

บทที่ 2 งานวิจัยที่เกี่ยวข้อง และทฤษฎี

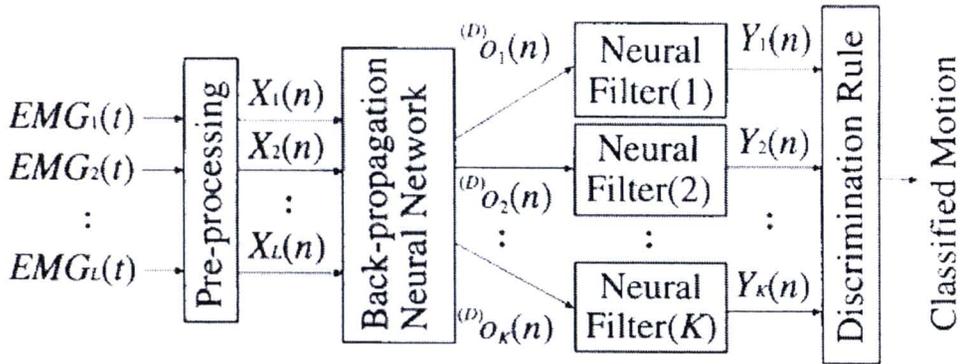
ในด้านการค้นหาวิธีสำหรับลดมิติข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อที่เหมาะสมนั้น งานวิจัยที่เกี่ยวข้องหรืองานวิจัยที่มีผู้ทดลองทำมาแล้วในอดีตที่ได้ทำการศึกษาในวิทยานิพนธ์นี้ประกอบไปด้วยเนื้อหาของการสกัดลักษณะเด่นในรูปแบบต่างๆ ทั้งเชิงเวลา และเชิงความถี่ รวมถึงเทคนิคการลดมิติข้อมูลและตัวจำแนกกลุ่มที่ได้ถูกนำมาประยุกต์ใช้ในรอบหลายปีที่ผ่านมา

สำหรับทฤษฎีนั้นจะประกอบไปด้วยสามหัวข้อซึ่งอธิบายถึงรายละเอียดของวิธีการลดมิติข้อมูลและหลักการคิดคำนวณของแต่ละเทคนิค วิธีการคำนวณเพื่อสกัดเอาลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อออกมา และส่วนสุดท้ายคือทฤษฎีการวัดค่าความแยกตัวของกลุ่มข้อมูลซึ่งใช้เป็นตัวชี้วัดในงานวิจัยชิ้นนี้

2.1 งานวิจัยที่เกี่ยวข้อง

คลื่นไฟฟ้ากล้ามเนื้อเมื่อถูกนำมาประยุกต์ใช้กับการควบคุมทางวิศวกรรมหลายด้าน โดยเฉพาะอย่างยิ่งแขนกล และกายอุปกรณ์เทียม เนื่องจากคลื่นไฟฟ้ากล้ามเนื้อจะบ่งบอกถึงความต้องการของผู้ใช้โดยตรง แต่การนำเอาข้อมูลดิบของคลื่นไฟฟ้ากล้ามเนื้อทั้งหมดมาใช้จำแนกนั้นเป็นสิ่งที่ทำให้การประมวลผลเพื่อจำแนกลักษณะมีเวลานาน ส่งผลถึงการตอบสนองของอุปกรณ์ต่อผู้ใช้ที่มีจะช้าไปตามลำดับ งานวิจัยที่มีการใช้คลื่นไฟฟ้ากล้ามเนื้อจึงเป็นต้องทำการเลือกลักษณะเด่นก่อนนำไปประมวลผลเพื่อสกัดเอาเฉพาะส่วนข้อมูลที่จำเป็นมาใช้ ทำให้สามารถลดเวลาในการประมวลผลลงได้ วิธีการลดมิติข้อมูลของแต่ละงานวิจัยได้เลือกใช้นั้นมีความเหมือนและแตกต่างกันไป ดังนี้

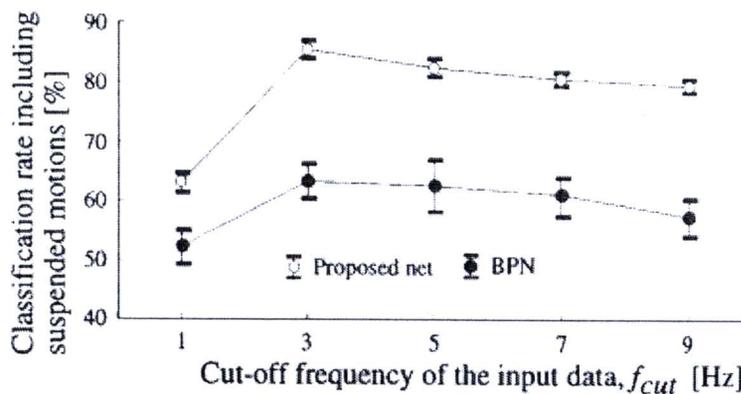
ในปี ค.ศ. 2000 Toshio Tsuji และคณะ [1] จาก Hiroshima University ได้ทำการเสนอวิธีการจำแนกรูปแบบของคลื่นไฟฟ้ากล้ามเนื้อในสภาวะที่ไม่อยู่นิ่ง โดยใช้ระบบโครงข่ายประสาทเทียม (Artificial Neural Network) สำหรับควบคุมกายอุปกรณ์เทียม ในขั้นตอนก่อนการประมวลผลจำแนกสัญญาณคลื่นไฟฟ้ากล้ามเนื้อของกล้ามเนื้อแต่ละมัดจะถูกวัดด้วยอิเล็กโทรดจำนวน L ตัววัดที่ความถี่ 1000 เฮิรตซ์และแปลงข้อมูลเป็นดิจิตอลจำนวน 12 บิต สัญญาณที่วัดได้จะถูกนำไปผ่านตัวกรองสัญญาณ (Filter) แบบบัตเตอร์เวิร์ทที่ความถี่ขอบเขตเท่ากับ f_{cut} เฮิรตซ์ หลังจากนั้นข้อมูลคลื่นไฟฟ้ากล้ามเนื้อที่เข้ามาถูกแปลงให้เป็นเวกเตอร์ลักษณะเด่น (Feature Vector) ซึ่งเป็นการลดมิติของข้อมูลลงจาก t ค่าไปให้เหลือเพียง N ค่า แล้วจึงนำข้อมูลที่ได้ส่งต่อไปเป็นข้อมูลขาเข้าของโครงข่ายประสาทเทียมแบบ Back-propagation ในลำดับถัดไป และท้ายที่สุดข้อมูลขาออกของระบบคือประเภทของการเคลื่อนที่ของผู้ใช้



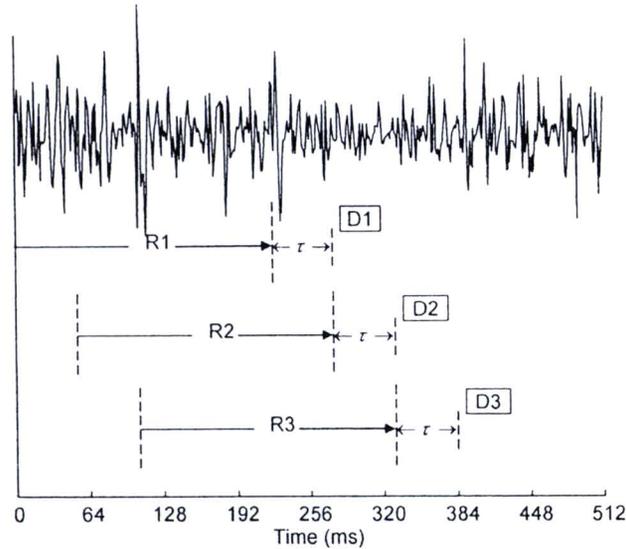
รูปที่ 2.1 รูปแบบของระบบโครงข่ายประสาทเทียมที่นำเสนอ [1]

จากการทดลองจำแนกสัญญาณในสภาวะไม่อยู่นิ่งของคนปกติ และผู้ป่วยแขนขาดได้ผลว่า ในรายของคนปกติระบบโครงข่ายประสาทเทียมที่นำเสนอมีความแม่นยำโดยเฉลี่ยเท่ากับ 76.5% และ ค่าความแม่นยำเท่ากับ 83.4% สำหรับผู้ป่วยแขนขาด

งานวิจัยชิ้นนี้มีผลความแม่นยำในการจำแนกข้อมูลที่ดี แต่ไม่ได้ให้รายละเอียดเกี่ยวกับขั้นตอนของการลดมิติการพิจารณา กล่าวถึงเพียงผลกระทบเกี่ยวกับการเลือกความถี่ขอบเขต ว่ามีผลทำให้เวกเตอร์ลักษณะเด่นที่ได้มีลักษณะเฉพาะแตกต่างกันออกไป ซึ่งส่งผลโดยตรงกับค่าความแม่นยำของการจำแนกที่จะได้ ดังในรูปที่ 2.2



รูปที่ 2.2 ผลกระทบของความถี่ขอบเขตต่อความสามารถในการจำแนกของโครงข่ายประสาทเทียม [1]



รูปที่ 2.3 แบบแผนการเคลื่อนหน้าต่างที่ให้ประสิทธิภาพในการคำนวณสูงสุด [2]

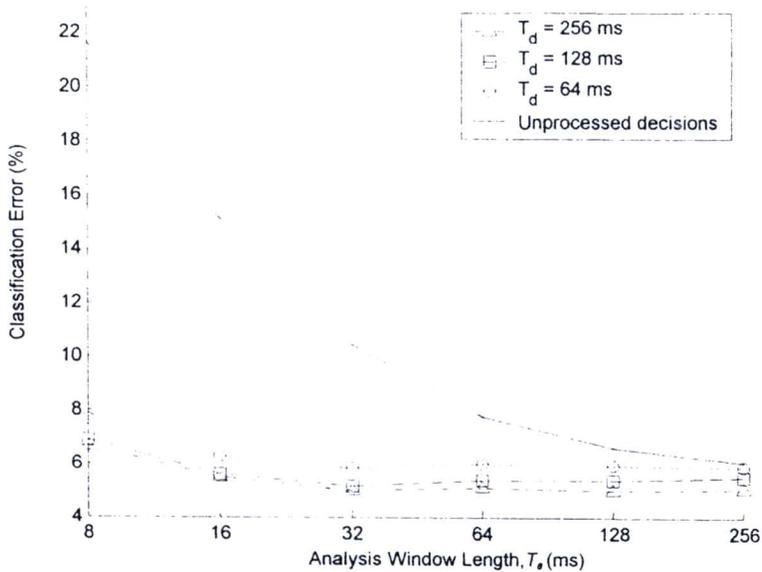
จากนั้นในปี 2003 Kevin Englehart และ Bernard Hudgins [2] ได้เสนอรูปแบบการควบคุมแบบ Real-Time ซึ่งมีแบบแผนการเคลื่อนหน้าต่างดังแสดงรูปที่ 2.3 ซึ่งเป็นวิธีที่ช่วยทำให้ประสิทธิภาพในการคำนวณลักษณะเด่นมีมากที่สุด อีกทั้งยังเพิ่มจำนวนผลการตัดสินใจของระบบให้มากขึ้นอีกด้วย

การทดลองดังกล่าวได้ใช้คลื่นไฟฟ้ากล้ามเนื้อจากการวัดค่าสัญญาณกระทำที่ความถี่เท่ากับ 1000 เฮิรซ์ แล้วแปลงค่าสัญญาณเป็นดิจิทัลจำนวน 16 บิต สัญญาณที่เข้ามาจะถูกประมวลผลบนหน้าต่างการพิจารณา (Analysis Windows) ขนาดตั้งแต่ 8 ถึง 256 มิลลิวินาที โดยแต่ละหน้าต่างการพิจารณาจะถูกคำนวณสร้างเป็นเซตของลักษณะเด่นซึ่งประกอบไปด้วยข้อมูลสถิติเชิงเวลา 4 ค่า คือ

1. จำนวนของครั้งที่กราฟผ่านค่าศูนย์ (Zero Crossing)
2. ความยาวรูปคลื่น (Waveform Length)
3. จำนวนของการเปลี่ยนเครื่องหมายของค่าความชัน (Slope Sign Changes)
4. ค่าอินทิเกรตของสัญญาณกล้ามเนื้อ (Mean Absolute Value)

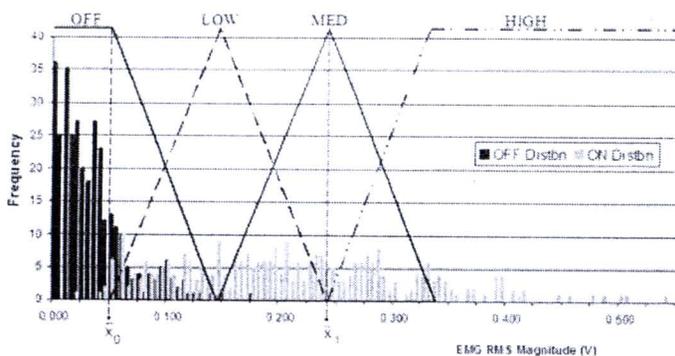
เมื่อค่าทั้งหมดถูกคำนวณเสร็จ แต่ละช่องสัญญาณข้อมูลจะถูกแปลงให้อยู่ในรูปของเวกเตอร์ลักษณะเด่น (Feature Vector) ขนาด 16 มิติ เพื่อนำไปใช้จำแนกข้อมูลโดยใช้ตัวจำแนกแบบ Linear Discriminate Analysis (LDA) ซึ่งเป็นวิธีที่ง่ายและรวดเร็ว อีกทั้งยังได้ความแม่นยำในการจำแนกที่ดี

