

จากรูปแบบข้อมูลข้างต้นจึงแบ่งกลุ่มข้อมูลที่ใช้ทดลองออกเป็น 2 กลุ่มคือ ข้อมูลที่ใช้สำหรับเรียนรู้ (Training Set) และข้อมูลที่ใช้ทดสอบ (Test Set) ตามวิธีการทดสอบประสิทธิภาพแบบไขว้ข้าม 10 กลุ่ม (10 fold Cross validation) ที่จะทดลองทั้งหมด 10 ครั้งด้วยการถดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอมไพร์

3.3.4 การเรียนรู้ของโครงข่ายประสาทเทียม

การเรียนรู้ของโครงข่ายประสาทเทียมมีขั้นตอนต่าง ๆ ดังแสดงดังต่อไปนี้

- 1) กำหนดค่าน้ำหนักเชื่อมต่อในโครงข่ายเริ่มต้นให้มีค่าน้อย ๆ หรืออาจสุ่มค่าตัวเลขที่มีค่าต่ำ ๆ (-1.0 ถึง +1.0 หรือ -0.5 ถึง +0.5)
- 2) วงการทำงาน (ขั้นตอนที่ 3-10) ตรวจสอบเงื่อนไขไม่ถูกต้องจะดำเนินการตามขั้นตอนที่ 3 – 10 ต่อไป
- 3) การฝึกฝน (ขั้นตอนที่ 4 – 10) เพื่อให้โครงข่ายประสาทเทียมได้เรียนรู้ซึ่งประกอบด้วยหลักคือ การเรียนรู้แบบป้อนไปข้างหน้า (ขั้นตอนที่ 3-5) คำนวณค่าความผิดพลาดที่เกิดขึ้นจากวิธีแพร่กระจายย้อนกลับ (ขั้นตอนที่ 6-7) ปรับปรุงค่าน้ำหนัก และค่าการกระตุ้น (ขั้นตอนที่ 8)
- 4) รับสัญญาณอินพุตที่เข้ามาในแต่ละอินพุตโหนด และคำนวณค่าสัญญาณอินพุตของชั้นอินพุต เพื่อส่งต่อไปยังโหนดในชั้นซ่อน
- 5) คำนวณผลรวมค่าน้ำหนัก และค่าอินพุตที่รับเข้ามาจากชั้นอินพุต โดยการคำนวณหาสัญญาณอินพุตของโหนดในชั้นซ่อนนั้นสามารถหาได้จากสมการ (2.1) จากนั้นปรับค่าสัญญาณออกจากโหนดในชั้นซ่อนด้วยฟังก์ชันกระตุ้นแล้วส่งออกเป็นสัญญาณอินพุตของโหนดในชั้นเอาต์พุตดังสมการ (2.2)
- 6) คำนวณผลรวมค่าน้ำหนัก และค่าอินพุตที่รับเข้ามาในชั้นซ่อน ขั้นตอนนี้จะคำนวณคล้ายกับในขั้นตอนที่ 5 โดยคำนวณตามสมการ (2.1) และหาผลรวมค่าสัญญาณอินพุตที่ถูกส่งมาจากชั้นซ่อน ค่าน้ำหนักประจำตัวของอินพุตโหนดในชั้นเอาต์พุตซึ่งเชื่อมต่อมาจากโหนดในชั้นซ่อน จากนั้นปรับค่าสัญญาณออกจากโหนดในชั้นเอาต์พุตด้วยฟังก์ชันกระตุ้นดังสมการ (2.2)
- 7) การหาค่าความคลาดเคลื่อน (Error Signal Term) ในชั้นเอาต์พุตระหว่างค่าที่ได้จริง (Actual Output Values) กับเป้าหมาย (Desired Output Values) เพื่อปรับค่าน้ำหนักโดยสัดส่วนความคลาดเคลื่อนคูณกับค่าอินพุตในชั้นเอาต์พุตดังสมการ

$$\delta_k = \frac{1}{2}(T_k - Z_k) f'(in L_k) \quad (3.2)$$

กำหนดให้

δ_k คือ ค่าความคลาดเคลื่อนในชั้น k

T_k คือ ค่าเอาต์พุตเป้าหมาย

Z_k คือ ค่าเอาต์พุตที่ได้มาจากชั้น k

$in L_k$ คือ ค่าผลรวมสัญญาณอินพุตของชั้น k

คำนวณค่าความถูกต้องของน้ำหนัก (Weight correction term) เพื่อใช้ในการปรับปรุงน้ำหนักที่เชื่อมต่อในชั้นซ่อนกับชั้นเอาต์พุตดังสมการ

$$\Delta W_{jk} = \alpha \delta_k Y_j \quad (3.3)$$

กำหนดให้

ΔW_{jk} คือ ค่าความถูกต้องของน้ำหนักเชื่อมต่อในชั้น j กับ k

α คือ ค่าอัตราการเรียนรู้ (Learning Rate)

