บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ได้ทบทวนทฤษฎี และงานวิจัยที่เกี่ยวข้อง เพื่อสร้างแบบจำลองที่สามารถทำนายระยะเวลาที่ใช้พิมพ์ตัวอักขระในคู่อักขระ โดยเน้นทบทวนองค์ความรู้ด้านการทำงานของสมอง การวิเคราะห์ถดถอย (Regression Analysis) ทั้งในส่วนของการวิเคราะห์ถดถอยแบบพื้นฐาน และการวิเคราะห์การถดถอยแบบหลายตัวแปร โครงข่ายประสาทเทียม (Artificial Neural Network) แบบจำลองต้นไม้เอ็มไฟว์พี (Model Tree: M5P) การแบ่งกลุ่มข้อมูลด้วยค่าเฉลี่ยเคกลุ่ม (K-Means Clustering) ข้อมูลอนุกรมเวลา (Time Series Data) และวิธีการทดสอบแบบไขว้ข้ามสิบกลุ่ม (10 Fold Cross-Validation)

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การทำงานของสมองและทักษะในการทำงาน

สมองมีความสามารถในการทำงานได้หลายประเภท ไม่ว่าจะเป็นงานที่เกี่ยวข้องกับการคำนวณ หรืองานที่มีความซับซ้อนเช่น การขับรถ หรือการพิมพ์ ซึ่งงานเหล่านี้จะใช้ทักษะการทำงานหลายส่วนประกอบกัน เพื่อให้ทำงานสำเร็จตามเป้าหมายที่ต้องการทั้งในส่วนกระบวนการคิด และส่วนควบคุมการเคลื่อนไหวของอวัยวะร่างกาย ซึ่งจะทำงานตอบสนองต่อข้อมูลที่ได้รับเข้ามาจากสภาพแวดล้อมภายนอก โดยที่การทำงานของสมองแบ่งออกเป็น 3 ส่วนคือ สมองส่วนหน้าทำหน้าที่เก็บข้อมูล สมองส่วนกลางทำหน้าที่เป็นทางผ่านของข้อมูลที่ผ่านการประมวลผลสู่ส่วนควบคุมการเคลื่อนไหว และสมองส่วนท้ายทำหน้าที่ควบคุมการทำงานของร่างกาย (อุบลรัตน์, 2535) ในขณะที่ทักษะเป็นสิ่งที่เกิดขึ้นจากการเรียนรู้ข้อมูล และลำดับขั้นตอนการทำงานอย่างใดอย่างหนึ่งหลาย ๆ ครั้งจนสามารถจดจำกระบวนการทำงานทั้งหมดของสิ่งที่เรียนรู้ได้ ซึ่งทักษะจะส่งผลให้ประสิทธิภาพในการทำงานของมนุษย์คือ ยิ่งมีทักษะในการทำงานมากเท่าใดประสิทธิภาพในการทำงานสูงมากขึ้นด้วย ทั้งในเรื่องความเร็ว และความถูกต้องแม่นยำในการทำงาน โดยที่ทักษะเกิดจากประสบการณ์ในการทำงานของผู้ใช้แต่ทักษะก็สามารถที่จะลบเลือน

ตามประสบการณ์เช่นกันคือ เมื่อไม่ได้ฝึกเป็นระยะเวลาหนึ่งทักษะก็จะลบเลือน แต่จะไม่หายไปถ้าฝึกเพิ่มเติมก็จะสามารถใช้งานทักษะดังกล่าวได้อย่างเต็มประสิทธิภาพเช่นเดิม

จากการทำงานของสมองที่กล่าวมาข้างต้นนี้ งานวิจัยนี้ได้นำแนวทางดังกล่าวมาประยุกต์ใช้ในการสร้างแบบจำลองการทำงานของสมองที่ใช้พิมพ์ตามรูปแบบการทำงานที่กล่าวมาในข้างต้น ซึ่งแบบจำลองสามารถอธิบายพฤติกรรมการพิมพ์ของผู้ใช้ และสามารถทำนายแนวโน้มของการพัฒนาทักษะการพิมพ์ได้

2.1.2 **โครงข่ายประสาทเทียม** (Artificial Neural Network: ANN)

โครงข่ายประสาทเทียมเป็นการศึกษาระบบการทำงานของเซลล์ประสาทภายในสมองที่ประกอบด้วยเซลล์ประสาท (Neuron) และเส้นประสาท (Nerve) โดยที่เซลล์ประสาทจะเชื่อมต่อกันในรูปแบบโครงข่าย ซึ่งการวิเคราะห์ และประมวลผลข้อมูลของระบบประสาทนั้นจะส่งข้อมูลผ่านระบบโครงข่ายของเซลล์ประสาท และทำงานในลักษณะขนานคือ ทำกิจกรรม หรืองานหลายอย่างในเวลาเดียวกันให้ได้มาซึ่งผลลัพธ์ที่ต้องการ โดยการทำงานของสมองในรูปแบบที่กล่าวมาในข้างต้นนั้นมีความสามารถหลายประการเช่น การสังเกต เรียนรู้ จดจำ ทำซ้ำ และแยกแยะสิ่งต่าง ๆ ซึ่งโครงข่ายประสาทเทียมได้จำลองรูปแบบการทำงาน และโครงสร้างการเชื่อมต่อดังกล่าวมา เพื่อทำให้ระบบคอมพิวเตอร์สามารถทำงานที่สมองทำได้ ซึ่งสามารถเพิ่มประสิทธิภาพในการทำงานให้มากขึ้น และช่วยลดข้อผิดพลาดที่เกิดขึ้นจากการทำงาน

นอกจากนี้โครงข่ายประสาทเทียมได้จำลอง รูปแบบการเคลื่อนที่ของข้อมูล โดยที่รูปแบบการเคลื่อนที่ของข้อมูลภายในสมองนั้นใช้สัญญาณไฟฟ้าเป็นสิ่งกระตุ้นการเคลื่อนที่ของข้อมูล เมื่อสัญญาณไฟฟ้าถึงระดับหนึ่งข้อมูลที่อยู่ในเซลล์ประสาทจะถูกส่งผ่านไปยังอีกเซลล์ประสาทหนึ่ง ในการทำงานลักษณะเดียวกันโครงข่ายประสาทเทียมใช้ค่าน้ำหนัก (Weight) แทนค่าของสัญญาณไฟฟ้า และโหนด (Node) แทนเซลล์ประสาท ซึ่งค่าน้ำหนักนั้นได้มาจากการคำนวณด้วยฟังก์ชันกระตุ้นที่อยู่ในแต่ละโหนดของโครงข่าย ซึ่งค่าน้ำหนักเป็นค่าที่กำหนดการเคลื่อนที่ของข้อมูลผ่านโหนด และสามารถปรับเปลี่ยนไปตามข้อมูลที่เข้ามาทางอินพุต เพื่อให้ประสิทธิภาพในการวิเคราะห์ข้อมูลของโครงข่ายประสาทเทียมใกล้เคียง และถูกต้องมากที่สุด (Russell and Norvig, 1995) โครงข่ายงานประสาทเทียมถูกพัฒนาขึ้นประยุกต์ใช้งานด้านการจำแนกรูปแบบ (Pattern recognition) และการจัดแบ่งประเภท (Classification) สำหรับงานวิจัยนี้ใช้โครงข่ายประสาทเทียมแบบหลายชั้น (Artificial Neural Network Multi-Layer Perceptron:

MLP) และวิธีการแพร่กระจายย้อนกลับ (Back-propagation Algorithm) เป็นอีกหนึ่งเครื่องมือที่ใช้สร้างแบบจำลองการพิมพ์

2.1.2.1 **ฟังก์ชันกระตุ้น** (Activate function)

การวิเคราะห์ข้อมูลด้วยโครงข่ายงานประสาทเทียมนั้นจะคำนวณข้อมูลที่เข้ามาทางโหนดอินพุตด้วย ค่าคุณลักษณะ (Attribute) และค่าน้ำหนักของข้อมูลผ่านสมการฟังก์ชันกระตุ้นที่จะคำนวณหาค่าการกระตุ้น ซึ่งเป็นค่าที่ใช้กำหนดการเคลื่อนของข้อมูลผ่านโหนดในชั้นต่าง ๆ ของโครงข่ายประสาทเทียมจากโหนดในชั้นหนึ่งไปสู่อีกโหนดอีกชั้นหนึ่ง โดยที่การคำนวณค่าการกระตุ้นด้วยฟังก์ชันกระตุ้นนั้นจะคำนวณผลรวมของค่าฟังก์ชันอินพุตทั้งหมดที่เข้ามาสู่โหนดในชั้นนั้น ๆ โดยค่าการกระตุ้นในแต่ละโหนดนั้นจะไม่ลดลงตามฟังก์ชันของโหนดอินพุตทั้งหมด ซึ่งโดยทั่วไปแล้วค่าการกระตุ้นจะใช้ในการจำกัดการผ่านของข้อมูล ซึ่งฟังก์ชันกระตุ้นมีอยู่หลายลักษณะตามความต้องการของงานที่แตกต่างกัน ในบางกรณีค่าข้อมูลที่ออกมาทางเอาต์พุตของโหนดจะเกิดจากการสุ่มของอินพุตทั้งหมดที่เข้ามาสู่โหนด โดยสมการฟังก์ชันอินพุตที่ใช้ในการคำนวณหาค่าการกระตุ้นแสดงให้เห็นดังต่อไปนี้

$$in_i = \sum_{j=0}^{n} W_{j,i} a_j \tag{2.1}$$

ดังนั้นจึงได้สมการ ฟังก์ชันกระตุ้น g ซึ่งเป็นผลรวมจากฟังก์ชันอินพุตเพื่อให้ได้ค่าเอาต์พุตของโหนด

$$a_i = g(in_i) = g\left[\sum_{j=0}^{n} W_{j,i} a_j \right] \tag{2.2}$$

กำหนดให้

g คือ ฟังก์ชันกระตุ้นของโหนด i

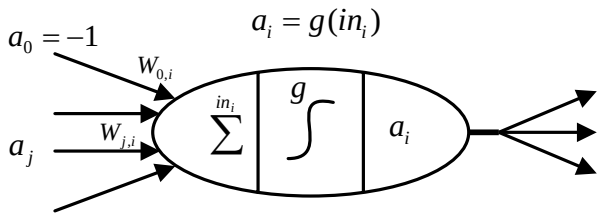
in_i คือ ฟังก์ชันผลรวมของอินพุตของโหนด i