ผลจากการทดลองทำให้เห็นว่าขนาดหน้าต่างที่ 32 ถึง 256 ไม่มีนัยทางความแตกต่างของค่าความแม่นยำ เพียงแต่การเพิ่มขนาดหน้าต่างจะทำให้ระบบต้องใช้หน่วยความจำสำหรับการคำนวณมากขึ้น ดังแสดงในกราฟรูปที่ 2.4



รูปที่ 2.4 ผลของการใช้งานหน้าต่างการพิจารณาที่ขนาดเวลา T_a และดีเลย์ที่ยอมรับได้ T_d [2]

เนื่องด้วยจากความไม่เป็นเชิงเส้นอย่างสูงของสัญญาณไฟฟ้ากล้ามเนื้อ จึงได้มีการประยุกต์ใช้ตรรกศาสตร์แบบฟัซซี (Fuzzy Logic) เข้ามาจัดการจำแนกกลุ่มข้อมูล โดยในปี ค.ศ. 2005 Abidemi Bolu Ajiboye และ Richard F. ff. Weir [3] ได้นำเสนอวิธีการใช้ชีวิตดังกล่าวสำหรับควบคุมกายอุปกรณ์เทียม โดยในขั้นตอนก่อนการประมวลผลข้อมูลสัญญาณที่เข้ามาจะผ่านการคำนวณค่ารากกำลังสองเฉลี่ย (Root-Mean-Square; RMS) ด้วยค่าสัญญาณ 64 ค่าในหน้าต่างการพิจารณา ซึ่งใช้เวลาทั้งหมด 45.7 มิลลิวินาที (64 samples/1400 Hz). ซึ่งวิธีการหาค่า RMS นี้เองที่เป็นตัวลดมิติการพิจารณาข้อมูลลงในการคำนวณ จากนั้นค่า RMS ที่ได้จะเป็นข้อมูลขาเข้าของฟังก์ชันการเป็นสมาชิก (Membership Function) ดังรูปที่ 2.3

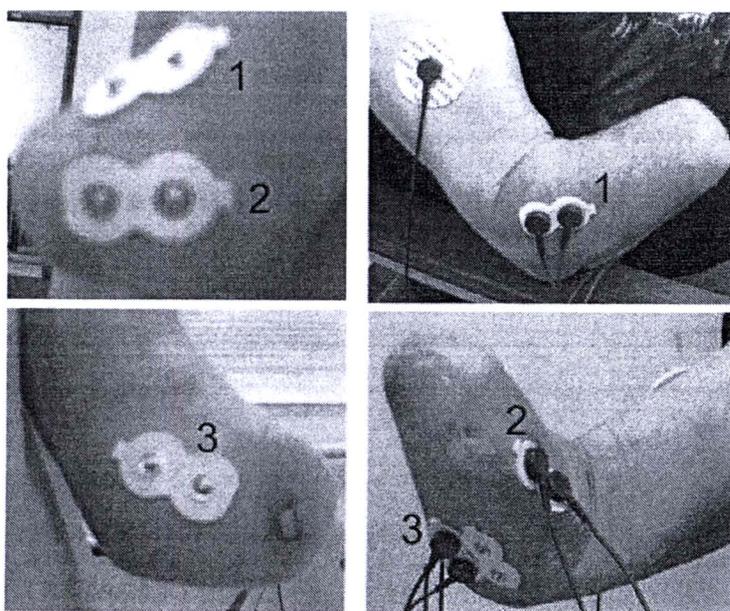


รูปที่ 2.5 ฮิสโทแกรมค่า RMS ของสัญญาณไฟฟ้ากล้ามเนื้อ ในช่วงก่อนการขยับกล้ามเนื้อ (OFF Distbn) และหลังขยับกล้ามเนื้อ (ON Distbn) ในฟังก์ชันการเป็นสมาชิก [3]

การทดลองนี้ได้ทำการวัดคลื่นไฟฟ้ากล้ามเนื้อจากผู้ป่วยพิการทางแขน จำนวน 4 คน ในตำแหน่งดังรูปที่ 2.6 ผู้ป่วยจะทำการขยับใน 4 ท่าทาง คือ

1. ยืดแขนหรือข้อมือ (Extension)
2. คว่ำมือ (Supination)
3. หงายมือ (Pronation)
4. ทำพักนิ่ง(Off)

เมื่อกำหนดได้ค่าองศาของความเป็นสมาชิก (Degree-Of-Membership) ของแต่ละช่องสัญญาณจากฟังก์ชันการเป็นสมาชิกแล้ว ค่านั้นจะถูกนำไปสรุปความหมายโดยใช้กฎการสรุปความแบบฟัซซี่ ดังแสดงในตารางที่ 2.1 กฎดังกล่าวสร้างขึ้นมาสืบขึ้นมาโดยใช้ Fuzzy C-Mean จำนวนของกฎที่ใช้ในการแปลความหมายนั้นขึ้นอยู่กับจำนวนของรูปแบบการขยับของกล้ามเนื้อที่ใช้ในการเรียนรู้ระบบ ผลจากการทดลองกับผู้ป่วย 4 คน พบว่าค่าความแม่นยำของระบบ เป็นไปดังตารางที่ 2.2 และมีความแม่นยำเฉลี่ยรวมอยู่ที่ 96.56%



รูปที่ 2.6 ตำแหน่งการวัดสัญญาณ EMG: (1) Extensors; (2) Supinators; (3) Flexors [3]

ตารางที่ 2.1 กฎการสรุปความที่ใช้ในระบบ [3]

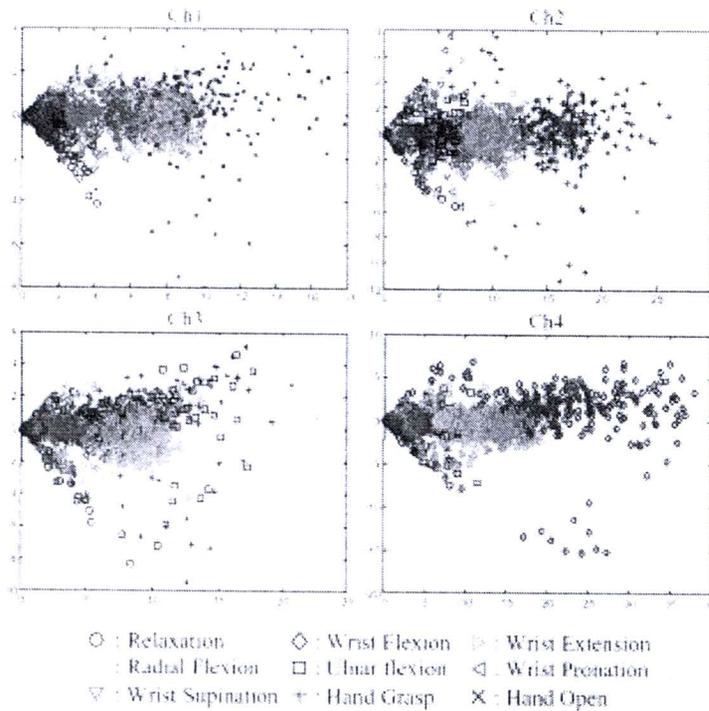
| | IF | | THEN | DOS |
|---------------|----------------|---------------|------------|-------|
| EMG Input1 | EMG Input 2 | EMG Input3 | Output | |
| Off | Off | Off | Off | 0.984 |
| Low | Off | Off | Extension | 1.00 |
| Med | Off | - | Extension | 0.961 |
| Med | Low | - | Extension | 0.984 |
| Med | Med | - | Extension | 0.969 |
| High | - | - | Extension | 1.00 |
| Off | Low | Off | Supination | 0.977 |
| Off | Med | Off | Supination | 0.977 |
| Off | Med | Low | Supination | 0.266 |
| Low | Med | Off | Supination | 0.422 |
| Low | Med | Low | Supination | 0.125 |
| Off | High | - | Supination | 0.094 |
| Low | High | - | Supination | 0.047 |
| Off | Off | Low | Flexion | 0.320 |
| Off | Off | Med | Flexion | 0.008 |
| Low | Off | Med | Flexion | 0.000 |
| Low | Low | Med | Flexion | 0.000 |
| - | Med | Med | Flexion | 0.078 |
| - | - | High | Flexion | 0.164 |

ตารางที่ 2.2 ค่าร้อยละความแม่นยำรวมของการจำแนกข้อมูลในผู้ป่วยแต่ละราย [3]

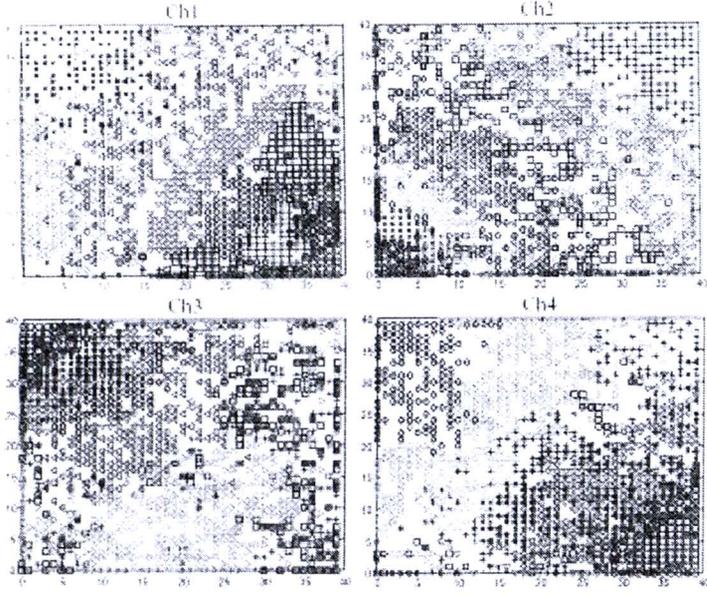
| | Subject 1 | Subject 2 | Subject 3 | Subject 4 |
|---------|-----------|-----------|-----------|-----------|
| OVERALL | 96.02% | 98.27% | 97.17% | 94.79% |

งานวิจัยเกี่ยวกับการใช้งานคลื่นไฟฟ้ากล้ามเนื้อในงานควบคุมช่วงระยะหลังที่ผ่านมา ได้เน้นประเด็นไปที่ความรวดเร็วในการประมวลผลแบบ Real-Time มากขึ้น เพื่อเพิ่มการตอบสนองที่ดีขึ้นให้กับผู้ใช้งาน ในปี 2006 Jun-Uk Chu และคณะ [4] ได้เสนอรูปแบบการจำแนกสัญญาณแบบใหม่โดยใช้วิธีการฉายลักษณะเด่นแบบเป็นเชิงเส้น และแบบไม่เป็นเชิงเส้น ซึ่งเป็นการรวมกันระหว่าง Principal Component Analysis (PCA) กับ Self-Organizing Feature Map (SOFM)

การลดมิติการพิจารณาโดยใช้ PCA ทำให้โครงสร้างของ classifier มีความง่ายขึ้น อีกทั้งยังช่วยลดเวลาที่ใช้ในการประมวลผลด้วย ส่วนการทำ Mapping แบบไม่เป็นเชิงเส้นด้วย SOFM จะเป็นการแปลงค่าที่ลดมิติลงมาจาก PCA ให้อยู่ในอีก พีเจอร์สเปส 2 มิติ มีขนาดเท่ากับ 40 x 40 ซึ่งช่วยเพิ่มความแตกต่างของแต่ละกลุ่มข้อมูลให้เพิ่มมากขึ้น ส่งผลต่อความแม่นยำที่ดีขึ้นในการจำแนกกลุ่ม

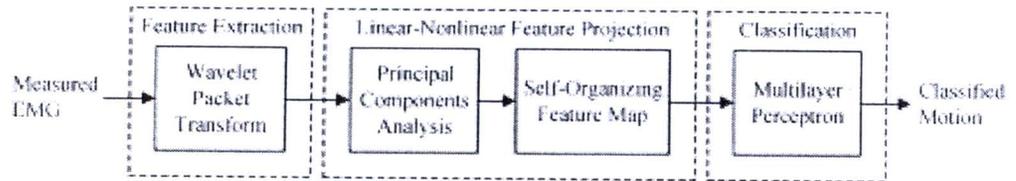


รูปที่ 2.7 กราฟแสดง 2 PCs ที่ลดมิติลงมาโดยใช้ PCA ในแต่ละช่องสัญญาณ [4]



รูปที่ 2.8 ลักษณะเด่นที่ถูกจัดกลุ่มแล้ว โดยใช้ SOFM [4]

แผนภาพบล็อกวิธีการจำแนกคลื่นไฟฟ้ากล้ามเนื้อ ในการทดลองสามารถเขียนได้เป็นดังรูปที่ 2.9 เริ่มต้นจากการแปลงสัญญาณที่เข้ามาโดยใช้ Wavelet Packet Transform (WPT) แล้วจึงทำการลดมิติของข้อมูลลงด้วย PCA และ SOFM ตามลำดับ และท้ายที่สุดข้อมูลจะถูกจำแนกด้วย Multilayer Perceptron (MLP) ผลออกมาเป็นประเภทการเคลื่อนที่ของผู้ใช้



รูปที่ 2.9 แผนภาพบล็อกวิธีการจำแนกสัญญาณไฟฟ้ากล้ามเนื้อ [4]

