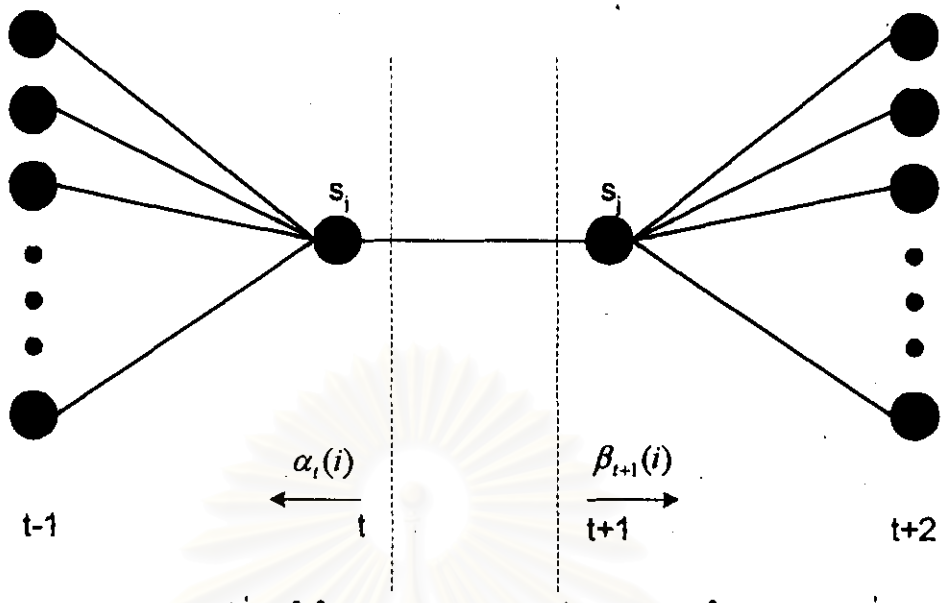
จากที่กล่าวมาทั้งขั้นตอนวิธีการไปหน้าและย้อนหลังใช้เพื่อคำนวณหา $\Pr(O|\lambda)$ เพื่อแก้ปัญหาในข้อที่ 1

ขั้นตอนวิธีการประมาณค่าซ้ำของ Buam-Welch (Buam-Welch Reestimation Algorithm)

ปัญหาที่ยากที่สุดในแบบจำลองฮิดเดนมาร์คอฟคือทำอย่างไรจึงจะปรับค่าพารามิเตอร์ของแบบจำลอง $\lambda=(A,B,\pi)$ เพื่อทำให้ความน่าจะเป็นของลำดับค่าสังเกตมีค่ามากที่สุดเมื่อกำหนดค่าพารามิเตอร์ของแบบจำลองมาให้ เนื่องจากไม่มีวิธีการที่แน่นอนในหาความน่าจะเป็นดังกล่าวให้ได้มากที่สุด ดังนั้นจึงได้มีการนำขั้นตอนวิธีการแบบวนซ้ำหรือเทคนิคเกรเดียนต์ (Gradient Technique) เพื่อหาค่าที่เหมาะสม ขั้นตอนวิธีการวนซ้ำที่นำมาใช้กับแบบจำลองฮิดเดนมาร์คอฟคือขั้นตอนวิธีการของ Buam-Welch ซึ่งเป็นเทคนิคการหาค่าเหมาะสมเหมือนกับขั้นตอนวิธีการหาค่าคาดหวัง (EM Algorithm) นั่นเอง

กำหนดให้ $\xi_t(i,j)$ เป็นความน่าจะเป็นของการอยู่ในสถานะ i ที่เวลา t และสถานะ j ที่เวลา $t+1$ เมื่อกำหนดให้แบบจำลองและลำดับค่าสังเกตมาให้สามารถหาได้ดังสมการที่ (2.79)

$$\xi_t(i,j)=\Pr(s_t=i,s_{t+1}=j|O,\lambda) \tag{2.79}$$



รูปที่ 2.12 ลำดับของเหตุการณ์ที่ต้องใช้ในการคำนวณเหตุการณ์ร่วมของระบบในสถานะ i ที่เวลา t และสถานะ j ที่เวลา $t+1$ (Rabiner, 1989)