a_i คือ สัญญาณอินพุตของโหนด i

$W_{j,i}$ คือ ค่าน้ำหนักของเส้นเชื่อมระหว่างโหนด i และ j

ภาพที่ 2.1

ฟังก์ชันการคำนวณในโหนดของโครงข่ายประสาทเทียม



ที่มา: “Artificial Intelligence a Modern Approach,” by Russell and Norvig, 1995

จากภาพที่ 2.1 สังเกตได้ว่ากำหนดค่าให้ $a_0 = -1$ และค่าน้ำหนักคือ $W_{0,i}$ เป็นค่าที่ป้อนเข้ามาคำนวณในสมการฟังก์ชันอินพุตของโหนด หลังจากนั้นผลรวมของข้อมูลอินพุตจะถูกคำนวณด้วยสมการฟังก์ชันกระตุ้น ซึ่งจะคัดกรองข้อมูลที่จะผ่านโหนดด้วยการคำนวณจากฟังก์ชัน ดังกล่าวคือ ข้อมูลจะผ่านโหนดก็ต่อเมื่อค่าการกระตุ้นเข้าใกล้ +1 และจะไม่ยอมให้ข้อมูลผ่านเมื่อค่าการกระตุ้นเข้าใกล้ 0 ซึ่งข้อมูลดังกล่าวเป็นข้อมูลที่ไม่ตรงตามความต้องการ โดยที่ข้อมูลจะคำนวณผ่านสมการฟังก์ชันกระตุ้นตามจำนวนชั้นของโครงข่ายประสาทเทียมจนกระทั่งได้ค่าผลลัพธ์ออกมาทางชั้นเอาต์พุต โดยที่การกระตุ้นไม่จำเป็นต้องอยู่ในลักษณะเส้นตรงที่มีทางเลือก 2 ทาง เพื่อใช้ในการแยกประเภทของข้อมูลอินพุต นอกจากนี้การเรียนรู้ของโครงข่ายประสาทเทียมนั้นคือ การคำนวณค่าน้ำหนักที่ได้จากค่าข้อมูลอินพุต และค่าน้ำหนักจะถูกปรับเปลี่ยนให้ได้มาซึ่งผลลัพธ์ที่เหมาะสมที่สุด

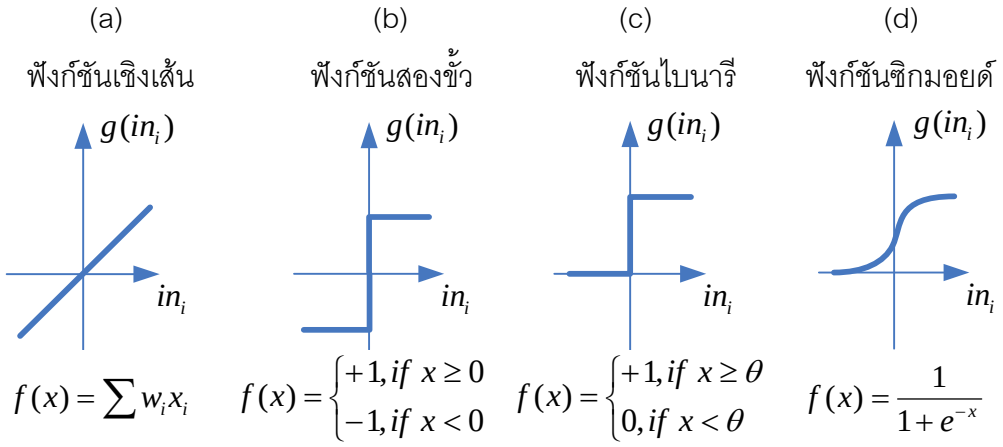
ค่าน้ำหนัก

ฟังก์ชันอินพุต

สมการของฟังก์ชันกระตุ้นมีความหลากหลาย และเหมาะสมกับงานที่หลากหลายทำให้โครงข่ายประสาทเทียมสามารถทำงานที่มีความซับซ้อนได้ดี โดยในแต่ละฟังก์ชันจะคำนวณหาค่าการกระตุ้นยกตัวอย่างเช่น แบบเชิงเส้น แบบสองขั้ว แบบไบนารี และแบบซิกมอยด์

ภาพที่ 2.2

ฟังก์ชันกระตุ้นแบบต่าง ๆ ของโครงข่ายประสาทเทียม



ที่มา:

2.1.2.2 โครงสร้างของโครงข่ายประสาทเทียม

โครงสร้างที่แสดงให้เห็นถึงรูปแบบการเชื่อมต่อกันระหว่างโหนดที่อยู่ภายในโครงข่ายประสาทเทียมมีอยู่ 2 แบบด้วยกัน ซึ่งโครงข่ายทั้ง 2 รูปแบบจะมีความสามารถในการวิเคราะห์ข้อมูลที่แตกต่างกัน และมีความเหมาะสมกับงานแต่ละประเภท

- **เน็ตเวิร์กแบบป้อนไปข้างหน้า** (Feed-Forward Networks)

+1

โครงข่ายประสาทเทียมที่มีโครงสร้าง และทิศทางการเคลื่อนของข้อมูลไปทางด้านหน้าเพียงอย่างเดียว ข้อมูลจะเข้าทางชั้นอินพุตผ่านชั้นซ่อน และออกไปทางชั้นเอาต์พุต การเชื่อมต่อกันของโหนดภายในโครงสร้างลักษณะนี้จะเป็นการเชื่อมต่อระหว่างโหนดในชั้นปัจจุบันกับโหนดชั้นถัดไป ซึ่งการเชื่อมต่อของโครงข่ายในลักษณะดังต่อไปนี้จะไม่มีอยู่คือ การย้อนกลับการวนรอบโหนดในชั้นเดียวกัน และการข้ามโหนดในชั้นใดชั้นหนึ่งของโครงข่าย ซึ่งในการคำนวณหาผลลัพธ์ที่ต้องการนั้นจะคำนวณแบบง่าย โดยฟังก์ชันที่ใช้คำนวณค่าจะกำหนดค่าน้ำหนักที่ได้จากค่าข้อมูลอินพุตทั้งหมดไม่มีฟังก์ชันกระตุ้นอื่นเพิ่มเข้ามาคำนวณ ซึ่งค่าเอาต์พุตที่ออกมาจากโหนดในชั้นใดก็ตามในโครงข่ายจะไม่ส่งผลต่อค่าเอาต์พุตที่ออกมาจากโหนดที่อยู่ชั้นเดียวกัน

+1

- **เน็ตเวิร์กแบบรีเคอร์เรนต์** (Recurrent Networks)

โครงข่ายประสาทเทียมที่มีโครงสร้าง ทิศทางการเคลื่อนที่ของข้อมูล และการคำนวณค่าของฟังก์ชันกระตุ้นที่แตกต่างไปจากโครงข่ายแบบทิศทางการป้อนไปข้างหน้า โครงข่าย

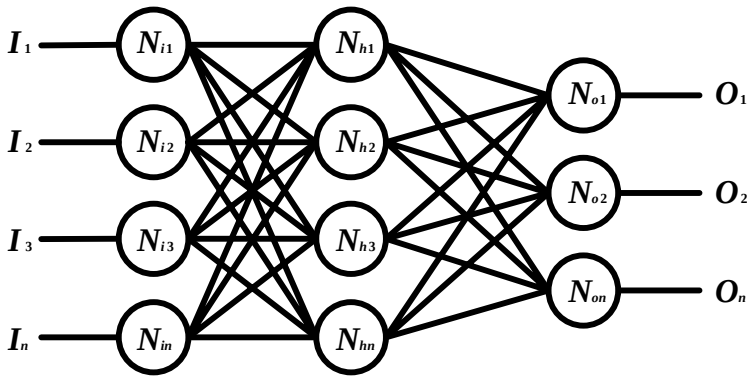
แบบรีเคอร์เรนต์นั้นจะมีการเชื่อมต่อที่มีความหลากหลายไม่ว่าจะเป็นการเชื่อมต่อจากชั้นเอาต์พุต หรือชั้นซ่อนป้อนกลับข้อมูลมายังโหนดในชั้นอินพุต ที่จะคำนวณ และปรับค่าการกระตุ้น ในการคำนวณค่ากระตุ้นจะคำนวณผ่านทางฟังก์ชันกระตุ้นข้อมูลที่ใช้ได้จากฟังก์ชันอินพุตที่ประกอบไปด้วย ค่าน้ำหนัก ค่าข้อมูลอินพุต และค่าสถานะภายในที่ได้จากการจัดเก็บระดับการกระตุ้นของโหนดในชั้นที่ป้อนกลับเข้ามาคำนวณต่อในชั้นนั้น ๆ การทำงานรูปแบบดังกล่าวจะทำให้ค่าการกระตุ้นมีความเปลี่ยนแปลงไปตามระดับของค่าน้ำหนักที่เปลี่ยนไป และในการปรับค่าน้ำหนักนั้นจะปรับจนกระทั่งค่าน้ำหนักอยู่ในสถานะคงที่ โดยที่คุณสมบัติประการหนึ่งของโครงข่ายลักษณะนี้คือ สามารถจำลองการทำงานหน่วยความจำระยะสั้นได้

2.1.2.3 โครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer Perceptron)

โครงข่ายประสาทเทียมแบบหลายชั้น ประกอบด้วยชั้นอินพุต (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นเอาต์พุต (Output Layer) โดยในแต่ละโหนดของแต่ละชั้นจะเชื่อมต่อกันทั้งหมด โดยการเรียนรู้ของโครงข่ายนั้นจะเรียนรู้จากกลุ่มข้อมูลตัวอย่าง เพื่อสอนให้โครงข่ายเข้าใจและสามารถแก้ปัญหาที่เกิดขึ้นจากกลุ่มข้อมูลตัวอย่างที่ใช้เรียนรู้ ในโครงสร้างของโครงข่ายประสาทเทียมจะประกอบด้วยโหนดที่ใช้ในการประมวลผลกระจายอยู่ในโครงสร้างแต่ละชั้น การประมวลผลของโครงข่ายประสาทเทียมจะอาศัยการส่งข้อมูลการทำงานผ่านโหนดในชั้นต่าง ๆ