ประสิทธิภาพของวิธีการจำแนกข้อมูลที่เสนอนี้เทียบกับวิธีการ PCA และ SOFM เพียงอย่างเดียว สำหรับกลุ่มผู้ทดลองจำนวน 10 คน เป็นไปดังตารางที่ 2.3 และผลการทดสอบการควบคุมแบบ Real-Time สำหรับแขนกลจำลองในคอมพิวเตอร์ ได้ผลว่าใช้เวลาในการประมวลผลรวมเท่ากับ 100 มิลลิวินาที

ตารางที่ 2.3 ค่าร้อยละความถูกต้องในการ โปรเจกซ์ข้อมูลของแต่ละ Feature Projection [4]

| Subject | PCA + SOFM | PCA | SOFM |
|---------|------------|---------|---------|
| Mean | 97.024% | 97.785% | 97.759% |

ถัดมาในปี ค.ศ. 2007 Jun-Uk Chu และคณะ [5] ก็ได้ทำการพัฒนาวิธีการจำแนกรูปแบบคลื่นไฟฟ้ากล้ามเนื้อ ต่อเพื่อนำไปใช้กับแขนกลจริงแบบ Real-Time โดยใช้ Linear Discriminate Analysis เป็นตัวคณิตในการพิจารณาข้อมูลแทนการใช้ PCA รวมกับ SOFM เนื่องจาก LDA สามารถอธิบายความแตกต่างระหว่างแต่ละกลุ่มข้อมูลได้ดีกว่า PCA ที่จะเป็นการอธิบายลักษณะของกลุ่มข้อมูลนั้นๆ จึงน่าจะสามารถใช้จำแนกกลุ่มรูปแบบคลื่นไฟฟ้ากล้ามเนื้อ ได้ดีกว่าแบบเดิม

ข้อมูลดิบของคลื่นไฟฟ้ากล้ามเนื้อที่รับมายังคงถูกแปลงด้วยวิธี Wavelet Packet Transform ได้ออกมาเป็นเวกเตอร์ลักษณะเด่นขนาดเท่ากับ 256 มิติ ต่อ 1 ช่องสัญญาณ แต่หลังจากนั้นจะนำข้อมูลที่แปลงแล้วไปลดมิติลงด้วย LDA ให้เหลือเพียง 8 มิติ เท่านั้น เพื่อส่งต่อไปเป็นข้อมูลขาเข้าของ Multilayer Perceptron ทำการจำแนกกลุ่มสัญญาณเป็นลำดับถัดไป

การทดลองเพื่อเปรียบเทียบประสิทธิภาพของการลดมิติการพิจารณาของ LDA ได้มีการเทียบกับ PCA Nonlinear Discriminate Analysis (NLDA) และ SOFM ซึ่งจากการเปรียบเทียบโดยใช้ Sammon's Stress และ Fisher's Index สามารถสรุปมาเป็นดังตารางที่ 2.3 นอกจากนั้นแล้วยังมีการเปรียบเทียบเวลาที่ใช้ในการประมวลผล และอัตราประสบความสำเร็จของแต่ละวิธีซึ่งได้ผลดังตารางที่ 2.4

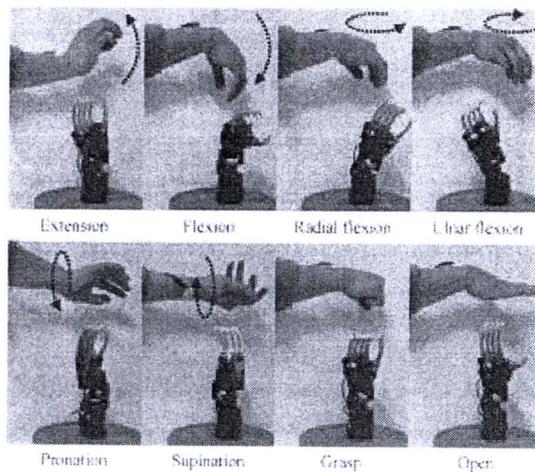
ตารางที่ 2.4 ค่าเฉลี่ยของ Sammon's Stress และ Fisher's Index [5]

| | LDA | LDA | NLDA | SOFM |
|--------------------|---------|--------|--------|---------|
| Sammon's Stress, E | 0.9764 | 0.3548 | 0.9816 | 0.9827 |
| Fisher's Index, J | 26345.8 | 0.0839 | 5248.8 | 47.2026 |

ตารางที่ 2.5 ค่าร้อยละความถูกต้องและเวลาในการประมวลผลในการโปรเจกต์ข้อมูลของแต่ละวิธีการลดมิติข้อมูล[5]

| | LDA | LDA | NLDA | SOFM |
|------------------------|------|------|------|------|
| Success rate[%] | 97.4 | 95.9 | 97.9 | 96.2 |
| Processing time [msec] | 2 | 2 | 150 | 300 |

ซึ่งจากข้อมูลผลการทดลองแสดงให้เห็นว่า LDA มีประสิทธิภาพที่ดี และเมื่อรวมเวลาประมวลผลทั้งหมดจนถึงขั้นตอนการตั้งงานแขนกลจริงแล้วพบว่าใช้เวลาไปทั้งหมด 97 มิลลิวินาที



รูปที่ 2.10 การควบคุมแขนกลโดยใช้คลื่นไฟฟ้ากล้ามเนื้อ [5]

ต่อมาในปี ค.ศ. 2010 ของ Dennis Tkach, Heuang และ Todd A Kuiken [6] เป็นการศึกษาความเสถียรของลักษณะเด่นเชิงเวลาของสัญญาณกล้ามเนื้อเพื่อนำมาใช้จำแนกรูปแบบของสัญญาณ โดยได้ทำการจำแนกสัญญาณกล้ามเนื้อจาก 4 กิจกรรม คือ

1. กระจกข้อมือขึ้น
2. กระจกข้อมือลง
3. คว่ำมือ
4. หายมือ

ซึ่งงานวิจัยชิ้นนี้ได้ศึกษาผลกระทบในการจำแนกสัญญาณจากการรบกวน 3 อย่าง อันประกอบไปด้วย

1. ความอ่อนล้าของกล้ามเนื้อ
2. สัญญาณรบกวนจากเครื่องมือวัด
3. การเลื่อนของตำแหน่งการติดอิเล็กโทรด

กลุ่มผู้วิจัยได้เลือกพิจารณาว่าหากเกิดการรบกวนจากทั้ง 3 ปัจจัยนี้แล้ว ลักษณะเด่นในเชิงเวลาตัวใดที่จะสามารถทนทานต่อสิ่งรบกวนทั้งหมดได้ดีที่สุด โดยลักษณะเด่นที่นำมาเปรียบเทียบนั้นมีทั้งหมด 11 ตัว ดังนี้

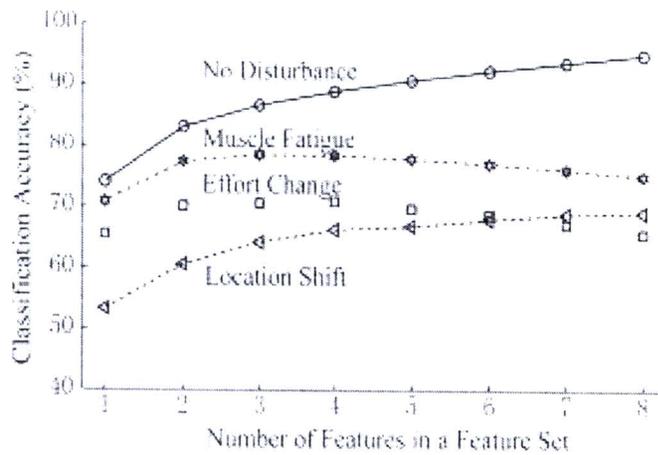
1. Mean Absolute Value (MAV)
2. Zero Crossings (ZC)
3. Slope Sign Changes (SSC)
4. Waveform Length (WL)
5. Willison Amplitude (wAmp)
6. Variance (VAR)
7. v-Order (vOrder)
8. log-Detector (logDetect)
9. EMG Histogram (emgHist)
10. Autoregression Coefficient (AR)
11. Cepstrum Coefficients (Ceps)



ตัวจำแนกที่ใช้ในงานวิจัยนี้คือ Support Vector Machine โดยผลสรุปที่ได้คือ ลักษณะเด่นที่ทนทานต่อสิ่งรบกวนมากที่สุดประกอบไปด้วย

1. Waveform Length
2. Autoregression Coefficient
3. Cepstrum Coefficients

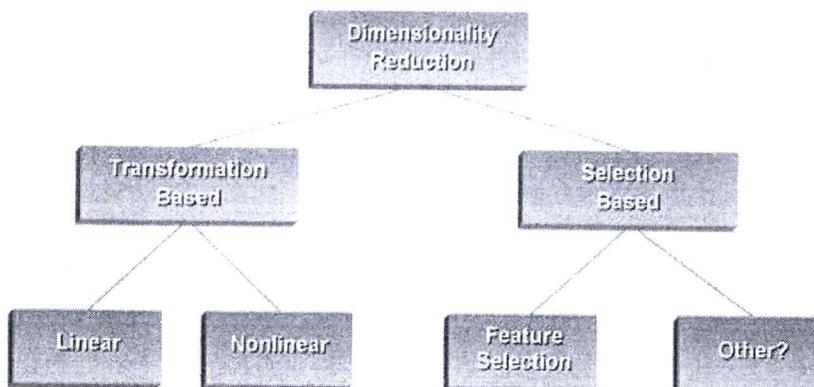
นอกจากนั้นแล้วงานวิจัยนี้ยังได้ทำการทดสอบประสิทธิภาพของการนำลักษณะเด่นมากกว่า 1 ตัว มารวมกันเพื่อบ่งบอกรูปแบบของสัญญาณ ซึ่งผลการทดลองได้พบว่าหากเรานำลักษณะเด่นจำนวน 3 ตัวมาพิจารณาร่วมกันแล้ว จะทำให้ประสิทธิภาพในการจำแนกรูปแบบสัญญาณนั้นดีขึ้น ดังแสดงให้เห็นในรูปที่ 2.11 เห็นได้อย่างชัดเจนว่าเมื่อเราเพิ่มจำนวนลักษณะเด่นในการพิจารณาขึ้น สำหรับสัญญาณที่ไม่มีสิ่งรบกวนจะทำให้ความแม่นยำในการจำแนกเพิ่มขึ้นเรื่อยๆ แต่สำหรับสัญญาณที่มีสิ่งรบกวนทั้ง 3 รูปแบบ จำนวนลักษณะเด่นที่ทำให้ได้ความแม่นยำสูงสุดอยู่ที่จำนวน 3 ค่า



รูปที่ 2.11 กราฟแสดงการเปลี่ยนแปลงของประสิทธิภาพในการจำแนกรูปแบบสัญญาณเมื่อเลือกใช้จำนวนลักษณะเด่นที่แตกต่างกัน [6]

2.2 การจำแนกประเภทของเทคนิคการลดมิติข้อมูล

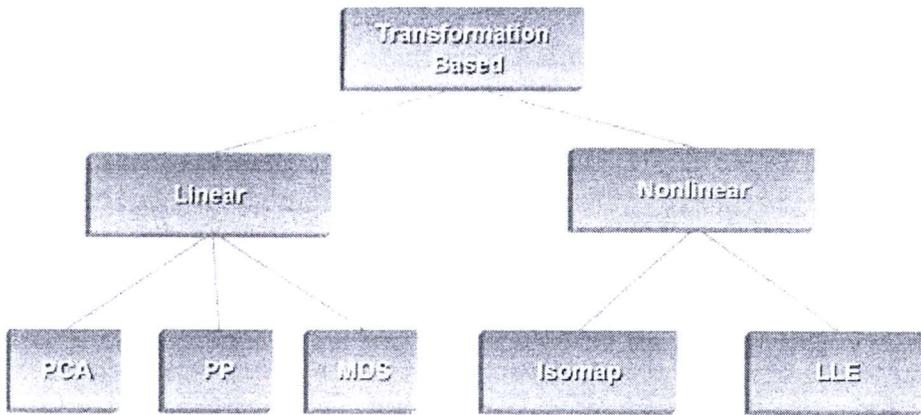
เทคนิคการลดมิติของข้อมูลมีอยู่หลายเทคนิค ซึ่งเราสามารถจัดกลุ่มประเภทของเทคนิคเหล่านั้นได้ออกเป็นดังรูปที่ 2.12 โดยประเด็นหลักที่ใช้ในการจัดกลุ่มดังกล่าวคือ การที่เทคนิคนั้นเป็นวิธีการที่ทำการแปลงรูปข้อมูลเดิมให้เป็นรูปแบบอื่น หรือว่า ยังคงสงวนลักษณะของข้อมูลเดิมไว้ ซึ่งการเลือกใช้เทคนิคในแต่ละแบบนั้นขึ้นอยู่กับว่า เราต้องการจะสงวนค่าลักษณะเด่นเดิมของข้อมูลที่ได้รับไว้ต่อ หรือเพียงแต่ต้องการดูความสัมพันธ์กันภายในกลุ่มข้อมูล ทั้ง 2 วิธีการยังสามารถแบ่งย่อยลงไปได้อีก ซึ่งมีรายละเอียดดังต่อไปนี้



รูปที่ 2.12 แผนภาพการจัดกลุ่มเทคนิคลดมิติข้อมูล [7]