ลำดับของเหตุการณ์ที่นำไปสู่เงื่อนไขที่จำเป็นในสมการที่ (2.79) แสดงในรูปที่ 2.12 จากนิยามของตัวแปรไปหน้าและย้อนกลับสามารถเขียน $\xi_t(i,j)$ ได้ดังนี้

$$\xi_t(i,j) = \frac{\alpha_t(i)b_j(O_{t+1})\beta_{t+1}(j)}{\Pr(O \mid \lambda)}$$
$$= \frac{\alpha_t(i)b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)b_j(O_{t+1})\beta_{t+1}(j)} \qquad (2.80)$$

เมื่อพจน์เศษเป็น $\Pr(s_t = i, s_{t+1} = j, O \mid \lambda)$ และหารด้วย $\Pr(O \mid \lambda)$

กำหนดให้ $\gamma_t(i)$ เป็นความน่าจะเป็นของการอยู่ในสถานะ i ที่เวลา t เมื่อกำหนดลำดับค่าสังเกตและแบบจำลองมาให้ ดังนั้นความสัมพันธ์ระหว่าง $\gamma_t(i)$ และ $\xi_t(i,j)$ เกิดจากผลรวมในสถานะ j ทั้งหมดดังนี้

$$\gamma_t(i) = \sum_{j=1}^{N}\xi_t(i,j) \qquad (2.81)$$

ถ้าผลรวมของ $\gamma_t(i)$ ในทางเวลาเป็นจำนวนของเวลาที่คาดว่าจะมีเข้าไปในสถานะ i หรือเป็นจำนวนการเปลี่ยนที่คาดว่าจะมีจากสถานะ i ผลรวมของ $\xi_t(i,j)$ บน t คือ $t=1$ ถึง $t=T-1$ เป็นจำนวนการเปลี่ยนที่คาดว่าจะมีจากสถานะ i ไปยังสถานะ j ดังนี้

$\sum_{t=1}^{T-1}\gamma_t(i)$ คือจำนวนการเปลี่ยนที่คาดว่าจะมีจากสถานะ i (2.82)

$\sum_{t=1}^{T-1}\xi_t(i,j)$ คือจำนวนการเปลี่ยนที่คาดว่าจะมีจากสถานะ i ไปยังสถานะ j (2.83)

ใช้สมการทั้งสองที่กล่าวมาเพื่อหาวิธีการประมาณค่าพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟที่ไปประกอบด้วย π, A และ B ได้ดังนี้

$$\bar{\pi}_i = \gamma_1(i) \tag{2.84}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{2.85}$$

$$\bar{b}_j(k) = \frac{\sum_{t=1 \text{ s.t. } O_t = v_k}^{T-1} \gamma_t(j)}{\sum_{t=1}^{T-1} \gamma_t(j)} \tag{2.86}$$

ถ้ากำหนดให้แบบจำลองในปัจจุบันคือ $\lambda = (A, B, \pi)$ เพื่อใช้คำนวณค่าทางด้านขวาของสมการที่ (2.86) ถึง (2.88) และกำหนดให้แบบจำลองที่ประมาณค่าใหม่คือ $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$ ที่อยู่ด้านซ้ายของสมการที่ (2.86) ถึง (2.88) นอกจากนี้สมการทั้งสามยังเป็นไปตามเงื่อนไขดังนี้ แบบจำลองเริ่มต้น λ เป็นตัวกำหนดจุดวิกฤตของฟังก์ชันความน่าเป็นจริงในกรณีที่ $\bar{\lambda} = \lambda$ หรือแบบจำลอง $\bar{\lambda}$ มีความน่าเป็นจริงมากกว่าแบบจำลอง λ เมื่อ $\Pr(O|\bar{\lambda}) > \Pr(O|\lambda)$ หมายความว่าลำดับค่าสังเกตที่มีความเป็นจริงมากกว่าจะให้กำเนิดแบบจำลองใหม่ $\bar{\lambda}$

จากกระบวนการข้างต้น ถ้ากระทำซ้ำโดยนำ $\bar{\lambda}$ แทนที่ λ และคำนวณการประมาณค่าใหม่เพื่อเป็นการปรับปรุงความน่าจะเป็นของ O ที่ถูกสังเกตจากแบบจำลองจนถึงจุดสิ้นสุดผลลัพธ์สุดท้ายของขั้นตอนวิธีการประมาณค่านี้เรียกว่า "การประมาณความน่าเป็นจริงสูงสุดของแบบจำลองฮิดเดนมาร์คอฟ"

สมการในการประมาณค่าคำนวณสามารถหาได้โดยตรงจากการหาค่าสูงสุดของฟังก์ชันเสริมของ Baum บน $\bar{\lambda}$ ดังนี้

$$Q(\lambda, \bar{\lambda}) = \sum_{Q} \Pr(Q|O,\lambda) \log\left[\Pr(O,Q|\bar{\lambda})\right] \tag{2.87}$$