ภาพที่ 2.3

โครงข่ายประสาทเทียมแบบหลายชั้น



ที่มา: “Artificial Intelligence a Modern Approach,” by Russell and Norvig, 1995

ภาพที่ 2.3 แสดงถึงกระบวนการเรียนรู้ของโครงข่ายประสาทเทียมแบบหลายชั้น ซึ่งรูปแบบในการเรียนรู้จะมีส่งผ่านข้อมูล 2 รูปแบบคือ การส่งผ่านไปข้างหน้า (Forward Pass) และการส่งผ่านย้อนกลับ (Backward Pass) ข้อมูลจะถูกส่งผ่านจากชั้นอินพุตไปยังชั้นต่อไปจนถึงชั้นเอาต์พุต และความผิดพลาดจะถูกส่งย้อนกลับ เพื่อปรับค่าผลลัพธ์ที่ได้ ทั้งนี้ โครงข่ายประสาทแบบหลายชั้นสามารถเรียนรู้การจำแนกข้อมูลที่มีความซับซ้อน และไม่เป็นเชิงเส้นได้

2.1.2.4 **วิธีการแพร่กระจายย้อนกลับ** (Back-propagation Algorithm)

วิธีการแพร่กระจายย้อนกลับเป็นอัลกอริทึมที่ใช้ในการเพิ่มประสิทธิภาพการเรียนรู้ของโครงข่ายประสาทวิธีหนึ่งที่นิยมใช้ เป็นกระบวนการปรับค่าน้ำหนักของโครงข่ายประสาทเทียมที่เกิดค่าความผิดพลาดจากการคำนวณของโครงข่าย ซึ่งค่าความผิดพลาดนี้ได้มาจากความแตกต่างระหว่างเอาต์พุตของโครงข่ายกับค่าเป้าหมาย โดยที่โครงข่ายประสาทเทียมแบบแพร่กระจายย้อนกลับจะปรับค่าน้ำหนัก เพื่อลดค่าความผิดพลาดที่เกิดขึ้นจากความแตกต่างดังกล่าวให้มีค่าน้อยที่สุด เพื่อให้โครงข่ายสามารถทำนาย หรือคำนวณค่าของข้อมูลให้ได้ใกล้เคียงกับค่าเป้าหมาย ซึ่งโครงข่ายจะเรียนรู้ไปข้างหน้า และส่งความผิดพลาดย้อนกลับ เพื่อปรับค่าน้ำหนักกลับไปยังโหนดในชั้นข้างหน้าคือ คำนวณค่าความผิดพลาดจากชั้นเอาต์พุต กลับมาจนกระทั่งถึงชั้นอินพุต การเรียนรู้วิธีแพร่กระจายย้อนกลับจากการทำนายของโครงข่ายสำหรับแต่ละค่าตัวอย่างด้วยค่าที่แท้จริง แต่สำหรับค่าตัวอย่างการฝึกฝนนั้น น้ำหนักการเชื่อมต่อจะถูกปรับเปลี่ยนนี้เป็นการส่งค่าย้อนกลับจากชั้นเอาต์พุต ชั้นซ่อน และกลับจนกระทั่งถึงโหนดในชั้นอินพุต

2.1.3 **การวิเคราะห์การถดถอย** (Regression Analysis)

ในการหาความสัมพันธ์ของตัวแปรหนึ่งกับตัวแปรอื่น ๆ มักจะใช้เครื่องมือทางสถิติที่เรียกว่า การวิเคราะห์การถดถอย (Regression Analysis) ซึ่งเป็นกระบวนการทั้งหมดที่จะสร้างความสมการความสัมพันธ์ระหว่าง 2 ตัวแปรขึ้นไป โดยที่ตัวแปรหนึ่งเป็นผลมาจากอีกตัวแปรหนึ่งหรือหลายตัวแปรก็ได้ โดยอาศัยข้อมูลจากการสุ่มตัวอย่างและการอนุมานอื่น ๆ เช่น ความสัมพันธ์ระหว่างยอดขายสินค้า (Y) กับค่าใช้จ่ายในการโฆษณา (X) หรือความสัมพันธ์ระหว่างเงินออม (Y) กับอัตราดอกเบี้ย (X) เป็นต้น

นั่นคือ $Y = f(X)$ ในกรณีที่มีเพียงตัวแปร 2 ตัว

หรือ $Y = f(X_1, X_2, \mathrm{K}, X_n)$ ในกรณีที่มีตัวแปรอินพุต n ตัว

ในการวิเคราะห์การถดถอยสามารถแบ่งความสัมพันธ์ได้เป็น 2 ลักษณะ คือ ความสัมพันธ์ในเชิงเส้นตรง (Linear Relationship) และความสัมพันธ์ในรูปแบบที่ไม่ใช่เส้นตรง

(Non-linear Relationship) ความสัมพันธ์เชิงเส้นตรงยังแบ่งออกได้เป็นความสัมพันธ์อย่างง่าย (Simple Relationship) และความสัมพันธ์เชิงซ้อน (Multiple Relationships)

2.1.3.1 การวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis)

เป็นการศึกษาเพื่อหาสมการซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรตาม (Y) 1 ตัวแปร กับตัวแปรอิสระ (X) 1 ตัวแปร โดยรูปแบบการถดถอยอย่างง่ายเชิงเส้นตรง (Linear Regression Model) สามารถเขียนในรูปสมการได้ดังนี้

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (2.3)$$

กำหนดให้

X_i คือ ค่าสังเกตตัวที่ i ของตัวแปรอิสระ

Y_i คือ ค่าสังเกตตัวที่ i ของตัวแปรตาม

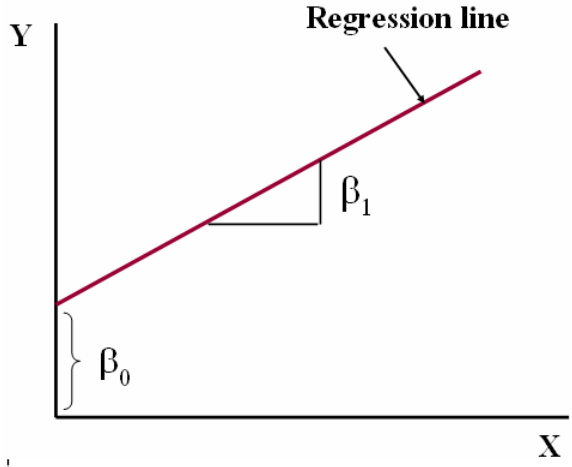
β_0 คือ Y-intercept หรือจุดที่เส้นถดถอยตัดแกน Y

β_1 คือ สัมประสิทธิ์การถดถอย (Regression coefficient) เป็นค่าที่บอกถึงความเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนไป 1 หน่วย

ε_i คือ ความคลาดเคลื่อนสุ่ม (Random error)

ภาพที่ 2.4

เส้นสมการถดถอยเชิงเส้น



ที่มา: “เศรษฐสถิติ,” โดยเรวัตร ธรรมาอภิรมย์, 2543

สำหรับการประมาณการโดยการใช้วิธีกำลังสองน้อยที่สุด หรือวิธีลีสท์สแควร์ (Least squares method) จะสามารถประมาณค่าพารามิเตอร์ β_0 และ β_1 โดยใช้สมการปกติ (Normal equation) ดังนี้

$$nb_0 + b_1 \sum X_i = \sum Y_i \tag{2.4}$$

$$b_0 \sum X_i + b_1 \sum X_i^2 = \sum X_i Y_i \tag{2.5}$$

ทั้งนี้ จากการแก้สมการ จะสามารถคำนวณหา b_0 และ b_1 ที่ใช้เป็นตัวประมาณค่าพารามิเตอร์ β_0 และ β_1 ได้หรืออาจสามารถสร้างเป็นสูตรในการคำนวณได้ดังนี้

$$b_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} \tag{2.6}$$

$$b_0 = \overline{Y} - b_1 \overline{X} \tag{2.7}$$

โดย b_0 หรือ Y-intercept เป็นค่าที่บอกให้ทราบว่าเมื่อ X เป็น 0 แล้ว Y จะมีค่าเป็นเท่าใด ถ้า b_0 เป็นบวก แสดงว่าเส้นถดถอยตัดแกน Y เหนือจุดเริ่มต้นแต่ถ้า b_0 เป็นลบแสดงว่าเส้นถดถอยตัดแกน Y ต่ำกว่าจุดเริ่มต้น

b_1 หรือ สัมประสิทธิ์การถดถอย (Regression Coefficient) เป็นค่าที่บอกให้ทราบว่าเมื่อ X เปลี่ยนไป 1 หน่วย จะทำให้ Y เปลี่ยนไปกี่หน่วย ถ้า b_1 เป็นบวก แสดงว่า เมื่อ X เพิ่มขึ้นจะทำให้ Y เพิ่มขึ้น(X และ Y มีความสัมพันธ์ในทางตามกัน) แต่ถ้า b_1 เป็นลบแสดงว่า เมื่อ X เพิ่มขึ้นจะทำให้ Y ลดลง (X และ Y มีความสัมพันธ์ในทางตรงข้าม)

โดยทั่วไปหากต้องการทราบว่าสมการถดถอยที่ประมาณขึ้นมามีความใกล้เคียงกับเส้นถดถอยที่แท้จริงในกลุ่มของข้อมูลมากน้อยพียงใด ในการวัดความใกล้เคียงกับเส้นถดถอยที่แท้จริงส่วนมากจะใช้สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination) หรือ R^2 โดยสามารถคำนวณได้จากสูตรดังต่อไปนี้

$$R^2 = \frac{b_1\left[\sum XY - \left(\sum X\right)\left(\sum Y\right)/n\right]}{\sum Y^2 - \left(\sum Y\right)^2/n} \tag{2.8}$$

กำหนดให้ $0 \leq R^2 \leq 1$

ถ้า $R^2 = 1$ แสดงว่า ความผันแปรทั้งหมดใน Y เป็นผลมาจากความเปลี่ยนแปลงของ X เพียงอย่างเดียว

ถ้า $R^2 = 0$ แสดงว่า ความผันแปรทั้งหมดใน Y ไม่ได้เป็นผลมาจากการเปลี่ยนแปลงของ X แต่มาจากสาเหตุหรือปัจจัยอื่น

ถ้า R^2 มีค่าเข้าใกล้ 1 หรือ 100% จะทำให้การใช้ ค่าของ X ไปช่วยคาดคะเน หรือทำนายค่าของ Y มีความแม่นยำมากขึ้น

นอกจากนี้สามารถใช้การทดสอบเอฟ (F-test - Analysis of Variance) ในการทดสอบว่าตัวแบบสามารถนำไปใช้ได้หรือไม่ ส่วนการใช้การวิเคราะห์หาค่าความแปรปรวน (Analysis of Variance: ANOVA) สามารถนำไปทดสอบว่ามีความสัมพันธ์ทางสถิติระหว่างตัวแปรตาม กับตัวแปรอิสระ 1ตัว หรือมากกว่า 1 ตัวได้