2.2.1 TRANSFORMATION-BASED REDUCTION

โดยทั่วไปแล้วสิ่งที่ได้ออกมาจากวิธีการลดมิติข้อมูล มักจะเป็นการแปลงข้อมูลโดยที่ไม่สามารถย้อนกลับไปยังลักษณะข้อมูลเดิมได้ ซึ่งจะถูกนำมาใช้กับสถานการณ์ที่ไม่ต้องการคงลักษณะข้อมูลเดิมเอาไว้สำหรับประมวลผลในขั้นตอนถัดไป วิธีการเหล่านี้สามารถแบ่งออกเป็น 2 ประเภท คือ วิธีการที่เป็นแบบเชิงเส้น (Linear) และ วิธีการที่ไม่เป็นแบบเชิงเส้น (Nonlinear) ดังรูปที่ 2.13



รูปที่ 2.13 กลุ่มประเภทของเทคนิคที่มีการเปลี่ยนแปลงลักษณะข้อมูล [7]

2.2.1.1 วิธีการแบบเชิงเส้น (Linear Methods)

วิธีการแบบเชิงเส้นนี้ได้รับการพัฒนามานานหลายปี ตัวอย่างวิธีการ ได้แก่ Principal Component Analysis (PCA) และ Multidimensional Scaling (MDS) วิธีการเหล่านี้ถูกใช้เพื่อค้นหาโครงสร้างยูคลิดีเนียนของความสัมพันธ์ภายในกลุ่มข้อมูล ซึ่งถึงแม้ว่าหากความสัมพันธ์ดังกล่าวมีมิติที่สูงมาก จะทำให้วิธีการเหล่านี้ไม่มีประสิทธิภาพได้ดี แต่ปัญหานี้ก็ยังสามารถประยุกต์ใช้ได้ในงานหลายๆด้าน วิธีการในกลุ่มนี้

PCA เป็นเทคนิคที่นิยมใช้มากเนื่องจากมีวิธีการที่ง่าย โดยจะแปลงข้อมูลดิบที่มีอยู่ให้มีมิติที่น้อยลง ด้วยการตัดตัวแปรที่ไม่มีความสัมพันธ์กันออกไป ข้อมูลที่ได้จะอยู่ในรูปของ Principal Component PCA อาศัยหลักการที่ว่า Feature ที่มีค่าความแปรปรวนที่มากเป็นข้อมูลที่มีประโยชน์มากกว่า

Linear Discriminant Analysis หรือ LDA เป็นการเพิ่มอัตราส่วนของค่าความแปรปรวนระหว่างกลุ่มข้อมูลกับค่าความแปรปรวนภายในกลุ่มข้อมูลให้สูงที่สุด เพื่อที่จะการันตีว่าข้อมูลที่เรารวบรวมแล้วนั้นจะแยกออกจากกันมากที่สุด สิ่งที่ต่างกันระหว่าง PCA และ LDA คือ LDA จะไม่เปลี่ยนแปลงตำแหน่งของข้อมูลเดิม แต่จะทำให้กลุ่มของข้อมูลแต่ละกลุ่มให้แยกออกจากกันให้มากที่สุด

Multidimensional Scaling (MDS) คือกลุ่มของเทคนิคที่ให้การประมาณค่าของวัตถุให้เป็นอินพุต และแสดงโครงสร้างออกมาเป็นภาพลักษณะรูปทรงต่างๆ การประมาณค่านี้คือการวัดความเหมือนหรือความต่างกันของข้อมูล ซึ่งผลที่ได้จากการลดมิติลงแล้วนั้นจะรักษาการประมาณค่านี้ให้ห่างออกจกกันให้ได้มากที่สุด

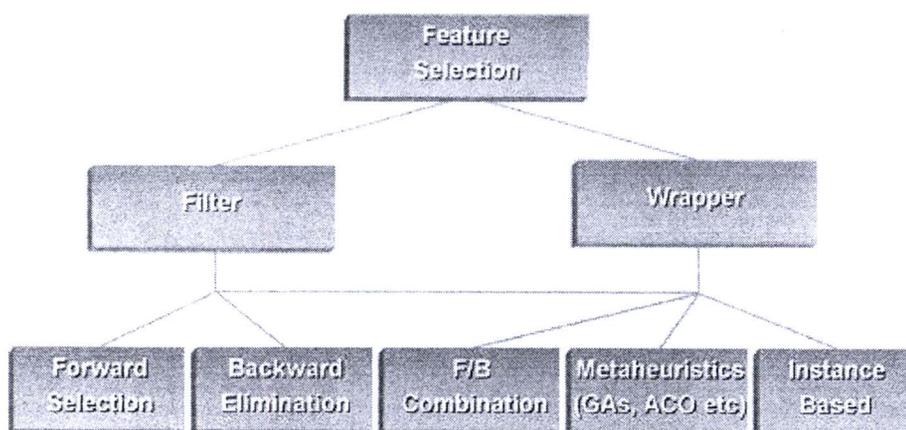
2.2.1.2 วิธีการแบบไม่เป็นเชิงเส้น (Nonlinear Methods)

การลดมิติของข้อมูลแบบวิธีก่อนหน้านี้นั้นมักจะใช้ได้ไม่ดีกับข้อมูลที่มีลักษณะความไม่เป็นเชิงเส้นอยู่ด้วย ทำให้จำเป็นต้องมีวิธีที่สามารถใช้งานกับข้อมูลประเภทนี้ได้

Isomap เป็นการต่อยอดออกมาจาก MDS โดยจะทำการรักษาระยะห่างทางรูปทรงของคู่ข้อมูล ซึ่งประสิทธิภาพของ Isomap นั้นขึ้นอยู่กับความสามารถในการเลือกขนาดของจำนวนข้อมูลข้างเคียงได้มากน้อยเพียงใด ถ้าหากขนาดที่ได้ใหญ่เกินไปจะทำให้แต่ละคู่ข้อมูลนั้นมีระยะห่างจากกันมากเกินไปจนทำให้ข้อมูลที่ลดมิติลงมานั้นผิดเพี้ยนไปได้

2.2.2 SELECTION-BASED REDUCTION

การลดมิติของข้อมูลด้วยเทคนิคนี้เป็นการหาเซตย่อยของข้อมูลที่เล็กที่สุดที่ยังคงให้ความหมายของชุดข้อมูลเดิมอยู่ได้ โดยที่ข้อมูลไม่ได้ถูกเปลี่ยนแปลง เป้าหมายหลักสำหรับวิธีการนี้คือจะอย่างไรให้ได้เซตข้อมูลย่อยที่เล็กที่สุดโดยที่ยังคงได้ค่าความแม่นยำสูงในการบ่งบอกถึงลักษณะข้อมูลเดิมอยู่ เราสามารถจำแนกวิธีการนี้ออกเป็นแผนภาพดังรูปที่ 2.14 จะเห็นได้ว่าสามารถแบ่งออกได้เป็น 2 แบบคือ ฟิลเตอร์ (Filter) และแรปเปอร์ (Wrapper)



รูปที่ 2.14 ลักษณะประเภทของการเลือกลักษณะเด่น [7]

2.3 เทคนิคการลดมิติข้อมูล (Dimensionality Reduction Technique)

ในการเก็บข้อมูลสำหรับงานวิจัยโดยเฉพาะอย่างยิ่งงานที่เกี่ยวกับทางด้านวิทยาศาสตร์ ข้อมูลที่เราได้มานั้นมักจะมีจำนวนมากตามจำนวนตัวแปรที่เราทำการสำรวจและจำนวนครั้งที่เราทำการสำรวจ กล่าวคือข้อมูลที่ได้นั้นมีจำนวนมิติที่มาก โดยคำว่ามิติในที่นี้หมายถึงตัวแปรที่เราทำการสำรวจในการวัดค่าแต่ละครั้ง ความสำเร็จของข้อมูลเหล่านี้ทำให้การคำนวณในการวิเคราะห์และแปลผลทำได้ยากและใช้เวลาในการประมวลผลนาน ทางแก้สำหรับปัญหาดังกล่าวคือใช้เทคนิคในการลดมิติในการพิจารณา ซึ่งเทคนิคที่ใช้อยู่ในปัจจุบันมีอยู่หลายเทคนิค ในส่วนนี้จะขอกล่าวถึงเทคนิคที่นิยมใช้หลักๆ

2.3.1 Principal Component Analysis (PCA)

PCA เป็นวิธีการลดมิติข้อมูลแบบเชิงเส้น ที่อาศัยเมตริกค่าความแปรปรวนร่วมเกี่ยว (Covariance Matrix) ของตัวแปรที่เราสนใจ PCA จะทำการลดมิติลงด้วยการหาผลรวมเชิงเส้นที่มีลักษณะเชิงตั้งฉากกัน เรียกว่า Principal Component (PC) ในจำนวนที่น้อยกว่ามิติของข้อมูลดิบ โดยให้ค่าความแปรปรวนของข้อมูลดิบมากที่สุดเป็น PC_1 คือ ผลรวมเชิงเส้นที่ให้ค่าความแปรปรวนมากที่สุด และ PC ลำดับถัดมาจะให้ค่าความแปรปรวนลดลงตามลำดับจนถึง PC ที่ n

เนื่องจากว่าค่าความแปรปรวนนั้นขึ้นอยู่กับขนาดของค่าที่นำมาคิด ดังนั้นขั้นตอนแรกของการคำนวณจะต้องลบค่าตัวแปรทุกตัวด้วยค่าเฉลี่ยของตัวแปรนั้นๆก่อนเสมอ สมมติว่าค่าข้อมูลดิบถูกปรับให้อยู่ในเมทริกซ์ X เสร็จแล้ว การคำนวณเมทริกซ์ค่าความแปรปรวน Σ สามารถทำได้จากสมการ

$$\Sigma = \frac{1}{n} X X^T \quad (2.1)$$

ใช้หลักการ Eigen-Decomposition สามารถเขียน Σ ให้อยู่ในรูป

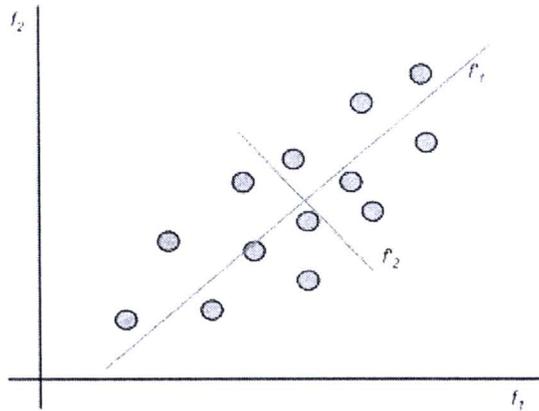
$$\Sigma = U \Lambda U^T \quad (2.2)$$

โดย $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ เป็นเมทริกซ์ทแยงมุมที่เรียงลำดับของ Eigenvalue จากน้อยไปมาก และ U คือ เมทริกซ์ตั้งฉากที่เก็บค่า Eigenvector มีขนาดเท่ากับ $p \times p$ ดังนั้น PC แต่ละตัวสามารถหาได้จากแถวที่ p ของเมทริกซ์ S ที่เป็นไปตามสมการที่ 2.3

$$S = U^T X \quad (2.3)$$

จะเห็นได้ว่า U^T คือเมทริกซ์น้ำหนัก W ที่จะทำการฉายฐานข้อมูลจากปริภูมิเดิม ไปยัง ปริภูมิต่ำกว่าที่มีขนาดมิติน้อยลงดังสมการที่ 2.4

$$S_{k \times n} = W_{k \times p} X_{p \times n} \quad (2.4)$$



รูปที่ 2.15 ตัวอย่างการจำแนกข้อมูลด้วย PCA ใน Feature Space 2 มิติ [8]

2.3.2 Linear Discriminant Analysis (LDA)

LDA คือการคำนวณเพื่อเพิ่มความแตกต่างเชิงเส้นของจุดข้อมูลที่อยู่ในกลุ่มที่ต่างกันให้มากที่สุด เป็นวิธีการแบบมีผู้สอน โดย LDA จะหา Linear Mapping M ที่ทำให้กลุ่มของข้อมูลที่ได้ในมิติที่ต่ำลง มีแยกกันเชิงเส้นในแต่ละกลุ่มมากที่สุด สิ่งที่จะนำมาพิจารณาความแยกกันเชิงเส้นดังกล่าวนี้คือ Within-Class Scatter (S_w) และ Between-Class Scatter (S_b) ซึ่งมีนิยามว่า

$$S_w = \sum_c p_c \text{COV}(\mathbf{X}^c - \bar{\mathbf{X}}^c) \quad (2.5)$$

$$S_b = \text{COV}(\mathbf{X} - \bar{\mathbf{X}}) - S_w \quad (2.6)$$

เมื่อ p_c เป็น ค่าความเป็นสมาชิกของข้อมูลนั้นในกลุ่ม c

$\text{COV}(\mathbf{X}^c - \bar{\mathbf{X}}^c)$ คือ เมทริกซ์ความแปรปรวนร่วมของข้อมูลในกลุ่ม c ที่ลบกับค่าเฉลี่ยข้อมูลในกลุ่ม c

$\text{COV}(\mathbf{X} - \bar{\mathbf{X}})$ คือ เมทริกซ์ความแปรปรวนร่วมของข้อมูลทั้งหมดที่ลบกับค่าเฉลี่ยข้อมูลทั้งหมด

LDA จะทำการปรับให้ได้ค่าอัตราส่วนระหว่าง Within-Class Scatter และ Between-Class Scatter ในข้อมูลที่ทำให้การลดมิติให้มีค่าสูงที่สุด โดยการหา Linear Mapping M ที่ให้ค่า Fisher Index สูงที่สุด ดังสมการ