การทำให้ $Q(\lambda, \bar{\lambda})$ มีค่ามากที่สุดเป็นการทำให้ความน่าเป็นจริงมีค่าสูงขึ้นด้วย โดยทำให้ฟังก์ชันความน่าเป็นจริงลู่เข้าสู่จุดวิกฤตจุดหนึ่งดังนี้

$$\max_{\bar{\lambda}}\left[Q(\lambda|\bar{\lambda})\right] \Rightarrow \Pr(O|\bar{\lambda}) \geq \Pr(O|\lambda) \tag{2.88}$$

ขั้นตอนวิธีการประมาณค่าใหม่สามารถแสดงให้เห็นเป็นขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุด (EM Algorithm) โดยขั้นตอนในการหาค่าคาดหวังคือการคำนวณหาฟังก์ชันเสริม $Q(\lambda, \bar{\lambda})$ และขั้นตอนในการหาค่ามากที่สุดก็คือการหาค่า $Q(\lambda, \bar{\lambda})$ ที่มีค่ามากที่สุดบน $\bar{\lambda}$ นอกจากนี้ขั้นตอนวิธีการประมาณค่าใหม่จะต้องเป็นไปตามเงื่อนไขเฟ้นสุ่ม (Stochastic Constraint) ของพารามิเตอร์แบบจำลองฮิดเดนมาร์คอฟ ดังต่อไปนี้

$$\sum_{i=1}^{N} \bar{\pi}_i = 1 \tag{2.89}$$

$$\sum_{j=1}^{N} \bar{a}_{ij} = 1 \quad , \quad 1 \le i \le N \tag{2.90}$$

$$\sum_{k=1}^{L} \bar{b}_j(k) = 1 \quad , \quad 1 \le j \le N \tag{2.91}$$

ขั้นตอนวิธีการ Viterbi (Viterbi Algorithm)

ขั้นตอนวิธีการ Viterbi เป็นขั้นตอนวิธีการหาลำดับสถานะของแบบจำลองฮิดเดนมาร์คอฟที่ถูกซ่อนเนื่องจากว่าลำดับสถานะนี้บอกถึงโครงสร้างของแบบจำลอง ค่าเฉลี่ยทางสถิติและลักษณะอื่นๆ ของแบบจำลองที่แสดงได้จากลำดับสถานะ วิธีในการหาลำดับสถานะที่เป็นไปได้เมื่อกำหนดลำดับค่าสังเกตมาให้คือการเลือกสถานะ s_t ด้วยวิธีที่เหมาะสมเพื่อให้ได้เส้นทางดีที่สุดและมีความน่าจะเป็น $\Pr(O,S \mid \lambda)$ มากที่สุด เทคนิคที่ใช้ในการหาลำดับสถานะที่ดีที่สุดเพียงลำดับเดียวเรียกว่า "ขั้นตอนวิธีการ Viterbi" ซึ่งคล้ายๆ กับขั้นตอนวิธีการเทียบทางเวลาแบบพลวัต (Dynamic Time Warping, DTW) โดยกำหนดให้ลำดับค่าสังเกต $O = \{ O_1\ O_2\ \cdots\ O_T \}$ และมีลำดับสถานะที่ดีที่สุดคือ $S = \{ s_1\ s_2\ \cdots\ s_T \}$ กำหนดตัวแปรได้ดังนี้

$$\delta_t(i) = \max_{s_1, s_2, \ldots, s_T} \Pr[s_1, s_2, \ldots, s_T = i, O_1\ O_2\ \cdots O_T \mid \lambda] \tag{2.92}$$

โดยที่ $\delta_t(i)$ เป็นความน่าจะเป็นที่มีค่าสูงสุดของเส้นทางเดี่ยวที่เวลา t ซึ่งเป็นค่าสังเกต t ค่าแรกและสิ้นสุดในสถานะ i ด้วยวิธีอุปนัยดังนี้

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij}\right] \cdot b_j(O_{t+1}) \tag{2.93}$$

ในการเรียกใช้ค่าลำดับสถานะจำเป็นต้องติดตามค่าอาร์กิวเมนต์ที่ทำให้สมการที่ (2.93) มีค่ามากที่สุดสำหรับแต่ละค่า t และ j โดยอาศัยแถวลำดับ $\psi_t(j)$ รายละเอียดของขั้นตอนวิธีการ Viterbi มีดังนี้