2.1.3.2 **การวิเคราะห์การถดถอยเชิงซ้อน** (Multiple Regression Analysis)

เป็นการศึกษาความสัมพันธ์อิทธิพลของตัวแปรอิสระตั้งแต่ 2 ตัวแปรขึ้นไปที่มีผลต่อตัวแปรตาม โดยมีแบบจำลองดังนี้

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon_i \qquad (2.9)$$

ในการคำนวณเส้นถดถอยเชิงซ้อนในกรณีที่มีตัวแปรอิสระมากกว่า 2 ตัวนั้นจะทำได้ยาก เนื่องจากความซับซ้อนของสมการ แต่เราสามารถใช้เครื่องคอมพิวเตอร์ในการหาค่าตัวแปรต่าง ๆ ได้ เช่น การประมาณเส้นถดถอยโดยใช้สมการปกติ (Normal Equations) การประมาณเส้นถดถอยโดยย้ายจุดกำเนิด และการประมาณเส้นถดถอยโดยใช้เมตริกซ์ (Matrix) (เรวัตร ธรรมาอภิรมย์, 2543)

2.1.4 **แบบจำลองต้นไม้เอ็มไฟว์พี** (Model Tree: M5P)

แบบจำลองต้นไม้เอ็มไฟว์พีเป็นต้นไม้ที่ใช้ทำนายผลข้อมูลที่เป็นตัวเลข ซึ่งพัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) แต่จะมีความแตกต่างกัน ในขั้นตอนการเลือกโหนดในแต่ละชั้นของต้นไม้ และค่าคุณลักษณะเอาต์พุตเป็นค่าตัวเลข แต่ค่าคุณลักษณะอินพุตนั้นจะเป็นไปได้ทั้งค่าต่อเนื่อง และไม่ต่อเนื่อง นอกจากนี้ที่โหนดใบ (Leaf Node) ของต้นไม้จะมีแบบจำลองเชิงเส้นที่ใช้ทำนายกลุ่มของข้อมูลที่เป็นค่าตัวเลข ในการทำนายค่าของกลุ่มข้อมูลจะวิเคราะห์ข้อมูลจากโหนด

ราก (Root Node) ลงมาจนกระทั่งถึงโหนดใบที่จะได้ค่าผลลัพธ์ ซึ่งจะใช้ค่าคุณลักษณะในการตัดสินใจเลือกเส้นทางจากโหนดสู่โหนดแต่ละชั้น

2.1.4.1 **การจัดกลุ่มค่าคุณลักษณะ** (Nominal Attribute)

ขั้นตอนการจัดกลุ่มข้อมูลจากค่าคุณลักษณะนั้นทำก่อนที่จะนำกลุ่มข้อมูลมาสร้างแบบจำลองต้นไม้เอ็มไฟว์พี โดยที่ค่าคุณลักษณะทั้งหมดจะถูกจัดกลุ่มเป็นคู่ตามค่าเฉลี่ยของกลุ่มข้อมูล และค่าคุณลักษณะที่ใกล้เคียงกัน ซึ่งค่าดังกล่าวเป็นค่าตัวเลขที่ได้มาจากการคำนวณหาค่าการเรียนรู้ และเรียงลำดับตามค่าเฉลี่ยของกลุ่มข้อมูล โดยเรียงลำดับจากค่าเฉลี่ยกลุ่มข้อมูลด้วยค่าคุณลักษณะนี้เป็นขั้นตอนเรียงลำดับของข้อมูลที่ถูกเลือกในแต่ละโหนด อย่างไรก็ตามค่ารบกวน (Noise) ที่เพิ่มขึ้นเมื่อใดก็ตามที่จำนวนของข้อมูลมีน้อย หรือค่าคุณลักษณะที่ไม่ครอบคลุมกลุ่มข้อมูล เป็นต้น

2.1.4.2 **การสร้างต้นไม้** (Building Tree)

การสร้างแบบจำลองต้นไม้เอ็มไฟว์พีนั้นจะสร้างคล้ายคลึงกับต้นไม้ตัดสินใจ แต่จะแตกต่างกันคือ แทนที่จะเลือกโหนดจากค่าสูงสุด แต่ใช้การแบ่งที่จะช่วยลดการผันแปรของกลุ่มข้อมูลในแต่ละกิ่ง (Branch) ของต้นไม้ให้มากที่สุด ต่อมาพิจารณาตัดเล็มต้นไม้จากแต่ละโหนดใบ การตัดเล็มนั้นจะแตกต่างจากต้นไม้ตัดสินใจคือ เมื่อตัดเล็มต้นไม้ในโหนดที่อยู่ชั้นบนจะต้องพิจารณาแทนที่โหนดที่ตัดออกด้วยระนาบถดถอยแทนที่จะเป็นค่าคงที่ และค่าคุณลักษณะจะใช้ในการตัดสินใจขยายโหนดออกจากโหนดนั้น ๆ การแบ่งโหนดในรูปแบบดังกล่าวช่วยแก้ไขปัญหาความเบี่ยงเบนของกลุ่มข้อมูล ซึ่งจะวัดค่าความผิดพลาดที่เกิดขึ้นในแต่ละโหนด และคำนวณหาการลดลงของค่าความผิดพลาดที่เกิดขึ้นมาด้วยค่าคุณลักษณะในแต่ละโหนด โดยที่ค่าคุณลักษณะใดทำให้ค่าความผิดพลาดลดลงมากที่สุดจะถูกนำมาใช้ในโหนดนั้น ซึ่งสามารถคำนวณโดยใช้สมการค่าเบี่ยงเบน (Standard deviation reduction: SDR) ดังต่อไปนี้

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \qquad (2.10)$$

โดยกำหนดให้ T_1, T_2 ... คือ กลุ่มข้อมูลได้จากการแบ่งโหนดตามค่าคุณลักษณะ

กระบวนการแบ่งโหนดของต้นไม้นั้นจะสิ้นสุดก็ต่อเมื่อค่าของกลุ่มข้อมูลนั้นคงที่คือจนกระทั่งแบ่งไปจนถึงโหนดที่มีความเปลี่ยนแปลงน้อยมาก ๆ ดังนั้นค่าเบี่ยงเบนเป็นเพียงตัวแปรตัวหนึ่งในค่าเปลี่ยนแปลงเริ่มต้นของกลุ่มข้อมูล

2.1.4.3 **การตัดเล็มต้นไม้** (Pruning Tree)

แบบจำลองเชิงเส้นนั้นจะอยู่ในแต่ละโหนดของต้นไม้ไม่เพียงจะใช้ในขั้นตอนการปรับให้เรียบ (Smoothing Process) ที่จะทำก่อนตัดเล็มต้นไม้โดยที่แบบจำลองจะถูกคำนวณในแต่ละโหนดของต้นไม้ที่จะถูกตัดตามสมการที่แสดงให้เห็นข้างล่างนี้

$$w_0 + w_1 a_1 + w_2 a_2 + \mathrm{K} + w_k a_k \qquad (2.11)$$

กำหนดให้

$a_1, a_2, \mathrm{K}, a_k$ คือ ค่าคุณลักษณะ

$w_1, w_2, \mathrm{K}, w_k$ คือ ค่าน้ำหนัก

อย่างไรก็ตามค่าคุณลักษณะที่ใช้ทดสอบส่วยย่อยของต้นไม้จะถูกใช้ในสมการถดถอย เพราะว่าค่าคุณลักษณะดังกล่าวส่งผลต่อการทำนายค่าข้อมูล ซึ่งจะขึ้นอยู่กับจำนวนของกลุ่มข้อมูลที่ใช้ทดสอบในแต่ละโหนด ในขั้นตอนการตัดเล็มต้นไม้จะประมาณค่าความผิดพลาดในแต่ละโหนดของกลุ่มข้อมูลที่นำมาทดสอบ ในขั้นแรกคำนวณค่าเฉลี่ยความแตกต่างระหว่างค่าที่ทำนายได้กับค่าข้อมูลจริง ซึ่งค่าที่ทำนายได้จากการเรียนรู้ข้อมูลของกลุ่มข้อมูลเรียนรู้ในแต่ละโหนด โดยที่แบบจำลองต้นไม้เอ็มไฟว์พีจะสร้างต้นไม้ตามกลุ่มข้อมูลเรียนรู้แต่ละชุด ค่าเฉลี่ยนี้จะประเมินค่าความผิดพลาดที่เกิดขึ้นต่ำกว่าค่าที่เป็นจริง ในกรณีข้อมูลที่ไม่เคยพบมาก่อนจึงได้แทนข้อมูลดังกล่าวด้วยการคูณ (n+v)/(n-v) เมื่อ n คือ จำนวนของตัวอย่างการเรียนรู้ในแต่ละโหนด และ v คือ จำนวนของตัวแปรในแบบจำลองเชิงเส้นที่ได้จากค่าของกลุ่มข้อมูลในแต่ละโหนด การประมาณค่าความผิดพลาดของแบบจำลองเชิงเส้นโหนดคือ การเปรียบเทียบกับค่าความผิดพลาดที่คาดหวังไว้จากส่วนย่อยของต้นไม้ เพื่อคำนวณค่าความผิดพลาดที่จะเกิดขึ้นต่อมาในภายหลัง ซึ่งจะใช้ค่าน้ำหนักในกิ่งของต้นไม้

2.1.4.4 **การปรับให้เรียบ** (Smoothing Process)

ขั้นตอนการปรับให้เรียบช่วยชดเชยความไม่ต่อเนื่องของข้อมูล ระหว่างสร้างแบบจำลองเชิงเส้นซึ้งเป็นผลมาจากการตัดเล็มต้นไม้ และมีจำนวนตัวอย่างการเรียนรู้มีน้อย โดยในขั้นแรกจะใช้แบบจำลองของโหนดใบคำนวณหาค่าประมาณ หลังจากนั้นจึงกรองข้อมูลจากเส้นทางไปยังโหนดราก ต่อจากนั้นจึงปรับให้เรียบในแต่ละโหนดด้วยการรวมค่าข้อมูลกับค่าที่ได้จากการประมาณของแบบจำลองในโหนดด้วยสมการดังต่อไปนี้

$$p' = \frac{np + kq}{n + k} \tag{2.12}$$

กำหนดให้

p' คือ การประมาณกลับไปยังโหนดที่สูงกว่า

p คือ การประมาณกลับมายังโหนดจากโหนดที่ต่ำกว่า

q คือ ค่าประมาณโดยแบบจำลองของโหนดในชั้นเดียวกัน

n คือ จำนวนตัวอย่างการเรียนรู้ของโหนดที่ต่ำกว่า

k คือ ค่าคงที่ (ค่าที่เลือกเท่ากับ 15) ทั้งนี้ จะเห็นได้ว่า การปรับให้เรียบจะเพิ่มความแม่นยำในการประมาณ