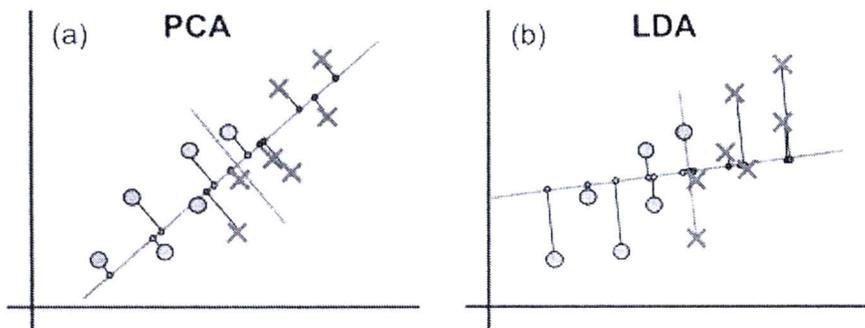
$$\phi(M) = \frac{M^T S_b M}{M^T S_w M} \quad (2.7)$$

เมื่อปรับได้ค่าที่สูงที่สุดแล้ว M จะถูกนำมาใช้เพื่อแปลงค่าข้อมูลดิบ M ให้เป็นข้อมูลที่มีมิติที่น้อยลง Y

$$Y_{k \times n} = M_{k \times p} X_{p \times n} \quad (2.8)$$

LDA เป็นวิธีการที่ง่ายและรวดเร็ว แต่ประสิทธิภาพของ LDA ขึ้นอยู่กับว่าข้อมูลที่นำมาใช้ในการเรียนรู้ระบบนั้นมีความแม่นยำมากน้อยเพียงใด อีกทั้งจำนวนของข้อมูลจำเป็นต้องมีมากอีกด้วยเพราะหากใช้ข้อมูลที่มีจำนวนน้อยเกินไป อาจจะทำให้ระบบของเราคือประสิทธิภาพลงได้ [6]

รูปที่ 2.16 แสดงการเปรียบเทียบการลดมิติของข้อมูลจาก 2 มิติให้เหลือเพียง 1 มิติ ด้วย PCA และ LDA แล้วพิจารณาจำแนกกลุ่มข้อมูล ในรูป (a) จะเห็นได้ว่า PCA ไม่สามารถแยกข้อมูลออกจากกันได้ดี แต่ในขณะที่ ในรูป (b) LDA จะสามารถหาการโปรเจกต์ที่แยกกลุ่มของข้อมูลได้ดีกว่า



รูปที่ 2.16 เปรียบเทียบการจำแนกกลุ่มข้อมูลด้วย PCA และ LDA [8]

2.3.3 Neighborhood Components Analysis (NCA)

เป็นวิธีการเรียนรู้แบบมีผู้สอน (Supervised Learning) ที่ J. Goldberger และคณะ ได้นำเสนอขึ้นในปี ค.ศ. 2004 [12] เพื่อใช้สำหรับจัดแบ่งกลุ่มข้อมูลที่แตกต่างกันให้เป็นกลุ่มที่แยกออกจากกันอย่างชัดเจนตามเมตริกระยะทางที่ได้กำหนดให้ ซึ่งคือการวัดระยะทาง Mahalanobis การใช้งาน NCA นั้น มีวัตถุประสงค์เช่นเดียวกับวิธีการจำแนกข้อมูลแบบ k-Nearest Neighborhood (k-NN)

เป้าหมายของวิธีการ NCA คือการเรียนรู้เมตริกระยะทางโดยค้นหาการแปลงเชิงเส้น (Linear Transformation) ของข้อมูลขาเข้าที่ทำให้ประสิทธิภาพการจำแนกประเภทเฉลี่ยแบบนำตัวหนึ่งออกไป (Leave-One-Out; LOO) มีค่าสูงที่สุด สำหรับปริภูมิของข้อมูลที่ถูกแปลงเสร็จแล้ว ใจความสำคัญ

ของขั้นตอนวิธีก็คือ เมทริกซ์ A ที่เกี่ยวข้องกับการแปลงค่า นั้น สามารถหาได้โดยการกำหนด ฟังก์ชันจุดประสงค์ (Objective Function) ที่หาอนุพันธ์ได้ให้ก่อนเมตริก A หลังจากนั้นจึงใช้ตัวแก้ ปัญหาแบบวนซ้ำ เช่น Conjugate Gradient Descent เป็นต้น ข้อดีของขั้นตอนวิธีการแบบนี้คือ จำนวน ของกลุ่มข้อมูล k นั้นสามารถประเมินได้ในลักษณะที่เป็นตั้งแต่ฟังก์ชันของ A ไปจนถึงค่าคงที่ค่า หนึ่ง

การที่จะนิยามเมทริกซ์ A นั้นจำเป็นต้องกำหนดฟังก์ชันจุดประสงค์ที่อธิบายถึงความแม่นยำในการ จำแนกกลุ่มในปริภูมิที่ได้ทำการแปลงไปแล้ว และพยายามหาค่า A ที่ทำให้ฟังก์ชันจุดประสงค์มีค่า มากที่สุด

$$\hat{A} = \operatorname{argmax}_A f(A) \quad (2.9)$$

พิจารณาการทำนายชนิดของกลุ่มของข้อมูลหนึ่งจุดโดยอาศัยความสัมพันธ์กันระหว่าง k-Nearest Neighborhood ของข้อมูลจุดนั้นๆ กับเมตริกระยะทางที่กำหนดไว้ วิธีการนี้เรียกว่าการจำแนกกลุ่ม แบบ Leave-One-Out อย่างไรก็ตาม เซตของข้อมูลใกล้เคียงที่ใกล้ที่สุดลำดับที่ C_i สามารถแตกต่างกัน ได้หลังจากทำการแปลงจุดทุกจุดด้วยการแปลงเชิงเส้น

ปัญหาเหล่านี้สามารถแก้ไขได้ด้วยการใช้ขั้นตอนวิธีทาง Stochastic Gradient Descent โดยจะพิจารณา เซตข้อมูลที่ถูกแปลงทั้งหมดให้อยู่ในลักษณะของ Stochastic Nearest Neighbors แทนการพิจารณา k-Nearest Neighborhood ที่แต่ละจุดของข้อมูลที่ทำกรแปลงแล้วในการจำแนกแบบ LOO แล้วกำหนด ให้ใช้ฟังก์ชัน Softmax ของระยะทางยูคลิเดียนยกกำลังสองระหว่างหนึ่งจุดใดๆในการจำแนกแบบ LOO กับแต่ละจุดที่อยู่ในปริภูมิที่ทำการแปลง ดังสมการที่ 2.10

$$p_{ij} = \begin{cases} \frac{e^{-\|x_i - x_j\|^2}}{\sum_k e^{-\|x_i - x_k\|^2}}, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases} \quad (2.10)$$

โดยความน่าจะเป็นของการจำแนกกลุ่มข้อมูลจุด i ได้ถูกต้องนั้นหมายถึง ค่าความน่าจะเป็นของการ จำแนกจุดแต่ละจุดที่อยู่ข้างเคียง C_i ดังสมการ

$$p_i = \sum_j p_{ij} \quad (2.11)$$

เมื่อ p_{ij} คือความน่าจะเป็นของการที่จำแนกได้ว่า j เป็นจุดใกล้เคียงกับ i

ณ จุดนี้สามารถกำหนดฟังก์ชันจุดประสงค์โดยใช้การจำแนกกลุ่มแบบ LOO โดยอาศัยข้อมูลทั้งหมด เหมือนกับ Stochastic Nearest Neighbours

$$f(A) = \sum_i \sum_{j \in C} p_{ij} = \sum_i p_i \quad (2.12)$$

ฟังก์ชันจุดประสงค์มักจะเลือกเอาเฉพาะที่สามารถหาอนุพันธ์เทียบกับ A ได้ดังสมการที่ 2.14

$$\frac{\partial f}{\partial A} = -2A \sum_i \sum_{j \in C} p_{ij} (x_{ij} x_{ij}^T - \sum_k p_{ik} x_{ik} x_{ik}^T) \quad (2.13)$$

$$= 2A \sum_i (p_i \sum_k p_{ik} x_{ij} x_{ij}^T - \sum_{i \in C} p_{ij} x_{ij} x_{ij}^T) \quad (2.14)$$

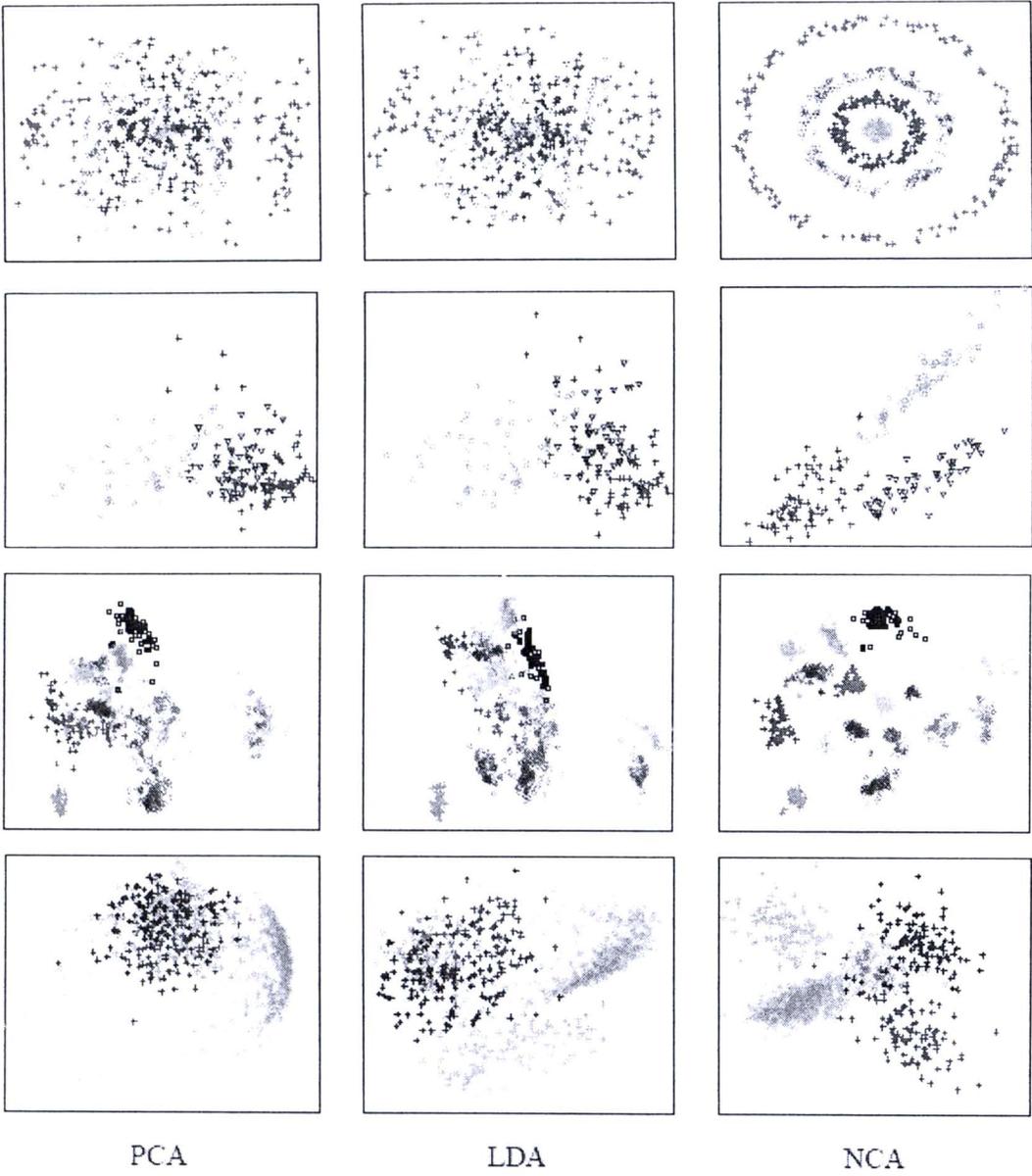
การที่จะหาค่าความชันของเมทริกซ์ A นั้นหมายถึงหาค่าความชันนั้นสามารถหาค่าได้ด้วยตัวแก้ ปัญหาแบบวนซ้ำเช่น Conjugate Gradient Descent โดยในทางปฏิบัตินั้น พจน์ข้างในสุดของค่าความชันถูกประเมินว่าไม่มีผลกับการหาค่า ดังนั้นพจน์รวมในสุดของค่าความชันจึงสามารถถูกตัดออกไปได้ ส่งผลให้เวลาที่ใช้ในการคำนวณลดลงแม้ว่าข้อมูลที่นำมาใช้คำนวณจะมีมากก็ตาม