ขั้นตอนที่ 1 กำหนดเงื่อนไขเริ่มต้นให้กับทุกๆ สถานะ i

$$\delta_t(i) = \pi_i b_i(O_1) \ , \ 1 \le i \le N \tag{2.94}$$

$$\psi_1(i) = 0 \tag{2.95}$$

ขั้นตอนที่ 2 การวนซ้ำ

$$\delta_t(j) = \max_i \left[\delta_{t-1}(i) a_{ij}\right] b_j(O_t) \tag{2.96}$$

$$\psi_t(j) = \arg\max_i \left[\delta_{t-1}(i) a_{ij}\right] \tag{2.97}$$

ขั้นตอนที่ 3 การสิ้นสุด

$$P^* = \max_{s \in S_T} [\delta_T(s)] \tag{2.98}$$

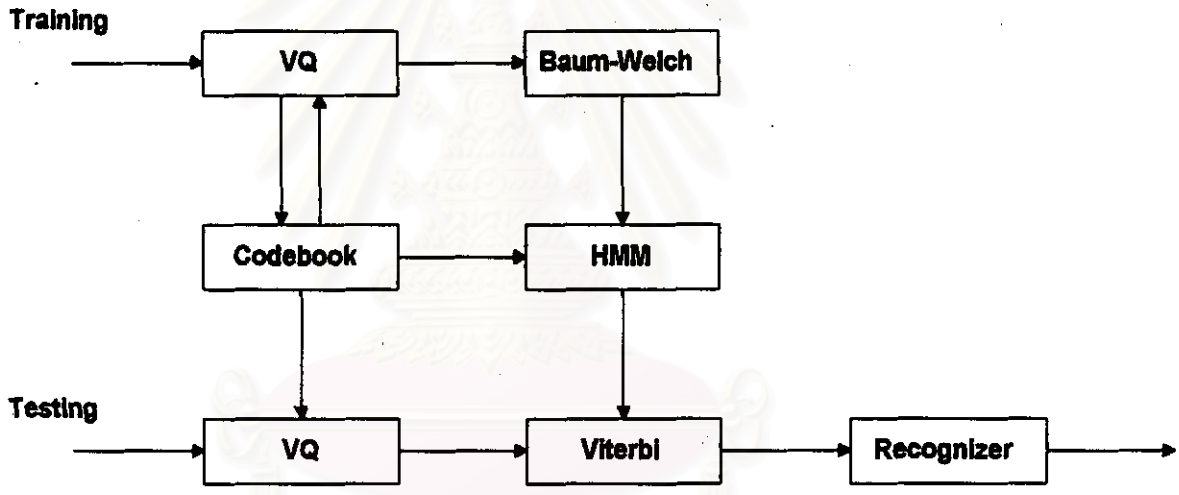
$$s_T^* = \arg\max_{s \in S_T} [\delta_T(s)] \tag{2.99}$$

ขั้นตอนที่ 4 การย้อนกลับเส้นทาง

$$s_t^* = \psi_{t+1}(s_{t+1}^*) \ , \ t = T-1, T-2, \ldots, 1 \tag{2.100}$$

จากขั้นตอนวิธีการไปหน้าและย้อนกลับที่กล่าวไปก่อนหน้านี้เป็นการหาความน่าจะเป็น $\Pr(O \mid \lambda)$ ซึ่งความน่าจะเป็นนี้เป็นผลรวมของ $\Pr(O, S \mid \lambda)$ บนทุกๆ ลำดับสถานะที่เป็นไปได้ S แต่ขั้นตอนวิธีการ Viterbi เป็นเพียงการหาประสิทธิภาพของค่าสูงสุดของความน่าจะเป็น $\Pr(O, S \mid \lambda)$ ในทุกๆ สถานะ S ดังนั้นจึงมองว่าขั้นตอนวิธีการ Viterbi เป็นขั้นตอนวิธีการในกรณีพิเศษของจากขั้นตอนวิธีการไปหน้าและย้อนกลับ

เมื่อนำขั้นตอนวิธีการที่กล่าวมาใช้แก้ปัญหาพื้นฐานทั้ง 3 ข้อของแบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต และการควอนไทซ์แบบเวกเตอร์มาประยุกต์ใช้งานร่วมกันได้เป็นระบบการรู้จำเสียงพูดดังแสดงในรูปที่ 2.13