อย่างไรก็ตามขั้นตอนการปรับให้เรียบสามารถที่จะทำได้โดยการรวมแบบจำลองเชิงเส้นเข้ากับแต่ละโหนดของต้นไม้ดังเช่น แบบจำลองของโหนดใบจะรวมกันหลังจากที่ต้นไม้ถูกสร้างเสร็จแล้วดังนั้นในการจำแนกข้อมูลจะมีเพียงแบบจำลองโหนดใบเท่านั้นที่ถูกใช้ ซึ่งข้อเสียคือแบบจำลองที่โหนดใบจะใหญ่ และยากที่จะเข้าใจได้เนื่องจากมีจำนวนตัวแปรมาก

2.1.4.5 **ค่าความผิดพลาด** (Missing Values)

ค่าความผิดพลาดหาด้วยการปรับเปลี่ยนสมการ SDR โดยที่ให้รวมเอาตัวแปรการชดเชยเข้าไปในสมการเบี่ยงเบนด้วยคือ

$$SDR = \frac{m}{|T|} \times \left[sd(T) - \sum_{j \in \{L,R\}} \frac{|T_j|}{|T|} \times sd(T_j) \right] \tag{2.13}$$

กำหนดให้

m คือ จำนวนของค่าข้อมูลที่ปราศจากความผิดพลาดของค่าคุณลักษณะ

T คือ กลุ่มของค่าคงที่ในโหนด

T_L และ T_R คือ กลุ่มของผลลัพธ์จากการแบ่งของค่าคุณลักษณะ

เมื่อประมวลผลจากกลุ่มข้อมูลเรียนรู้ และทดสอบ ซึ่งค่าคุณลักษณะที่เลือกใช้แบ่งกลุ่มข้อมูลเข้าสู่ส่วนย่อยของต้นไม้ตามค่าคุณลักษณะ โดยที่แนวทางพื้นฐานนั้นจะใช้ค่ากลุ่มข้อมูลที่แทนด้วยค่าคุณลักษณะที่เกี่ยวข้องกัน ซึ่งสาเหตุของความผิดพลาดเกิดขึ้นขณะประมวลผลข้อมูลในกลุ่มตัวอย่างที่ไม่เคยพบมาก่อน ซึ่งวิธีการพื้นฐานที่ใช้การแทนข้อมูลด้วยค่าเฉลี่ยของค่าคุณลักษณะของกลุ่มข้อมูลเรียนรู้ โดยเรียนรู้จนกระทั่งถึงโหนดที่ได้รับผลกระทบในกรณีที่ค่าคุณลักษณะเกี่ยวข้องกัน ซึ่งการพิจารณารายละเอียดแทนค่ากลุ่มด้วยข้อมูลค่า

คุณลักษณะที่เรียนรู้จากกลุ่มตัวอย่าง โดยในขั้นตอนแรกค่าข้อมูลทั้งหมดจะถูกแบ่งค่าคุณลักษณะด้วยค่าขีดแบ่ง (Threshold) ซึ่งเป็นทางเลือกหนึ่ง โดยเรียงลำดับค่าข้อมูลแต่ละความเป็นไปได้ในการแบ่งคำนวณด้วยสมการ SDR หลังจากนั้นจึงคำนวณตามด้วยสมการเลือกจุดแบ่งเพื่อลดความผิดพลาด และช่วยตัดสินใจเลือกจุดแบ่งที่ดีที่สุด

แบบจำลองต้นไม้เอ็มไฟว์พีได้รวบรวมสมการถดถอยกับต้นไม้ถดถอยเข้าด้วยกันเป็นต้นไม้ที่ทุกโหนดของต้นไม้ประกอบไปด้วยแบบจำลองเชิงเส้น แบบจำลองของต้นไม้เป็นสิ่งที่เหมาะสมกับฟังก์ชั่นต่อเนื่องที่มีการค่าของข้อมูลเป็นเชิงเส้นทำงานมากกว่า สมการถดถอยเชิงเส้น และต้นไม้ถดถอยซึ่งมีขนาดเล็กจึงทำให้สามารถเข้าใจได้ง่ายกว่า นอกจากนี้ค่าเฉลี่ยความผิดพลาดที่ได้จากการเรียนรู้ข้อมูลในกลุ่มเรียนรู้น้อยกว่า (Witten and Frank, 2005)

2.1.5 **การแบ่งกลุ่มข้อมูล** (Database Clustering)

การแบ่งกลุ่มข้อมูลถูกใช้จัดเตรียมข้อมูลก่อนนำข้อมูลไปวิเคราะห์ด้วยเครื่องมือที่ใช้ในการเรียนรู้ ซึ่งเป็นกระบวนการหนึ่งของการทำเหมืองข้อมูลที่ใช้จัดการกลุ่มข้อมูลขนาดใหญ่ โดยนำข้อมูลมาวิเคราะห์ดึงความรู้ หรือสิ่งสำคัญที่ซ่อนอยู่ภายในกลุ่มข้อมูลออกมา ซึ่งมีประโยชน์มากในการค้นหาข้อมูลที่ดีที่สุดภายในกลุ่มข้อมูล นอกจากนี้ยังช่วยลดขนาดของข้อมูลที่เกี่ยวข้องกันให้อยู่ในกลุ่มเดียวกัน โดยแต่ละกลุ่มของข้อมูลนั้นจะประกอบไปด้วยความคล้ายคลึงกันของข้อมูลตัวหนึ่งกับข้อมูลอีกตัวหนึ่งในกลุ่มข้อมูล ซึ่งจะวัดความใกล้เคียงกันของข้อมูลกับกลุ่มข้อมูลเมื่อข้อมูลถูกแบ่งกลุ่มค่าคุณลักษณะของข้อมูลแต่ละกลุ่มจะแตกต่างกัน

2.1.5.1 **การแบ่งกลุ่มด้วยค่าเฉลี่ยเคกลุ่ม** (K-Means Algorithm)

ค่าเฉลี่ยเคกลุ่มเป็นกระบวนการที่ใช้ในการแบ่งกลุ่มข้อมูล โดยใช้ความแตกต่างของค่าคุณลักษณะที่อยู่ในข้อมูลแต่ละตัว ซึ่งจะแยกข้อมูลที่มีค่าคุณลักษณะที่ใกล้เคียงกันให้อยู่กลุ่มเดียวกัน ซึ่งในการกำหนดข้อมูลใดอยู่ในกลุ่มใดของค่าเฉลี่ยเคกลุ่มนั้นจะแบ่งกลุ่มข้อมูลด้วยการกำหนดจำนวนกลุ่มข้อมูล (k) ซึ่งเมื่อกำหนดจำนวนกลุ่มแล้วนั้นจึงหาจุดค่าจุดศูนย์กลางของกลุ่มข้อมูลตามที่ได้กำหนดไว้ ซึ่งค่าศูนย์กลางเริ่มต้นนั้นได้จากการสุ่ม หลังจากนั้นจึงใช้สมการวัดค่าความห่างของยูคลิด (Euclidian Distance) ระหว่างข้อมูลกับข้อมูลศูนย์กลาง ดังต่อไปนี้

$$J = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \qquad (2.14)$$

กำหนดให้

J คือ ผลรวมของผลต่างระหว่างจุดศูนย์กลางของข้อมูลกำลังสอง

n คือ จำนวนกลุ่มที่จะแบ่ง

x_i คือ เซตของข้อมูลกลุ่มที่ i เมื่อ $i = 1, 2, 3, \mathrm{K}\ , k$

y_i คือ เซตของข้อมูลกลุ่มที่ i เมื่อ $i = 1, 2, 3, \mathrm{K}\ , k$

สมการผลรวมกำลังสองของความห่างระหว่างข้อมูล และข้อมูลศูนย์กลางของกลุ่มถูกคำนวณ เพื่อบ่งบอกว่าข้อมูลดังกล่าวอยู่ในกลุ่มใด จากการแบ่งกลุ่มด้วยวิธีที่กล่าวมาข้างต้น ช่วยลดขนาดของข้อมูลที่จะใช้ในการวิเคราะห์ เพื่อแก้ปัญหาที่เกิดขึ้นจากข้อมูลดังกล่าวให้ได้ความถูกต้องมากขึ้น และลดสิ่งรบกวนต่อข้อมูล

2.1.6 **ข้อมูลอนุกรมเวลา** (Time Series Data)

ข้อมูลอนุกรมเวลา (Time Series data) คือ ชุดของข้อมูลที่เก็บรวบรวมตามระยะเวลาเป็นช่วง ๆ อย่างต่อเนื่องกัน เช่น ข้อมูลยอดขายสินค้าที่เก็บรวบรวมต่อเนื่องกันไปเป็นระยะเวลาหลาย ๆ เดือน ข้อมูลรายได้ประชาชาติปีต่าง ๆ ที่เก็บรวบรวมต่อเนื่องกันไปเป็นระยะเวลาหลาย ๆ ปี เป็นต้น

ในการวิเคราะห์อนุกรมเวลา ผู้วิเคราะห์จะแยกองค์ประกอบต่าง ๆ ที่ประกอบกันขึ้นเป็นอนุกรมเวลา โดยจะมีการเปลี่ยนแปลงไปตามอิทธิพลต่าง ๆ เช่น การเปลี่ยนแปลงการผลิต เทคโนโลยี สภาวะอากาศ เป็นต้น ในการหาคุณลักษณะของอนุกรมเวลาเราสามารถใช้แบบจำลองได้หลายแบบ แบบจำลองที่ใช้โดยนักเศรษฐศาสตร์แบบหนึ่งคือ แบบจำลองแบบคลาสสิก (Classical model) เป็นการอธิบายถึงองค์ประกอบของการแปรผันของอนุกรมเวลา 4 ส่วน ดังต่อไปนี้