การนำเอา NCA ไปใช้สำหรับลดมิติของข้อมูลแบบเชิงเส้นสามารถทำได้เช่นเดียวกับกระบวนการ จำแนกกลุ่ม โดยกำหนดให้เมทริกซ์ A ที่จะทำการหานั้นให้มีมิติเป็นสี่เหลี่ยมจัตุรัสขนาด $d \times D$ โดยที่ $d \ll D$ แล้วทำการเรียนรู้เซตของข้อมูล หลังจากทำการเรียนรู้เสร็จสิ้น ค่าของเมทริกซ์ A จะถูกเก็บไว้เพื่อใช้ในการแปลงข้อมูลที่จะนำมาทดสอบ x_{test} ให้เป็นข้อมูลในปริภูมิ ใหม่ y_{test} ดังสมการที่ 2.15

$$y_{test} = A \cdot x_{test} \quad (2.15)$$

รูปที่ 2.15 แสดงตัวอย่างการเปรียบเทียบประสิทธิภาพในการลดมิติของเซตข้อมูลให้เหลือ 2 มิติ ระหว่าง 3 วิธีได้แก่ PCA LDA และ NCA โดยเซตข้อมูลที่นำมาทดสอบประกอบไปด้วย 4 กลุ่มข้อมูล คือ “Concentric Rings” “Wine” “Faces” และ “Digits” ที่มีขนาดมิติเดิมเท่ากับ 3 13 560 และ 256 ตามลำดับ ซึ่งจากผลการทดสอบพบว่าข้อมูลในแต่ละกลุ่มที่ทำการแปลงโดยใช้วิธี NCA นั้น สามารถแยกกันได้ดีกว่าวิธีที่ใช้ PCA และ LDA





รูปที่ 2.17 ผลการทดลองลดมิติข้อมูลให้เหลือ 2 มิติ ด้วยวิธี PCA LCA และ NCA [12]

2.4 วิธีการสกัดค่าลักษณะเด่นข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อทางสถิติเชิงเวลา

เป็นวิธีการคำนวณค่าทางสถิติของข้อมูลโดยพิจารณาข้อมูลที่อยู่ในโดเมนเวลา ซึ่งมีวิธีการวัดค่าอยู่หลายวิธีดังต่อไปนี้

2.4.1 จำนวนของ zero crossing

เป็นการวัดค่าความถี่ของสัญญาณแบบง่าย ๆ โดยนับจำนวนครั้งที่สัญญาณตัดผ่านค่าศูนย์ และมีค่า threshold (ϵ) มาใช้ในการคำนวณเพื่อลดสัญญาณรบกวน สมการในการคำนวณจะกำหนดให้ในการวัดค่า x_k และ x_{k+1} จะทำการเพิ่มจำนวน zero crossing ถ้าหาก

$$\{x_k > 0, x_{k+1} < 0\} \text{ หรือ } \{x_k < 0, x_{k+1} > 0\} \quad (2.16)$$

และ

$$|x_k - x_{k+1}| \geq \epsilon$$

2.4.2 ความยาวรูปคลื่น (Waveform length, WL)

เป็นข้อมูลที่แสดงถึงความซับซ้อนของรูปคลื่นอันประกอบไปด้วยค่า amplitude ความถี่ และช่วงความยาวภายในตัวแปรเดียว โดยสามารถคำนวณได้จาก

$$l_0 = \sum_{k=1}^L |\Delta x_k| \quad (2.17)$$

เมื่อ $\Delta x_k = x_k - x_{k-1}$

2.4.3 จำนวนของการเปลี่ยนเครื่องหมายของค่าความชัน

เป็นข้อมูลที่แสดงค่าในเชิงความถี่อีกข้อมูลหนึ่ง คล้ายกับจำนวนของ zero crossing โดยกำหนดใช้ในการคำนวณ x_{k-1} x_k และ x_{k+1} จำนวนของการเปลี่ยนเครื่องหมายของค่าความชันจะเพิ่มขึ้นถ้าหาก

$$\{x_k > x_{k-1}, x_k > x_{k+1}\} \text{ หรือ } \{x_k < x_{k-1}, x_k < x_{k+1}\} \quad (2.18)$$

และ

$$|x_k - x_{k+1}| \geq \epsilon \text{ หรือ } |x_k - x_{k-1}| \geq \epsilon \quad (2.19)$$

2.4.4 ค่าอินทีเกรตของสัญญาณกล้ามเนื้อ (Mean Absolute Value, MAV)

เป็นปริมาณที่หาได้จากค่าสัมบูรณ์ของสัญญาณ ที่ได้จากการสุ่มภายในช่วงเวลาจำนวน L ค่า สามารถคำนวณได้จาก

$$\bar{x}_i = \frac{1}{L} \sum_{k=1}^L |x_k| \quad \text{for } i = 1, \dots, I \quad (2.20)$$

เมื่อ x_k คือการวัดค่าสัญญาณครั้งที่ k ในกลุ่มข้อมูล i และ I คือจำนวนของกลุ่มข้อมูล ทั้งหมดที่วัด

2.4.5 ค่าความแปรปรวน (Variance, VAR)

เป็นค่าที่ใช้วัดข้อมูลความหนาแน่นของพลังงานของสัญญาณกล้ามเนื้อ ซึ่งหากค่านี้มีปริมาณสูงแล้ว จะหมายถึงว่า สัญญาณกล้ามเนื้อมีการเปลี่ยนแปลงมาก และบ่งชี้ว่าสัญญาณมีความหนาแน่นพลังงาน มาก สามารถคำนวณได้จาก

$$VAR = \frac{1}{L-1} \sum_{k=1}^L x_k^2 \quad \text{for } i = 1, \dots, I \quad (2.21)$$

เมื่อ x_k คือการวัดค่าสัญญาณครั้งที่ k ในกลุ่มข้อมูล i และ I คือจำนวนของกลุ่มข้อมูล ทั้งหมดที่วัด

2.4.6 ค่าขนาดวิลลิสัน (Willison Amplitude, WAMP)

ลักษณะเด่นนี้นิยามไว้ว่าเป็นจำนวนครั้งที่มีการเปลี่ยนแปลงของขนาดของสัญญาณ EMG ที่มาก เกินกว่าเส้นขีดแบ่ง (ϵ) ที่กำหนดเอาไว้ ซึ่งค่านี้เป็นตัวที่บ่งบอกการทำงานของกล้ามเนื้อส่วนที่ทำการ เคลื่อนไหว และเป็นตัววัดระดับของการหดของกล้ามเนื้ออีกด้วย ค่าเส้นขีดแบ่งที่เหมาะสมคือ ประมาณ 50 และ 100 มิลลิโวลต์ [10] สามารถคำนวณได้จาก

$$WAMP = \sum_{k=1}^N f(|x_k - x_{k+1}|) \quad (2.22)$$

$$\text{โดย} \quad f(x) = \{1 \text{ if } x > \epsilon\} \quad (2.23)$$

2.4.7 ค่า v-Order

ค่าวัดนี้บ่งบอกถึงการประมาณของแรงที่กล้ามเนื้อใช้อยู่ สามารถคำนวณได้จาก

$$vOrder = \sqrt[v]{E\{|x_k|^v\}} \quad (2.24)$$

เมื่อ E คือ ค่าคาดหวัง

จากการศึกษาพบว่าค่า v ที่เหมาะสมที่สุดคือ 2 ซึ่งทำให้สมการนี้มีค่าเช่นเดียวกับความแปรปรวน

2.4.8 ค่า log-Detector

ลักษณะเด่นนี้จะใช้การประมาณแรงที่ใช้ในการเกร็งกล้ามเนื้อเช่นเดียวกับค่า vOrder สามารถคำนวณได้จาก

$$\log Detect = e^{\frac{1}{N} \sum_{i=1}^N \log(|x_i|)} \quad (2.25)$$

2.4.9 ฮิสโตแกรมของ EMG (EMG Histogram)

ลักษณะเด่นนี้ให้ข้อมูลเกี่ยวกับความถี่ของสัญญาณกล้ามเนื้อในขนาดต่างๆ โดยข้อมูลจะถูกแบ่งเป็นช่วงตามกำหนด อาสาสมัครแต่ละคนก็จะมีค่าของแต่ละช่วงไม่เหมือนกัน เนื่องจากขนาดของสัญญาณที่ได้นั้นไม่เท่ากัน แต่ถือว่าไม่มีผลกับการจำแนกสัญญาณเพราะว่าโมเดลในการจำแนกสัญญาณนั้นแยกเฉพาะบุคคลอยู่แล้ว

2.4.10 ค่าสัมประสิทธิ์ออโตรีเกรสชัน (Autoregression Coefficient)

ลักษณะเด่นนี้จะโมเดลสัญญาณกล้ามเนื้อแต่ละตัวให้เป็นสัญญาณทางเวลาแบบเส้นตรงของออโตรีเกรสชัน และให้ข้อมูลเกี่ยวกับสถานะการหดตัวของกล้ามเนื้อ สามารถคำนวณได้จาก

$$x_k = \sum_{i=1}^p a_i x_{k-i} + e_k \quad (2.26)$$

เมื่อ a_i คือ ค่าสัมประสิทธิ์ของออโตรีเกรสชันลำดับที่ i

p คือ ค่าลำดับของโมเดลออโตรีเกรสชัน

e_k คือ สัญญาณรบกวนสีขาว (White Noise)

2.4.11 ค่าสัมประสิทธิ์เคปสตรัม (Cepstrum Coefficients)

เคปสตรัมของสัญญาณคือผลที่ได้จากการหาการแปลงกลับการแปลงฟูริเยร์ของค่าลอการิทึมของขนาดของการแปลงฟูริเยร์ในสัญญาณหนึ่งๆ ซึ่งค่าที่ได้จะแสดงถึงอัตราการเปลี่ยนแปลงของสัญญาณในแต่ละช่วงความถี่ ค่าสัมประสิทธิ์เคปสตรัมสามารถคำนวณได้จากโมเดลของโอโตรีเกรสชันดังนี้

$$c_1 = -a_1 \quad (2.27)$$

$$c_i = -a_i - \sum_{l=1}^{i-1} \left(1 - \frac{l}{i}\right) a_l c_{i-l} \quad (2.28)$$

- เมื่อ a_i คือ ค่าสัมประสิทธิ์ของออโตรีเกรสชัน i
 c_i คือ ค่าสัมประสิทธิ์เคปสตรัมลำดับที่ i
 i คือ มิติของโมเดล

จะเห็นได้ว่าค่าสัมประสิทธิ์นี้ไม่จำเป็นต้องหาการแปลงฟูริเยร์ดังนั้นจึงจัดได้ว่าเป็นลักษณะเด่นเชิงเวลา

2.5 การสกัดลักษณะเด่นข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อเชิงความถี่

ลักษณะเด่นหรือ Feature ของสัญญาณไฟฟ้ากล่อมเนื้อสามารถสกัดออกมาได้จากข้อมูลที่อยู่ในหลากหลายโดเมน เช่น โดเมนเวลา หรือ โดเมนความถี่ เป็นต้น ซึ่งข้อมูลเชิงความถี่ที่จะนำไปสกัดลักษณะเด่นนั้นเป็นผลที่ได้มาจากการแปลงฟูริเยร์ (Fourier transform) ของข้อมูลสัญญาณเชิงเวลา จากการศึกษเพิ่มเติมพบว่าเราสามารถสกัดเอาลักษณะเด่นจากข้อมูลที่อยู่ในรูปเชิงความถี่ หรือ โดเมนความถี่ได้อีก 2 วิธี [1] คือ

- 1). Median Frequency (MDF)
- 2). Mean Frequency (MNF)

2.5.1 ความถี่มัชฌิมฐาน (Median Frequency)

คือ ค่าความถี่ ณ ตำแหน่งที่สเปกตรัมนั้นถูกแบ่งออกเป็นสองบริเวณที่มีกำลังเท่าๆกัน สามารถเขียนออกมาเป็นได้ดังสมการที่ 2.29

$$\sum_{j=1}^{MDF} P_j = \sum_{i=MDF}^M P_j = \frac{1}{2} \sum_{i=1}^M P_i \quad (2.29)$$

เมื่อ P_j คือค่าสเปกตรัมกำลัง ณ ความถี่ลำดับที่ j และ M คือลำดับความถี่สุดท้ายที่มีในข้อมูล

2.5.2 ความถี่เฉลี่ย (Mean Frequency)

คือ ค่าความถี่เฉลี่ยของสัญญาณไฟฟ้ากล่อมเนื้อ สามารถคำนวณได้จากผลรวมของผลคูณระหว่างค่าสเปกตรัมกำลังกับค่าความถี่ของสเปกตรัมนั้นๆ แล้วหารด้วยความเข้มของสเปกตรัมกำลังทั้งหมด ดังสมการที่ 2.30

$$MNF = \frac{\sum_{j=1}^M f_j P_j}{\sum_{j=1}^M P_j} \quad (2.30)$$

เมื่อ f_j คือค่าความถี่ของสเปกตรัม ณ ความถี่ลำดับที่ j

2.6 การสกัดลักษณะเด่นเชิงเวลาและสเกล

ในการแปลงฟูริเยร์แบบทั่วไปแล้ว สัญญาณที่ได้ออกมานั้นจะสูญเสียข้อมูลในเชิงเวลาออกไป กล่าวคือ เราไม่สามารถทราบได้ว่าค่าความถี่ที่เกิดขึ้นนั้น ได้เกิดขึ้น ณ ช่วงเวลาใด ซึ่งการแปลงแบบนี้เป็นที่ยอมรับได้สำหรับสัญญาณที่ไม่มีเปลี่ยนแปลงเมื่อเทียบกับเวลา แต่สำหรับสัญญาณกล่อมเนื้อแล้วนั้นรูปแบบสัญญาณจะอยู่ในลักษณะของสัญญาณที่มีการเปลี่ยนแปลงเมื่อเทียบกับเวลา ดังนั้นหากเราต้องการจะทราบทั้งข้อมูลของความถี่และข้อมูลทางเวลา จำเป็นต้องใช้ข้อมูลที่เป็นลักษณะเด่นเชิงเวลา-ขนาด ซึ่งมีอยู่ 3 ตัวดังนี้

2.6.1 การแปลงฟูริเยร์ช่วงเวลานั้น (Short Time Fourier Transform)