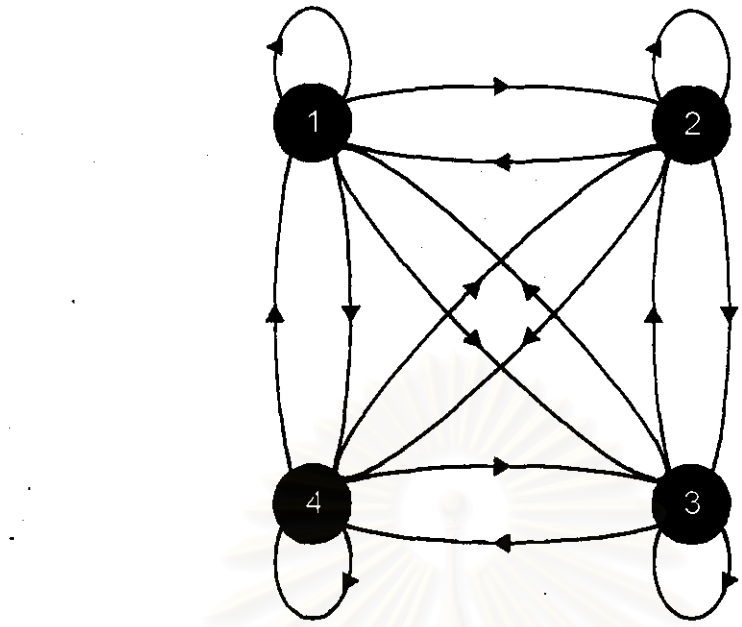


รูปที่ 2.13 ระบบการรู้จำเสียงพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (Zhang, 1995)

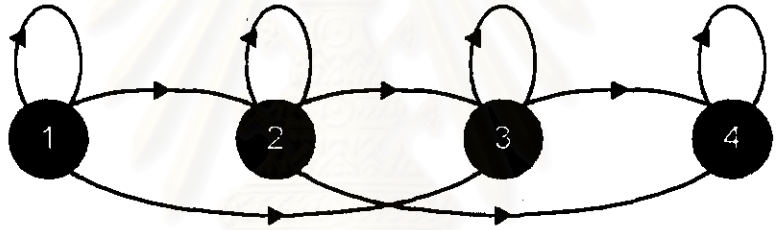
2.4.4 ประเภทของแบบจำลองฮิดเดนมาร์คอฟ

แบบจำลองฮิดเดนมาร์คอฟส่วนใหญ่เป็นการพิจารณาเพียงกรณีพิเศษของแบบจำลองฮิดเดนมาร์คอฟประเภทเออร์กอดิกมาตรฐาน ซึ่งเป็นแบบจำลองฮิดเดนมาร์คอฟที่ทุกสถานะต่อเชื่อมถึงกันหมด โดยทุกสถานะของแบบจำลองสามารถเข้าถึงสถานะอื่นๆ ได้ในเส้นทางเดียวดังแสดงในรูปที่ 2.14(ก) เมื่อจำนวนสถานะของแบบจำลอง $N = 4$ สถานะและมีคุณสมบัติเฉพาะของสัมประสิทธิ์ a_{ij} มีค่าเป็นบวก จากรูปที่ 2.14(ก) จะได้สมการของค่าสัมประสิทธิ์ a_{ij} ได้ดังนี้

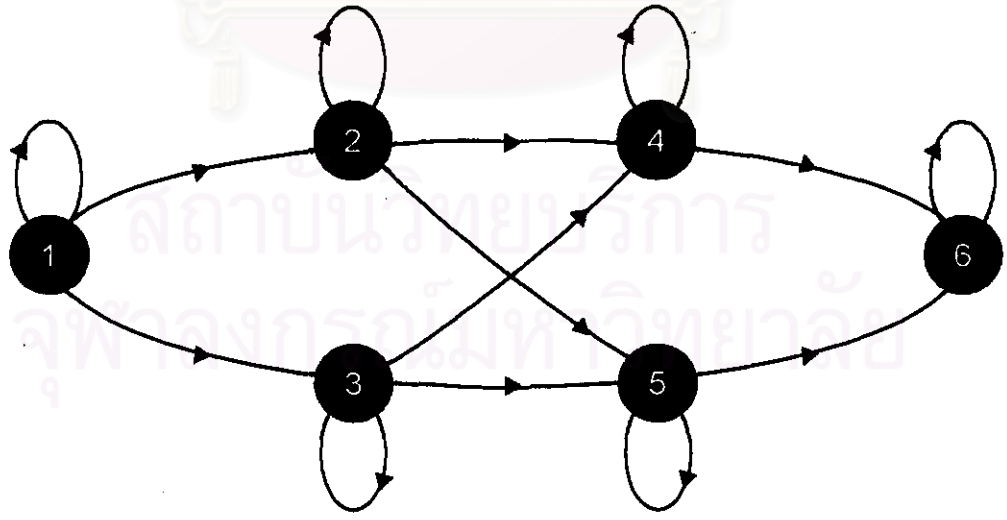
$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \tag{2.101}$$