1. ค่าแนวโน้ม (Secular trend) แทนด้วย T_t

2. การเปลี่ยนแปลงหรือความแปรผันตามฤดูกาล (Seasonal Variation) แทนด้วย S_t

3. การเปลี่ยนแปลงหรือความผันแปรตามวัฏจักร (Cyclical Variation) แทนด้วย C_t

4. การเปลี่ยนแปลงหรือความผันแปรเนื่องจากเหตุการณ์ผิดปกติ (Irregular Variation) แทนด้วย I_t

- **การเปลี่ยนแปลงหรือความแปรผันตามฤดูกาล** (Seasonal Variation) **แทนด้วย** S_t

เป็นการเปลี่ยนแปลงข้อมูลมีลักษณะการเพิ่มขึ้น หรือลดลงในลักษณะเดียวกันของรอบระยะเวลาหนึ่งที่แน่นอนเรียกว่า การเปลี่ยนแปลงตามฤดูกาล หน่วยของระยะเวลาสำหรับข้อมูลอาจเป็นรายชั่วโมง รายวัน รายสัปดาห์ รายเดือน รายไตรมาส สำหรับข้อมูลรายปีไม่มีการแปรผันตามฤดูกาล การเปลี่ยนแปลงตามฤดูกาลนั้นกำหนดระยะเวลาการเกิดซ้ำในรอบหนึ่ง ๆ ได้ค่อนข้างแน่นอน ตัวอย่างเช่น ยอดขายรายเดือนของห้างสรรพสินค้าแห่งหนึ่ง

- **การเปลี่ยนแปลงหรือความแปรผันตามวัฏจักร** (Cyclical Variation) **แทนด้วย** C_t

การเปลี่ยนแปลงตามวัฏจักร มีการเปลี่ยนแปลงเคลื่อนไหวในลักษณะซ้ำ ๆ กัน และจะมีลักษณะคล้ายคลึงกับการเปลี่ยนแปลงตามฤดูกาล แต่จะต่างกันก็ตรงที่การเปลี่ยนแปลงตามวัฏจักรแต่ละรอบจะใช้ระยะเวลาที่นานกว่าคือ ตั้งแต่ 5 ปีขึ้นไป ข้อมูลที่มีการเปลี่ยนแปลงตามวัฏจักรในทางธุรกิจ เรียกว่า "วัฏจักรธุรกิจ" (Business Cyclical) โดยทั่วไปประกอบด้วย ระยะเจริญรุ่งเรือง (prosperity) ระยะฝืดเคือง (recession) ระยะตกต่ำ (depression) และระยะขยายตัว (recovery)

- **การเปลี่ยนแปลงหรือความแปรผัน เนื่องจากเหตุการณ์ผิดปกติ** (Irregular Variation) **แทนด้วย** I_t

เป็นการเปลี่ยนแปลงของข้อมูลอนุกรมเวลาที่เกิดจากเหตุการณ์ที่ไม่สามารถคาดการณ์ได้ล่วงหน้า เช่น การเกิดไฟไหม้ในโรงงาน การเกิดอุทกภัย การนัดหยุดงานของคนงาน แผ่นดินไหว เป็นต้น ซึ่งเหตุการณ์เหล่านี้เป็นสิ่งที่เกิดขึ้นโดยบังเอิญไม่คาดคิดมาก่อน เป็นการเปลี่ยนแปลงที่เป็นเชิงสุ่ม (Random variation) เพราะไม่ได้อยู่ภายใต้เงื่อนไขที่กำหนด

จากองค์ประกอบของอนุกรมเวลาทั้ง 4 อย่าง คือ T S C และ I ในข้อมูลอนุกรมชุดหนึ่ง ๆ ไม่จำเป็นต้องครบองค์ประกอบข้างต้นก็ได้ ทั้งนี้ ขึ้นอยู่กับชนิดของข้อมูล

จากปัจจัยทั้ง 4 ข้างต้น ถ้า Y แทนข้อมูลอนุกรมเวลาชุดหนึ่ง ๆ เราสามารถกำหนดแบบจำลองได้ 2 แบบ ดังนี้

1. แบบจำลองผลบวก (Additive model) ถือว่าข้อมูลในแต่ละอนุกรมเวลาประกอบด้วยผลบวกขององค์ประกอบทั้ง 4 อย่าง

$$Y_t = T_t + S_t + C_t + I_t \qquad (2.15)$$

2. แบบจำลองผลคูณ (Multiplicative model) ถือว่าข้อมูลในแต่ละอนุกรมเวลาประกอบด้วยผลคูณขององค์ประกอบทั้ง 4 อย่าง

$$Y_t = T_t \times S_t \times C_t \times I_t \quad (2.16)$$

โดยทั่วไปข้อมูลอนุกรมเวลาจะมีความสัมพันธ์ในรูปแบบจำลองผลคูณ เนื่องจากเป็นการพิจารณาการเปลี่ยนแปลงในรูปอัตราร้อยละ ซึ่งจะทำให้ผลการวิเคราะห์ใกล้เคียงความเป็นจริงมากกว่าการใช้แบบจำลองผลบวก

2.1.7 **วิธีการทดสอบแบบไขว้ข้าม 10 กลุ่ม** (10 - Fold Cross Validation)

วิธีการทดสอบแบบไขว้ข้ามเป็นเทคนิคที่ใช้ประเมินผลประสิทธิภาพของแบบจำลองที่สร้างมาด้วยเครื่องมือต่าง ๆ ที่ถูกเลือกใช้ ซึ่งวิธีการไขว้ข้ามกลุ่มทดลองแบ่งกลุ่มข้อมูลออกเป็น 2 กลุ่ม คือส่วนของกลุ่มข้อมูลเรียนรู้ (Training Set) เป็นกลุ่มข้อมูลที่ถูกใช้ในการสร้างแบบจำลองที่ออกแบบไว้ และข้อมูลอีกกลุ่มหนึ่งที่ไม่ได้ใช้เรียนรู้นั้นจะถูกนำไปใช้เป็นกลุ่มข้อมูลทดสอบ (Test Set) ซึ่งเป็นกลุ่มข้อมูลที่ใช้วัดประสิทธิภาพการทำงานของแบบจำลองที่ออกแบบมาด้วยการให้แบบจำลองจำแนกข้อมูลในกลุ่มข้อมูลทดสอบ

วิธีการทดสอบแบบไขว้ข้าม n กลุ่ม กลุ่มข้อมูลได้ถูกแบ่งออกเป็น n ส่วน โดยแต่ละส่วนที่แบ่งจะมีขนาดเท่า ๆ กัน และกลุ่มที่แบ่งออกมานั้นจะถูกเรียกว่า “โฟลด์” (Fold) ซึ่งจะนำมาเป็นกลุ่มข้อมูลสำหรับเรียนรู้จำนวน n-1 และกลุ่มข้อมูลที่ใช้ทดสอบ 1 ส่วน ในวิธีการนี้จะถ่ายทอดความแตกต่างของข้อมูลในกลุ่มย่อยที่ถูกสร้างขึ้น ในบางครั้งข้อมูลได้มาจากการสุ่ม โดยจะสุ่มข้อมูลเริ่มต้นในกลุ่มตามจำนวนกลุ่มที่ถูกสร้างจะแตกต่างกันจากการสุ่ม ซึ่งในบางครั้งอาจได้ข้อมูลที่ไม่เหมือนกัน เทคนิคการแบ่งกลุ่มสำหรับเรียนรู้ และทดสอบดังกล่าวใช้วัดประสิทธิภาพการเรียนรู้ของข้อมูลทั้งหลายด้วยการประมาณตามสมมติฐานที่จะทำนายข้อมูลในอนาคตที่มองไม่เห็น และใช้ทดสอบประสิทธิภาพการทำนายข้อมูลตามสมมติฐานที่ต้องการจะได้จากข้อมูลกลุ่มทดสอบ (Witten and Frank, 2005)

ในการศึกษาครั้งนี้กลุ่มข้อมูลที่ใช้เรียนรู้สร้างแบบจำลองด้วยการถดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอ็มไฟว์พี ซึ่งข้อมูลจะถูกใช้ทดสอบว่าแบบจำลองที่สร้างด้วยเครื่องมือใดสามารถทำนายค่าเวลาการพิมพ์ได้ใกล้เคียงกับค่าเวลาที่ผู้ใช้สามารถทำได้ ซึ่งงานวิจัยนี้ใช้วิธีการทดสอบแบบไขว้ข้าม 10 กลุ่ม ที่แบ่งกลุ่มข้อมูลเรียนรู้ และกลุ่มข้อมูลทดสอบจำนวน 10 ชุด จากนั้นหาค่าเฉลี่ยความผิดพลาดของแต่ละแบบจำลองที่ทำนายได้ ซึ่งทำ

ให้สามารถมองเห็นภาพรวมความผิดพลาดที่เกิดขึ้น หรือประสิทธิภาพที่แบบจำลองแต่ละแบบทำได้ด้วยการทดสอบข้อมูลแต่ละชุดจะได้ค่าความผิดพลาดของแต่ละชุดข้อมูล เมื่อทดสอบจนครบทั้ง 10 ครั้ง ก็จะนำค่าความผิดพลาดทั้งหมดมาเฉลี่ย เพื่อให้ได้ภาพรวมของค่าความผิดพลาดทั้งหมด

2.2 งานวิจัยที่เกี่ยวข้อง

การศึกษาด้านทักษะการพิมพ์ได้รับการพัฒนากันอย่างต่อเนื่อง ทั้งการพัฒนารูปแบบการจัดวางตัวอักขระบนแป้นพิมพ์ และการพัฒนาแบบฝึกพิมพ์ ดังนั้นจึงมีงานวิจัยที่สนใจศึกษากระบวนการทำงานของสมองขณะพิมพ์ของมนุษย์เป็นจำนวนมาก เพื่อศึกษาทำความเข้าใจกระบวนการทำงาน และสามารถพัฒนาทักษะการทำงานให้มีประสิทธิภาพสูงขึ้น โดยกระบวนการทำงานของสมองจะประมวลผลข้อมูลที่ได้เรียนรู้ เพื่อหาขั้นตอนการทำงานที่จะนำมาซึ่งผลลัพธ์ตอบสนองกลับไปยังสภาพแวดล้อม และยังมีกระบวนการตรวจสอบความถูกต้องของผลลัพธ์ดังกล่าว นอกจากนี้ยังมีงานวิจัยที่จำลองการทำงานส่วนกระบวนการคิดของสมอง โดยให้ความสนใจในเรื่องความเร็วในการตอบสนองข้อมูล และความสามารถในการจดจำข้อมูล