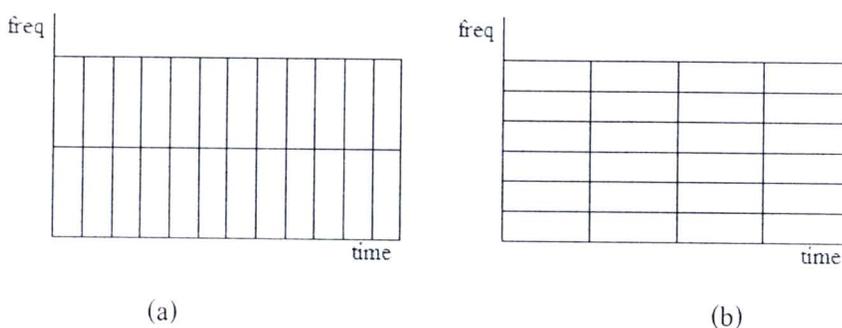
การแปลงสัญญาณแบบฟูริเยร์ช่วงเวลานั้น จะทำการแปลงสัญญาณไปเป็นฟังก์ชันขนาด 2 มิติ ระหว่าง เวลาและค่าความถี่ โดยการแปลงนั้นจะใช้ฟังก์ชันวินโดว์มาคูณกับข้อมูลสัญญาณในช่วงเวลานั้นๆ สามารถเขียนเป็นสมการได้ดังนี้

$$STFT(x(t)) \equiv X(\tau, \omega) \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-j\omega t} dt \quad (2.31)$$

เมื่อ $w(t)$ คือ ฟังก์ชันวินโดว์

$x(t)$ คือ สัญญาณที่จะนำมาแปลง

แต่ข้อเสียของการแปลงแบบนี้คือปัญหาเรื่องความละเอียดของสัญญาณที่ได้นั้นจะมีค่าตายตัว ทำให้เราจำเป็นต้องเลือกว่าจะให้ฟังก์ชันที่แปลงออกมาได้นั้นมีความละเอียดในเชิงเวลาหรือความถี่มากกว่ากัน ซึ่งค่าความละเอียดนั้นจะถูกกำหนดโดยความยาวของฟังก์ชันวินโดว์ที่นำมาคูณ หากทำการแปลงในช่วงเวลาสั้นๆ ข้อมูลที่ได้ออกมาก็จะมีความละเอียดในเชิงเวลามากกว่าเชิงความถี่ ดังแสดงในรูปที่ 2.18(a) และในทางตรงกันข้าม หากเราเลือกใช้ฟังก์ชันวินโดว์ที่มีขนาดยาวก็จะได้ความละเอียดเชิงความถี่ มากกว่าข้อมูลเชิงเวลา รูปที่ 2.18 (b)

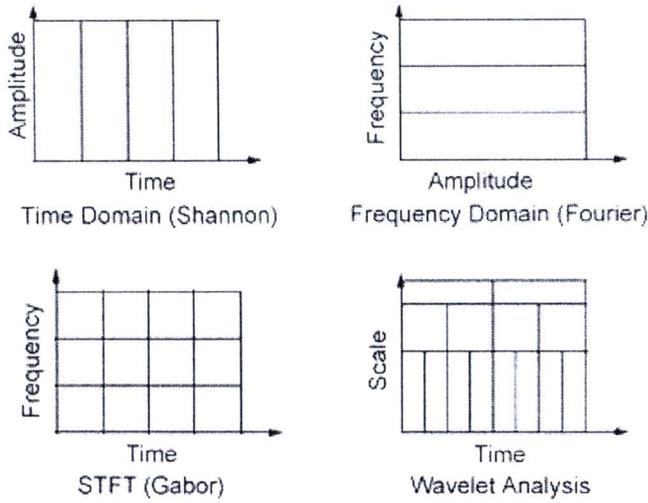


รูปที่ 2.18 แสดงการเปรียบเทียบความละเอียดของการแปลงฟูรีเยร์ช่วงเวลาสั้น[9]

2.6.2 การแปลงเวฟเลต (Wavelet Transform)

การแปลงเวฟเลตทำให้เราสามารถวิเคราะห์สัญญาณในบริเวณเฉพาะช่วงสั้นๆของสัญญาณทั้งหมดได้ โดยการวิเคราะห์สัญญาณด้วยเวฟเลตนั้นจะทำให้เราได้รูปแบบของข้อมูลที่วิธีการอื่นๆไม่สามารถทำได้ เช่น แนวโน้ม จุดหักมุม ความไม่ต่อเนื่องในลำดับอนุพันธ์ที่สูงขึ้นไป และความคล้ายกันของสัญญาณ รูปที่ 2.19 แสดงความแตกต่างของโดเมนในการวิเคราะห์สัญญาณในแบบต่างๆ

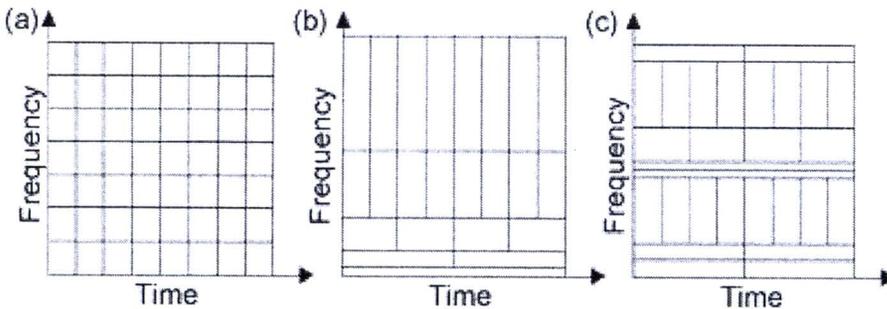
ในการวิเคราะห์เวฟเลตนั้น ค่าสเกลและความถี่มีความสัมพันธ์กันอยู่ กล่าวคือ ในช่วงที่สเกลมีขนาดเล็กจะแสดงให้เห็นถึงรายละเอียดการเปลี่ยนแปลงที่รวดเร็วของสัญญาณที่มีความถี่สูง และช่วงที่สเกลมีขนาดใหญ่จะแสดงการเปลี่ยนแปลงที่ช้าของสัญญาณความถี่ต่ำ การแปลงเวฟเลตจึงเปรียบเสมือนกล้องจุลทรรศน์ทางคณิตศาสตร์ที่สามารถมองไปยังส่วนต่างๆของสัญญาณได้ด้วยการปรับโฟกัส



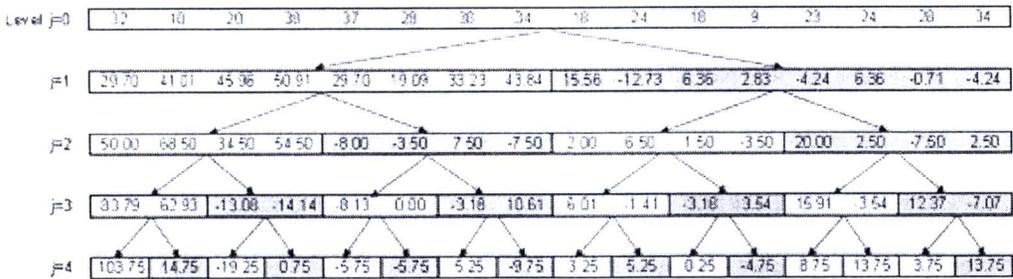
รูปที่ 2.19 รูปแบบการแบ่งช่วงวิเคราะห์สัญญาณในแต่ละโดเมน[9]

2.6.3 การแปลงเวฟเลตแพ็คเกจ (Wavelet Packet Transform)

การแปลงเวฟเลตแพ็คเกจนั้นจะทำให้เราสามารถวิเคราะห์สัญญาณใน โดเมนของเวลาและสเกลได้ดีที่สุด เนื่องจากการแปลงแบบนี้จะทำให้การแบ่งช่วงการพิจารณานั้นสามารถปรับเปลี่ยนได้มากกว่าการแปลงแบบ 2 วิธีข้างต้น ซึ่งขึ้นอยู่กับว่าจะเลือกใช้ในแต่ละงานอย่างไร รูปที่ 2.20 แสดงการเปรียบเทียบการแบ่งของวิธีการแปลงค่าทั้ง 3 วิธีซึ่งจะเห็นได้ว่าการแปลงแบบเวฟเลตแพ็คเกจจะทำให้การแบ่งช่วงนั้นมีความหลากหลายมากกว่า ในรูปที่ 2.20 (c) แสดงนี้เป็นเพียงการแบ่งแบบใดๆเท่านั้น การแปลงแบบเวฟเลตแพ็คเกจสามารถแบ่งได้อีกหลายรูปแบบ



รูปที่ 2.20 แสดงการเปรียบเทียบการแบ่งช่วงของ (a) STFT (b) WT และ (c) WPT [10]



รูปที่ 2.21 แผนผังต้นไม้ของการแปลงแบบ Wavelet Packet [11]

ในปี ค.ศ. 2009 ของ K. Kiatpanichagij และ N. Afzulpurkan[11] งานวิจัยชิ้นนี้ได้มีการใช้วิธีที่เรียกว่าการแปลงเวฟเลตแพ็คเกจ (WPT) เพื่อแปลงข้อมูลสัญญาณกล้ามเนื้อให้อยู่ในรูปเชิงเวลา-ความถี่หลายระดับดังแสดงในรูปที่ 2.21 ช่วงสัญญาณย่อยที่มีพื้นหลังสีขาวคือช่วงสัญญาณย่อยในส่วนของความถี่ต่ำ และพื้นหลังสีเทาคือช่วงสัญญาณย่อยที่เป็นสัญญาณความถี่สูง งานวิจัยชิ้นนี้ได้เลือกใช้ช่วงสัญญาณย่อยในระดับที่ 3 สำหรับการคำนวณในลำดับต่อไป ซึ่งเมื่อแปลงค่าสัญญาณออกมาแล้วจะทำให้ได้ค่าสัมประสิทธิ์พลังงานของ WPT นิยามให้เป็น

$$X = \{x_i | x_i = (s_i)^2, i = 1, \dots, n\} \quad (2.32)$$

เมื่อ s_i คือ ค่าสัมประสิทธิ์ลำดับที่ i ของช่วงสัญญาณย่อย

n คือ จำนวนสัมประสิทธิ์ของช่วงสัญญาณย่อย

แต่การใช้ค่าสัมประสิทธิ์พลังงานมาจำแนกกลุ่มสัญญาณโดยตรงนั้นมีปัญหาในเรื่องสัญญาณรบกวน อันเนื่องมาจากกระบวนการแปลงค่า และเพื่อให้ได้ค่าที่มีความเสถียรมากกว่ากลุ่มผู้วิจัยจึงได้เลือกใช้ลักษณะเด่นทางสถิติของค่าสัมประสิทธิ์พลังงานอีก 3 เซต เพิ่มเติม มีรายละเอียดของแต่ละเซต

เซตที่ 1: ค่าพลังงานเฉลี่ย

คำนวณได้จากสมการ

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.33)$$

ดังนั้นในแต่ละช่องสัญญาณจะประกอบไปด้วยค่าพลังงานเฉลี่ยนี้ทั้งหมด 8 ค่า

เซตที่ 2: ค่าพลังงานเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน(S.D.) ค่าความเบ้ (Skewness) และภาวะยอดมน (Kurtosis)

ในเซตนี้ค่าทางสถิติอีก 3 ค่าที่เพิ่มเติมมาสามารถคำนวณได้จากสมการต่อไปนี้



$$S.D. = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.34)$$

$$Skewness = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3 \quad (2.35)$$

$$Kurtosis = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (2.36)$$

ดังนั้นในเซตที่ 2 นี้แต่ละช่องสัญญาณจะประกอบไปด้วยข้อมูลทั้งหมดจำนวน 32 ค่า

เซตที่ 3: ค่าโมเมนต์เชิงเวลา ลำดับที่ 1 ถึง 5

ค่าโมเมนต์เชิงเวลาในแต่ละลำดับสามารถคำนวณได้จาก

$$M_m = \sum_{i=1}^n \left(\frac{i}{n} \right)^m x_i \quad (2.37)$$

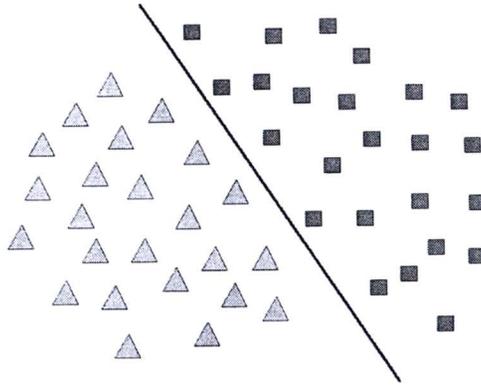
เมื่อกำหนดให้ m คือลำดับของโมเมนต์เริ่มจาก 0 ถึง 5

ดังนั้นเซตที่ 3 จะมีจำนวนข้อมูลของแต่ละช่องสัญญาณเท่ากับ 6 ค่า

ข้อมูลทั้ง 3 เซตจะถูกนำไปใช้ในการจำแนกสัญญาณโดยแยกกันพิจารณาเพื่อเปรียบเทียบการใช้งานว่ากลุ่มใดให้ประสิทธิภาพในการจำแนกสูงที่สุด