(ก)



(ข)



(ค)

รูปที่ 2.14 แบบจำลองฮิดเดนมาร์คอฟ 3 ประเภท (ก) แบบจำลองเออร์กอดิก 4 สถานะ (ข) แบบจำลองซ้ายไปขวา 4 สถานะ (ค) แบบจำลองเส้นทางขนานซ้ายไปขวา 6 สถานะ (Rabiner, 1989)

ในการประยุกต์ใช้งานแบบจำลองฮิดเดนมาร์คอฟกับงานเฉพาะอย่างนั้น ได้มีการเสนอแบบจำลองฮิดเดนมาร์คอฟประเภทอื่นที่เหมาะสมกับคุณสมบัติของสัญญาณเสียงพูดได้ดีกว่าแบบจำลองประเภทเออร์กอดิกมาตรฐาน คือแบบจำลองซ้ายไปขวา (Left-Right Model) หรือแบบจำลอง Bakis ดังแสดงในรูปที่ 2.14(ข) เพราะว่าลำดับสถานะของแบบจำลองมีคุณสมบัติตามเวลา คือเมื่อเวลาเพิ่มขึ้นดัชนีของสถานะจะเพิ่มขึ้นหรืออยู่ที่เดิมเปรียบเสมือนดำเนินจากซ้ายไปขวา ดังนั้นแบบจำลองซ้ายไปขวาจึงเหมาะสมในการจำลองสัญญาณที่มีคุณสมบัติเปลี่ยนแปลงตามเวลา เช่น สัญญาณเสียงพูด เป็นต้น

คุณสมบัติพื้นฐานของแบบจำลองฮิดเดนมาร์คอฟประเภทซ้ายไปขวานั้น คือค่าสัมประสิทธิ์ของการเปลี่ยนสถานะต้องมีคุณสมบัติดังนี้

$$a_{ij} = 0, \quad j < i \tag{2.102}$$

จากสมการที่ (2.102) หมายความว่าไม่มีการเปลี่ยนสถานะไปยังสถานะที่มีดัชนีต่ำกว่าดัชนีของสถานะปัจจุบัน และมีความน่าจะเป็นเริ่มต้นมีคุณสมบัติตามสมการที่ (2.103)

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \tag{2.103}$$

เนื่องจากลำดับสถานะต้องเริ่มต้นจากสถานะที่ 1 และสิ้นสุดในสถานะที่ N ดังนั้นแบบจำลองซ้ายไปขวาต้องเพิ่มเงื่อนไขของสัมประสิทธิ์การเปลี่ยนสถานะ เพื่อไม่ให้เกิดการเปลี่ยนแปลงของดัชนีมากเกินไปดังนี้

$$a_{ij} = 0, \quad j > i + \Delta \tag{2.104}$$

จากรูปที่ 2.14(ข) จะกำหนดให้ $\Delta = 2$ และไม่มีการข้ามสถานะเกินกว่า 2 สถานะ สามารถเขียนการเปลี่ยนสถานะให้อยู่ในรูปเมตริกซ์ได้ดังนี้

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix} \tag{2.105}$$

และที่สถานะสุดท้ายของแบบจำลองประเภทซ้ายไปขวามีสัมประสิทธิ์ของการเปลี่ยนสถานะดังสมการต่อไปนี้

$$a_{NN} = 1 \tag{2.106}$$

$$a_{Ni} = 0, \quad i < N \tag{2.107}$$

นอกจากนี้ยังมีแบบจำลองฮิดเดนมาร์คอฟประเภทอื่นๆ ที่เป็นไปได้อีกมากขึ้นอยู่ลักษณะการนำไปประยุกต์ใช้งาน ตัวอย่างดังในรูปที่ 2.14(ค) เป็นแบบจำลองฮิดเดนมาร์คอฟประเภทซ้ายไปขวาต่อขนานกันสองชุด ซึ่งแบบจำลองประเภทนี้มีความยืดหยุ่นมากกว่าแบบจำลองประเภทซ้ายไปขวา