ในปี 1938 Card et al. (อ้างถึงไว้ใน Preece et al., 1994) ได้เสนอแนวทางการสร้างแบบจำลองการทำงานของสมองมนุษย์ (Model Human Processor: MHP) อธิบายถึงรายละเอียด ขั้นตอน หน้าที่ และส่วนเก็บข้อมูลต่าง ๆ ภายในสมองมนุษย์ ซึ่งแบ่งส่วนการทำงานตามหน้าที่ของแต่ละส่วนที่เชื่อมโยงกัน และส่งผ่านข้อมูลจากส่วนหนึ่งไปยังอีกส่วนหนึ่ง จนงานที่รับเข้ามาจากสภาพแวดล้อมเสร็จสมบูรณ์ตามความต้องการ จากการทำงานรูปแบบดังกล่าวสามารถแบ่งออกเป็น 3 ส่วนดังต่อไปนี้ ส่วนประมวลผลการรับรู้ (Perceptual Processor) ทำหน้าที่รับข้อมูลจากสภาพแวดล้อมที่กระทำต่อมนุษย์ซึ่งในการเลือกรับข้อมูลที่เข้ามานั้นจะเลือกตามความสนใจในขณะนั้นว่าได้ให้สนใจทำงานใดอยู่ ในส่วนนี้มีความจำรับความรู้สึก (Sensory Memory) มีอยู่ 3 ประเภทคือ ความจำสิ่งกระตุ้นทางภาพ (Iconic Memory) ความจำสิ่งกระตุ้นทางเสียง (Echoic Memory) และความจำสิ่งกระตุ้นทางการสัมผัส (Haptic Memory) ซึ่งจะเก็บข้อมูลที่รับเข้ามาไว้ชั่วคราวหลังจากนั้นข้อมูลจะถูกส่งไปยัง ส่วนประมวลผลกระบวนการคิด (Cognitive Processor) วิเคราะห์ และประมวลผลข้อมูล ซึ่งจะแยกแยะ เรียนรู้ และค้นหากระบวนการทำงานที่จะได้มาซึ่งผลลัพธ์ในงานนั้น ๆ ซึ่งในส่วนนี้จะมีความจำระยะสั้น (Short Term Memory) ที่จะเก็บข้อมูลขณะทำงาน โดยความจำดังกล่าวจะเกิดขึ้นอย่างทันทีทันใดขณะทำงาน และสามารถเก็บข้อมูลได้เพียงระยะเวลาอันสั้นจำนวนเนื้อที่ในการเก็บข้อมูลมีอยู่อย่าง

จำกัด รูปแบบในการจัดเก็บข้อมูลจะแบ่งออกเป็นกลุ่ม ๆ ซึ่งในแต่ละกลุ่มจะเก็บข้อมูลได้ประมาณ 4 ถึง 5 ตัวอักขระ เมื่อข้อมูลหรืองานนั้นได้รับเข้ามาเรียนรู้บ่อยครั้งในระยะเวลาหนึ่งข้อมูล และขั้นตอนการทำงานจะส่งไปจัดเก็บในความจำระยะยาว (Long Term Memory) เมื่อข้อมูลถูกจัดเก็บเข้าสู่ส่วนนี้แล้วจะส่งผลให้ระยะเวลาที่ใช้ในการทำงานลดลง เนื่องจากผู้ใช้มีความชำนาญในการทำงานแล้ว และทักษะการทำงานจะค่อย ๆ ลบเลือนลงเมื่อผู้ใช้ไม่ได้ทำงานชิ้นนั้นนานเป็นระยะเวลาหนึ่ง ซึ่งหลังจากข้อมูลผ่านการประมวลผลข้างต้นเรียบร้อยแล้วจะถูกส่งไปยังส่วนสั่งการอวัยวะต่าง ๆ ของร่ายกายทำงานตอบสนองคือ ส่วนประมวลผลการเคลื่อนไหว (Motor Processor) ซึ่งจะควบคุมให้อวัยวะทำงานตอบสนองต่อข้อมูลที่ได้รับเข้ามา

จากงานวิจัยที่กล่าวมาในข้างต้นได้มีการพัฒนาแบบจำลองตามแนวทางข้างต้น “ในการวิเคราะห์ความรู้สำหรับใช้ในการทำงานด้วยการกำหนดส่วนการทำงานออกเป็น 4 ส่วนได้แก่ เป้าหมาย (Goals) ส่วนควบคุมการทำงาน (Operators) ส่วนขั้นตอนการทำงาน (Methods) และส่วนเลือกกฎการทำงาน (Selection rules)” โดยงานวิจัยของ John and Kieras (1996) เป็นงานวิจัยที่พัฒนาแบบจำลอง GOMS เพื่อศึกษาพฤติกรรมการทำงานของผู้ใช้ตามขั้นตอนที่กล่าวมาในข้างต้นครอบคลุมการทำงานให้บรรลุเป้าหมายที่สนใจอยู่ในขณะนั้นให้เสร็จตามระยะเวลาที่กำหนด โดยที่ส่วนควบคุมการทำงานจะควบคุมกระบวนการคิด และการเคลื่อนไหวของกล้ามเนื้อ ส่วนขั้นตอนการทำงานจะจัดลำดับขั้นตอนการทำงานที่กำหนดมาจากส่วนควบคุม เพื่อทำงานให้สำเร็จตามเป้าหมาย และส่วนเลือกกฎการทำงานจะเลือกกฎที่จะใช้ในแต่ละงานต้องการ ในบางกรณีมีได้มากกว่า 1 กฎที่จะทำให้งานเสร็จ ซึ่งการเลือกนั้นเลือกตามความเหมาะสมในแต่ละสถานการณ์ และตามประสบการณ์ของผู้ใช้ ในงานวิจัยที่พัฒนาเพิ่มเติมในแนวทางเดียวกันแต่จะแตกต่างกันในกระบวนการทำงานบางประการ ได้แก่ CMN-GOMS KLM NGOMSL และ CPM-GOMS เป้าหมายในการพัฒนางานที่กล่าวมาทั้งหมดนี้ เพื่อสร้างแบบจำลองทางคอมพิวเตอร์ที่ช่วยลดระยะเวลา และต้นทุน ที่ใช้ในการออกแบบระบบการติดต่อระหว่างเครื่องคอมพิวเตอร์กับผู้ใช้ (User Interface) และยังสามารถทำนายระยะเวลาที่ใช้ในการทำงานระหว่างผู้ใช้กับระบบ

งานวิจัยพัฒนาสถาปัตยกรรมด้วยการประยุกต์การทำงานในแนวทางเดียวกัน โดย Kieras and Mayer (1997) ได้พัฒนาสถาปัตยกรรมในส่วนกระบวนการคิดคือ สถาปัตยกรรม EPIC (Executive Process-Interaction Control) ซึ่งนำวิธีการทำงานของ MHP ที่แสดงในข้างต้น อีกทั้งยังศึกษาการทำงานในส่วนของอารมณ์ และประสิทธิภาพในการทำงานที่หลากหลายของมนุษย์ เพื่อสร้างแบบจำลองคำนวณความเปลี่ยนแปลงของปฏิสัมพันธ์ระหว่างมนุษย์กับคอมพิวเตอร์ที่จะเปลี่ยนไปตามสถานการณ์ และได้ออกแบบโครงสร้างซอฟต์แวร์ (Software

Framework) ไว้สำหรับพัฒนาแบบจำลองการทำงานด้วยคอมพิวเตอร์ที่แสดงถึงความต้องการขั้นตอนการทำงานพื้นฐานในงานที่มีความหลากหลายและซับซ้อน ซึ่งอยู่ในส่วนของกฎการทำงาน (Procedural Rule) เมื่อแบบจำลองเป็นส่วนรองรับสิ่งกระตุ้นที่เข้ามาจากภายนอก จุดมุ่งหมายในการพัฒนาสถาปัตยกรรมนี้แสดงถึงข้อจำกัดของความสามารถในการทำงาน และคุณลักษณะในการประมวลผลข้อมูลของสมองมนุษย์ โดยทดสอบความถูกต้องของสถาปัตยกรรม และจำลองกระบวนการที่มนุษย์ใช้ทำงาน การรับข้อมูลจากสภาพแวดล้อมได้มาจากอุปกรณ์ที่จำลองการทำงานติดต่อกับผู้ใช้ ส่วนความทรงจำที่เก็บข้อมูลลำดับขั้นตอนการทำงาน และข้อมูลของงานต่าง ๆ รวมไปถึงส่วนตอบสนองข้อมูลที่รับเข้ามาจากสภาพแวดล้อม และภายในส่วนของกระบวนการคิดนั้นจะมีส่วนกำหนดกฎ และลำดับขั้นตอนการทำงาน

งานวิจัยพัฒนาเพิ่มเติมในสิ่งที่ไม่ได้ออกแบบไว้ใน EPIC คือ ส่วนของขีดจำกัดในการเก็บข้อมูลของหน่วยความจำระยะสั้น ดังนั้น Anderson el at. (2004) จึงได้พัฒนาสถาปัตยกรรม ACT-R (The Adaptive Control of Thought Rational) ซึ่งเป็นแบบจำลองที่ใช้แก้ปัญหาการทำงาน การเรียนรู้ และความทรงจำของผู้ใช้ เป็นต้น ซึ่งได้ตั้งสมมติฐานของส่วนการทำงานของสถาปัตยกรรมของแบบจำลองอยู่ 3 ส่วนคือ ส่วนการประกาศ (Declarative Module) ส่วนขั้นตอนการทำงาน (Procedural Module) และส่วนเป้าหมาย (Goal Module) ซึ่งในส่วนการประกาศจะเกี่ยวข้องกับการเก็บข้อมูลในรูปแบบของหน่วยความจำย่อย (Chunk) ความสามารถในการเก็บข้อมูล และระยะเวลาที่ใช้ในการเก็บข้อมูล

ในการศึกษาขีดจำกัดและเวลาในการเก็บข้อมูลนั้น Fu el at. (2006) ได้ศึกษาด้วยการจัดเก็บระยะเวลา และจำนวนของข้อมูลที่จะถูกจัดเก็บไว้ในหน่วยความจำย่อย ซึ่งทดลองโดยป้อนชุดคำเข้าไปในแบบจำลอง และจับเวลาเมื่อเริ่มต้นพิมพ์ตัวอักขระตัวแรก จนกระทั่งพิมพ์ตัวอักขระของคำนั้นหมดจนกดแป้นตกลง (Enter) ในส่วนขั้นตอนการทำงานจะสร้างกฎการทำงาน (Production Rule) ที่ตอบสนองต่อรูปแบบของข้อมูลที่เข้ามาทางส่วนการประกาศ ซึ่งจะหากฎ และกระบวนการทำงานที่จะทำให้งานนั้นสำเร็จตามเป้าหมาย ในการกำหนดเป้าหมายจะแบ่งออกเป้าหมายย่อย ๆ และจะทำงานให้บรรลุเป้าหมายย่อยตามลำดับจนบรรลุเป้าหมายหลักที่กำหนดไว้