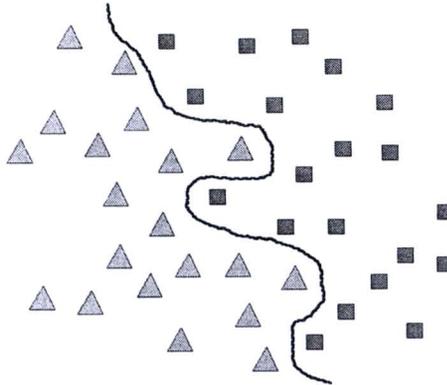
2.7 Support Vector Machine (SVM)

SVM เป็นวิธีการจำแนกกลุ่มข้อมูลที่มีพื้นฐานมาจากหลักการของระนาบการตัดสินใจ (Decision Planes) ที่ทำหน้าที่เป็นเส้นขอบของการตัดสินใจ โดยระนาบการตัดสินใจคือสิ่งที่แบ่งแยกระหว่างเซตของวัตถุที่อยู่ในกลุ่มที่แตกต่างกันดังแสดงตัวอย่างในรูปที่ 2.22 จะเห็นได้ว่าวัตถุจะถูกแบ่งให้อยู่ในกลุ่มสี่เหลี่ยมหรือสามเหลี่ยม เส้นแบ่งสี่ด้านบอกถึงขอบเขตว่าด้านขวาของเส้นจะเป็นวัตถุสี่เหลี่ยมและด้านซ้ายคือวัตถุสามเหลี่ยม ซึ่งเมื่อมีจุดใดๆเข้ามาใหม่ ระบบก็สามารถแยกแยะรูปทรงวัตถุได้โดยอาศัยเส้นดังกล่าว



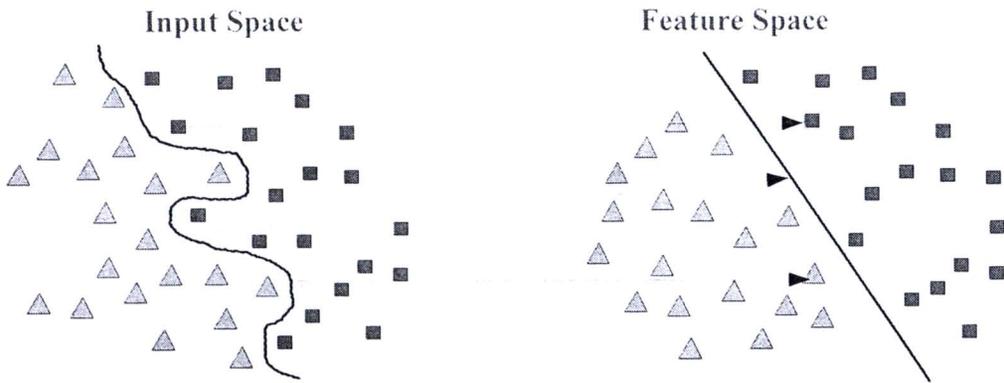
รูปที่ 2.22 ตัวอย่างเส้นตัดสินใจแบบเชิงเส้น สำหรับจำแนกกลุ่มวัตถุสามเหลี่ยมและสี่เหลี่ยม

แต่สำหรับการจำแนกกลุ่มข้อมูลโดยทั่วไปนั้นจะไม่ง่ายดังที่แสดงในรูปที่ 2.23 และมักจะจำเป็นต้องใช้โครงสร้างที่มีความซับซ้อนมากขึ้นเพื่อให้สามารถจำแนกกลุ่มข้อมูลได้ดีที่สุด เมื่อเปรียบเทียบกับรูปแบบที่ผ่านมานั้นจะเห็นได้ชัดว่าในรูปที่ 2.23 สิ่งที่สามารถแบ่งกลุ่มข้อมูลสี่เหลี่ยมและสี่แฉงออกจากกันได้ดีที่สุดจำเป็นต้องใช้เส้นโค้งที่มีความซับซ้อนมากกว่าเส้นตรง



รูปที่ 2.23 ตัวอย่างเส้นตัดสินใจที่มีความซับซ้อน ไม่เป็นเชิงเส้น

งานในการจำแนกข้อมูลจะเกี่ยวกับการวาดเส้นตรงที่ใช้ในการแบ่งวัตถุที่อยู่ใน กลุ่มต่างๆนี้เป็นที่รู้จักกันในชื่อ “Hyperplane Classifiers” ซึ่งวิธีการ SVM นั้นเหมาะสมอย่างยิ่งที่จะนำมาใช้ในงานจำแนกข้อมูลในลักษณะดังกล่าว



รูปที่ 2.24 การแปลงจุดข้อมูลของ SVM

ในรูปที่ 2.24 แสดงถึงแนวคิดของ SVM ซึ่งจะเห็นได้ว่าจุดเริ่มต้นของวัตถุ (ด้านซ้ายของภาพ) ถูกจับคู่โดยใช้เซตของฟังก์ชันทางคณิตศาสตร์ ที่เรียกว่า “Kernels” กระบวนการของการจัดเรียงวัตถุเรียกว่า “Mapping” หรือ “Transformation” วัตถุที่ถูกจับคู่หรือวัตถุที่อยู่ด้านขวาของภาพนั้น เป็นวัตถุที่สามารถแบ่งแยกได้ในเชิงเส้น ดังนั้นแทนการใช้เส้นโค้ง สิ่งที่ระบบทำคือ หาเส้นที่ดีที่สุดในการแบ่งวัตถุสามเหลี่ยมและสี่เหลี่ยม

SVM เป็นวิธีการจำแนกกลุ่มลำดับแรกๆ ที่ทำการจำแนกกลุ่มข้อมูลโดยสร้าง Hyperplanes ในพื้นที่หลายมิติ ซึ่งแบ่งกรณีของกลุ่มที่แตกต่างกันได้ นอกจากนั้นแล้ว SVM ยังสามารถทำได้ทั้งการถดถอย (Regression) และ การจำแนกกลุ่ม (Classification)

การสร้าง Hyperplane ที่เหมาะสมที่สุดนั้น SVM จะใช้กระบวนการวิธีแบบวนซ้ำ เพื่อใช้สำหรับลดค่าของฟังก์ชันความผิดพลาดให้ได้ค่าต่ำที่สุด และการจำแนกกลุ่มแบบ SVM สามารถแบ่งออกได้เป็น 2 ประเภทตามรูปแบบของฟังก์ชันความผิดพลาด ดังนี้

2.7.1 การจำแนกกลุ่ม SVM ประเภทที่ 1

สำหรับการจำแนกกลุ่มประเภทที่ 1 นี้เรียกว่า C-SVM เป็นการเรียนรู้เพื่อลดค่าฟังก์ชันความผิดพลาด ดังสมการที่ 2.38

$$\frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \quad (2.38)$$

$$y_i (w^T \varphi(x_i) + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

- โดยที่ C คือ ค่าความจุคงที่ (Capacity Constant)
 w คือ เวกเตอร์ของค่าสัมประสิทธิ์
 b คือ ค่าคงที่
 ξ คือ ค่าพารามิเตอร์สำหรับจัดการกับข้อมูลที่ไม่สามารถจำแนกได้
 i คือ จำนวนกรณีในการเรียนรู้ มีค่าตั้งแต่ 1 ถึง N
 y คือ สมาชิกของ $+1$ และ -1 ซึ่งบ่งบอกถึงกลุ่มของข้อมูล
 x_i คือ ตัวแปรอิสระ
 ϕ คือ ฟังก์ชัน Kernel สำหรับแปลงข้อมูลจากปริภูมิอินพุต (Input Space) ไปยังปริภูมิลักษณะเด่น (Feature Space)

จากสมการที่ 2.38 จะสังเกตได้ว่ายิ่งค่า C มากขึ้น ความผิดพลาดก็จะมีมากขึ้นตามไปด้วย ดังนั้นควรเลือกค่า C ให้เหมาะสมเพื่อหลีกเลี่ยงการเกิด Over Fitting

2.7.2 ฟังก์ชัน Kernel

ฟังก์ชัน Kernel ($K(x_i, x_j)$) ที่นิยมใช้ใน SVM มีอยู่ 4 ฟังก์ชัน ประกอบด้วย

$$1. \text{ Linear: } K(x_i, x_j) = x_i \cdot x_j \quad (2.39)$$

$$2. \text{ Polynomial: } K(x_i, x_j) = (\gamma x_i \cdot x_j + \text{coefficient})^{\text{degree}} \quad (2.40)$$

$$3. \text{ Radial Basis Function (RBF): } K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2.41)$$

$$4. \text{ Sigmoid: } K(x_i, x_j) = \tanh(\gamma x_i \cdot x_j + \text{coefficient}) \quad (2.42)$$

2.8 การวัดค่าความแยกตัวของกลุ่มข้อมูล (Class Separability)

ในการวิเคราะห์กลุ่มข้อมูลนั้น การวัดค่าความแยกตัวของกลุ่มข้อมูลสามารถบ่งบอกถึงข้อมูลเชิงลึกที่ช่วยให้เราสามารถเลือกใช้ตัวจำแนกกลุ่มสำหรับข้อมูลชุดนั้นๆ ได้อย่างเหมาะสม นอกจากนั้นแล้ว การวัดค่าดังกล่าวยังเป็นตัวแปรอีกตัวที่ช่วยบ่งชี้ถึงประสิทธิภาพของวิธีการลดมิติข้อมูล เพราะไม่เพียงแต่การลดจำนวนมิติลงของข้อมูลที่เป็นการทำงานหลักแล้ว แต่เมื่อชุดข้อมูลเดิมถูกแปลง หรือฉายให้ไปอยู่ในอีกปริภูมิซึ่งมีจำนวนมิติที่น้อยกว่าแล้วนั้น ข้อมูลใหม่ในแต่ละกลุ่มควรที่มีค่าความแยกตัวของกลุ่มเพิ่มมากขึ้นเพื่อช่วยให้ตัวจำแนกสามารถจำแนกกลุ่มได้อย่างแม่นยำมากยิ่งขึ้น หรือกล่าวได้ว่ากลุ่มข้อมูลใหม่ควรจะแยกห่างออกจากกันมากกว่าเดิมนั่นเอง ซึ่งวิธีการวัดค่าความแยกตัวนั้นมีอยู่หลายวิธี ดังต่อไปนี้

2.8.1 Fisher's Index (J)

เป็นการวัดค่าความแยกตัวของกลุ่มข้อมูลหลังจากผ่านการฉายไปยังปริภูมิใหม่ โดยทำการเปรียบเทียบอัตราส่วนระหว่างเทรซของเมทริกซ์การกระจายระหว่างกลุ่ม (\tilde{S}_B) กับดีเทอร์มิแนนต์ของเมทริกซ์การกระจายภายในกลุ่ม (\tilde{S}_W) ดังสมการที่ 2.47

$$J = \frac{\text{Trace}(\tilde{S}_B)}{\text{Trace}(\tilde{S}_W)} \quad (2.47)$$

โดยที่
$$\tilde{S}_W = \sum_{i=1}^N m_i \quad (2.48)$$

และ
$$\tilde{S}_B = \sum_{i=1}^N \sum_t r_i^t (m_i - m)(m_i - m)^T \quad (2.49)$$

เมื่อ m_i คือ ค่าเฉลี่ยของข้อมูลในกลุ่ม i ทั้งหมด
 m คือ ค่าเฉลี่ยของผลรวม m_i ทุกกลุ่ม

ซึ่งหากแต่ละกลุ่มข้อมูลมีลักษณะที่เกาะกันเป็นกลุ่มมากค่า $\text{Trace}(\tilde{S}_W)$ ที่ได้ก็จะมีค่าน้อยลง และหากข้อมูลระหว่างสองกลุ่มใดๆอยู่ห่างกันมากก็จะทำให้ค่า $\text{Trace}(\tilde{S}_B)$ ของสองกลุ่มนั้นมีค่าสูงขึ้นตาม จะสังเกตได้ว่าหากค่า J ของชุดข้อมูล A คำนวณออกมามีค่าสูงกว่าชุดข้อมูล B แล้วละก็ จะแสดงว่าชุดข้อมูล A นั้นมีการกระจายตัวที่เหมาะสมดีกว่าชุดข้อมูล B

2.8.2 Thornton's Separability Index (SI)

วิธีการวัดนี้คือการวัดค่าองศาของความเกี่ยวข้องกันระหว่างข้อมูลขาเข้ากับข้อมูลขาออกที่อยู่ในกลุ่มเดียวกัน ซึ่งใน [13] ได้แสดงให้เห็นว่าวิธีการวัดค่า SI เป็นวิธีการที่มีประสิทธิภาพอย่างยิ่งสำหรับวัดค่าความแยกตัวของกลุ่มข้อมูล โดยผลลัพธ์ของ SI ที่คำนวณได้จะมีค่าอยู่ระหว่างช่วง 0 ถึง 1 ถ้ากรณีของแต่ละข้อมูลแต่ละกลุ่มนั้นสามารถแยกออกจากกันได้เป็นอย่างดี ค่า SI จะมีค่าเข้าใกล้ 1 และจะมีค่าเข้าใกล้ 0 เมื่อแต่ละกลุ่มนั้นขยับเข้ามาใกล้กันมากขึ้น ในการคำนวณค่า SI สามารถทำได้ดังในสมการที่ 2.50

$$SI = \frac{\sum_{i=1}^n (f(x_i) + f(x'_i) + 1) \bmod 2}{n} \quad (2.50)$$

เมื่อกำหนดให้ f เป็นฟังก์ชันเป้าหมาย

x'_i เป็นจุดใกล้เคียงมากที่สุดของ x_i

n เป็นจำนวนของจุดทั้งหมด