งานวิจัยอีกแนวทางได้ศึกษาการจำลองการเคลื่อนไหวของมือ (Movement Time) ซึ่งเป็นการศึกษาแบบแยกส่วนการทำงานเมื่อรับข้อมูลเข้ามา และหากระบวนการเคลื่อนไหวตอบสนองต่อข้อมูลดังกล่าว โดยสร้างสมการที่สามารถทำนายระยะเวลาที่ใช้ในการเคลื่อนที่ไปยังเป้าหมาย ในปี 1954 Fitts (อ้างถึงไว้ใน MacKenzie, 1992 และ McGuffin and Balakrishnan, 2005) ได้พัฒนากฎของฟิตต์ (Fitts' Law) ซึ่งเป็นสมการจำลองการเคลื่อนไหว เนื่องจาก

พัฒนาการในด้านทักษะของมนุษย์มีความเปลี่ยนแปลงตามลำดับเวลา และมีแนวโน้มที่สามารถคาดเดาได้ ซึ่งสมการนี้จะทำนายระยะเวลาที่ใช้ในการเคลื่อนที่ของมือจากตำแหน่งเริ่มต้นที่ประจำอยู่ไปยังตำแหน่งเป้าหมายที่ต้องการ โดยที่เวลาดังกล่าวจะสามารถใช้พัฒนาอุปกรณ์นำเข้าข้อมูลซึ่งมีพื้นฐานอยู่บนความเร็ว และความสนใจทำงาน ณ ขณะนั้นแต่ข้อจำกัดคือ ปัจจัยที่นำมาวิเคราะห์นั้นมีเพียงขนาด และความห่างของเป้าหมายจากจุดเริ่มต้น ในการทำนายความเร็ว ความแม่นยำการเคลื่อนที่ไปถึงเป้าหมาย และสนใจเพียงงานระดับล่างคือ การชี้ตำแหน่ง และการเคลื่อนที่ ซึ่งไม่ได้พิจารณางานเวลาการตอบสนอง หรือเวลาที่ใช้ในการเลือกกฎสำหรับงานที่เป็นทางเลือก

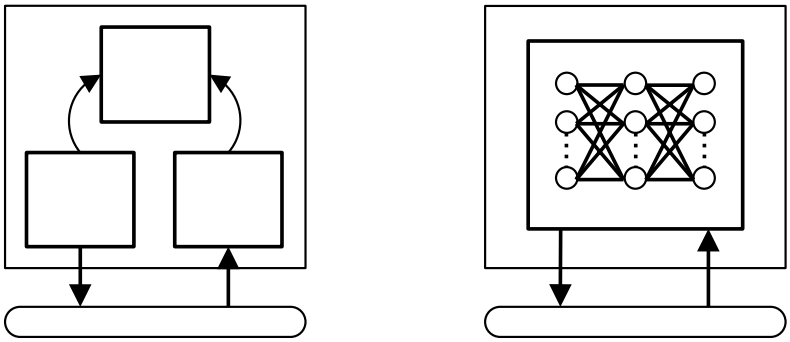
จากงานวิจัยที่ศึกษาเรื่องการเคลื่อนที่ดังกล่าว ต่อมาได้มีการศึกษาเวลาที่มนุษย์ใช้ในการตอบสนอง (Reaction Time) ต่อข้อมูลที่รับเข้ามา เพื่ออนุมานโครงร่างภายนอกของระบบจิตใจ และเสนอกลุ่มของแบบจำลองด้วยโครงข่ายแบบคิว (Queuing Network) Liu (1996) ที่จำลองส่วนการประมวลผลทางจิตใจแบบง่าย ซึ่งศึกษาข้อมูลอินพุตลักษณะเป็นค่าที่มีความต่อเนื่อง และไม่ต่อเนื่อง เพื่อใช้ทำนายพฤติกรรม และเวลาการตอบสนองของมนุษย์ Liu el at. (2006) เสนอสถาปัตยกรรมที่จำลองการทำงานสมองคือ Queuing Network-Model Human Processor (QN–MHP) โดยประยุกต์ 2 เทคนิคเข้าด้วยกันคือ โครงข่ายแบบคิว และสัญลักษณ์ (Symbolic Approach) ซึ่งทั้ง 2 เทคนิคมีลักษณะเด่นในการทำงานที่แตกต่างกันคือ โครงข่ายแบบคิวเป็นส่วนการทำงานหลาย ๆ อย่างในเวลาเดียวกัน และส่วนสัญลักษณ์เป็นการแสดงออกทางการเคลื่อนไหวที่มีลักษณะเฉพาะในแต่ละงาน ซึ่งการผสมผสาน 2 แนวทางดังกล่าว เพื่อสร้างแบบจำลองที่สามารถทำงานที่มีความซับซ้อน และจำลองการเคลื่อนไหวในแต่ละสถานการณ์ที่เปลี่ยนไป ในการพัฒนา 3 ส่วนตามแนวทางของ MHP คือ กระบวนการรับรู้ กระบวนการคิด และกระบวนการเคลื่อนไหวของกล้ามเนื้อ ประสิทธิภาพการทำงานที่ได้จากแบบจำลองมีความใกล้เคียงกับประสิทธิภาพที่มนุษย์สามารถทำได้ในงานเดียวกัน

Wu and Liu (2004, 2008) ได้ประยุกต์ใช้สถาปัตยกรรม QN-MHP ในงานด้านการพิมพ์จำลองพฤติกรรมการพิมพ์ (Transcription Typing) ซึ่งประกอบด้วย การพิมพ์ผิด การเคลื่อนไหวของตา และจินตนาการของสมอง โดยประสบความสำเร็จในการจำลองพฤติกรรมการพิมพ์ ในแต่ละพฤติกรรมการพิมพ์ที่จำลองเสมือนเป็นผลลัพธ์ของการกระทำที่จัดการกับข้อมูลที่รับเข้ามา ทั้งนี้แบบจำลองคำนวณการพิมพ์ไม่ได้เสนอแต่มุมมองทางทฤษฎีในประสิทธิการพิมพ์เท่านั้น แต่ยังพัฒนาการออกแบบทางสรีระศาสตร์ และการวิเคราะห์หาเครื่องมือที่จะใช้ในการออกแบบส่วนติดต่อกับผู้ใช้

จากงานวิจัยที่ศึกษาข้อมูลหากระบวนการสร้างแบบจำลองการทำงานของสมองขณะพิมพ์ที่สามารถทำนายระยะเวลาที่ใช้ในการพิมพ์ด้วยวิธีการต่าง ๆ ที่กล่าวมาในข้างต้นจะเห็นได้ว่าแบบจำลองที่พัฒนาออกมานั้นเป็นแบบจำลองที่วิเคราะห์การทำงานของสมองแยกออกเป็นส่วนตามหน้าที่ และความสามารถเฉพาะของแต่ละส่วนการทำงาน

ภาพที่ 2.5

รูปแบบของแบบจำลองการทำงานสมอง



ภาพที่ 2.5 แสดงแบบจำลองมีความแตกต่างเนื่องจากการศึกษาในครั้งนี้ใช้แบบจำลองที่มีความแตกต่างในขั้นตอนการสร้าง ข้อมูลที่ใช้สร้าง และเครื่องมือที่เลือกใช้มาสร้างแบบจำลอง ซึ่งเครื่องมือต่าง ๆ ที่งานวิจัยนี้เลือกใช้ได้แก่ แบบจำลองต้นไม้เอ็มไฟว์พี โครงข่ายงานประสาทเทียม และการถดถอยเชิงเส้น โดยเครื่องมือทั้งหลายที่งานวิจัยนี้เลือกใช้มีความสามารถในการทำนายข้อมูลที่เป็นตัวเลข และเนื่องจากที่งานวิจัยนี้ต้องการสร้างแบบจำลองของผู้ใช้แต่ละคนที่มีความแตกต่างกันในเรื่อง พฤติกรรมการทำงาน ทักษะการทำงาน และความเป็นเอกลักษณ์เฉพาะบุคคลที่แตกต่างกัน จึงได้นำข้อมูลเวลาการพิมพ์ของผู้ใช้แต่ละคนนำมาเป็นข้อมูลในการสร้างแบบจำลอง ซึ่งข้อมูลเวลาดังกล่าวเป็นข้อมูลตัวแทนการเรียนรู้ที่ได้จากการฝึกพิมพ์ดีดด้วยแบบฝึก หรือแป้นพิมพ์ต่าง ๆ อีกทั้งยังมีการเพิ่มการจัดการข้อมูลการพิมพ์ที่มีคุณลักษณะพิเศษที่มีความเปลี่ยนแปลงตามลำดับเวลา และในเรื่องของการพิมพ์ที่มีลักษณะการทำงานเป็นกลุ่ม ยกตัวอย่างเช่น การก้าวนิ้วในการพิมพ์คู่อักขระ เป็นต้น แบบจำลองที่สร้างด้วยกระบวนการเก่าที่แบ่งส่วนการทำงานเป็นส่วน ๆ ที่ทำหน้าที่วิเคราะห์ข้อมูลที่ได้รับมาจากสภาพแวดล้อม และส่งต่อกัน เพื่อให้ได้ผลลัพธ์ แต่ในงานวิจัยนี้จะเป็นแบบจำลองการเรียนรู้ของแต่ละบุคคลที่ใช้ทำงานนั้น ๆ ตามสมมติฐานไว้ว่าคนแต่ละคนมีการเรียนรู้ไม่เหมือนกันมีความเร็ว และประสิทธิภาพในการเรียนรู้ที่ไม่เท่าเทียมกัน ดังนั้นมนุษย์แต่ละคนจะมีสมการในการเรียนรู้ที่แตกต่างกันดังนั้น วิธีการที่

ใช้ในการสร้างแบบจำลองจึงแตกต่างกันด้วยการกำหนดค่าคุณลักษณะจากเวลาที่ใช้ในการพิมพ์ของผู้ใช้แต่ละคนรวมกับค่าคุณลักษณะอื่นที่จะแสดงในส่วนวิธีดำเนินงานวิจัย