δ_k คือ ค่าความคลาดเคลื่อนในชั้น k

Y_j คือ ค่าสัญญาณเอาต์พุตของชั้น j

- 8) หาผลรวมเดลต้าอินพุต (Delta input) ในชั้นซ่อนชั้นตอนนี้แตกต่างไปจากในชั้นตอนที่ 7 เนื่องจากชั้นซ่อนจะไม่มีค่าเอาต์พุตเป้าหมาย (Output Target) จึงสมการดังต่อไปนี้

$$\delta_j = \sum_{k=1}^N \delta_k W_{jk} f'(in L_j) \quad (3.4)$$

กำหนดให้

δ_j คือ ค่าความคลาดเคลื่อนในชั้น j

W_{jk} คือ ค่าความถูกต้องของน้ำหนักเชื่อมต่อในชั้น j กับ k

δ_k คือ ค่าความคลาดเคลื่อนในชั้น k

$in L_j$ คือ ค่าผลรวมสัญญาณอินพุตของชั้น j

คำนวณค่าความถูกต้องของน้ำหนักเพื่อใช้ในการปรับปรุงค่าน้ำหนักที่เชื่อมต่อในชั้นอินพุตกับชั้นซ่อนดังสมการ (3.3)

9) ปรับปรุงค่าน้ำหนักแต่ละโหนดของชั้นเอาต์พุตด้วยสมการ

$$W_{jk}^t = W_{jk}^{t-1} + \Delta W_{jk} \quad (3.5)$$

กำหนดให้

W^t คือ ค่าน้ำหนักผ่านการปรับค่าในชั้น j กับ k

W^{t-1} คือ ค่าน้ำหนักก่อนปรับค่าในชั้น j กับ k

ΔW_{jk} คือ ค่าความถูกต้องของน้ำหนักเชื่อมต่อในชั้น j กับ k

จากนั้นปรับค่าน้ำหนักใหม่แต่ละตัวในชั้นซ่อนด้วยสมการ (3.5)

10) ทดสอบการทำงานของโครงข่ายประสาทเทียมหลังจากที่ผ่านกระบวนการฝึกฝนเสร็จสิ้นแล้ว หลังจากนั้นจึงป้อนข้อมูลทดสอบเข้าสู่โครงข่ายประสาทเทียมแบบหลายชั้น และคำนวณหาค่าเอาต์พุตจากค่าน้ำหนักที่มีการคำนวณไว้ในขั้นตอนการเรียนรู้

เงื่อนไขการหยุดฝึกกำหนดเงื่อนไขในการหยุดฝึกนั้นสามารถทำได้ 2 กรณีคือ เมื่อผลรวมของค่าผิดพลาดเฉลี่ยระหว่างเอาต์พุตกับเป้าหมายทั้งหมดมีค่าลดลงน้อยกว่าค่าที่ได้จากเอาต์พุตของโครงข่ายประสาทเทียมเทียบกับค่าเป้าหมายที่ใช้สำหรับบ่งชี้ผลของการฝึกฝนเพียงพอกับการใช้งานหรือไม่ อีกกรณีหนึ่งคือ การกำหนดจำนวนรอบของการฝึกฝนให้ดำเนินไปจนถึงค่าที่รอบการเรียนรู้ที่ตั้งไว้ ซึ่งจำนวนที่เหมาะสมได้จากการทดลอง

3.4 วิธีวัดผลการทดลอง

กระบวนการตรวจสอบค่าความผิดพลาดในการทำนาย หรือการคำนวณของแบบจำลองที่สร้างขึ้นด้วยวิธีการต่าง ๆ ที่กล่าวถึงในข้างต้น เพื่อวัดประสิทธิภาพในการทำงานของแบบจำลองค่าความผิดพลาดดังกล่าวได้จากการทำนายระยะเวลาการพิมพ์ของแบบจำลองกับค่าระยะเวลาการพิมพ์ที่ผู้ใช้ทำได้นำมาคำนวณตามสมการดังต่อไปนี้

$$RMSE = \sqrt{\frac{\sum_{patterns} (Desired - Actual)^2}{\# patterns}} \quad (3.6)$$

กำหนดให้

RMSE (Root Mean Squared Error) คือ ค่าเฉลี่ยของความผิดพลาดที่ได้จาก

การทำนายของโครงข่ายประสาทเทียม

Desired คือ ค่าที่ได้จากโครงข่ายประสาทเทียม

Actual คือ ค่าข้อมูลเป้าหมาย

Pattern คือ จำนวนของข้อมูลทั้งหมด

แบบจำลองจะทดสอบหาค่าเฉลี่ยความผิดพลาดกำลังสองที่ประเมินประสิทธิภาพด้วยวิธีการทดสอบแบบไขว้ข้ามสิบกลุ่ม ซึ่งการหาค่าความผิดพลาดหลังจากแบ่งเป็น 10 กลุ่ม และนำข้อมูลที่ได้ไปสร้างแบบจำลองจากกลุ่มข้อมูลเรียนรู้ และทดสอบด้วยกลุ่มข้อมูลทดสอบ หลังจากทำนายด้วยแบบจำลองจึงวัดประสิทธิภาพของแบบจำลองด้วยค่าเฉลี่ยความผิดพลาดกำลังสองของทั้ง 10 กลุ่ม เพื่อเปรียบเทียบประสิทธิภาพการทำงานของแบบจำลองที่สร้างขึ้นด้วยการถดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอมไพร์ฟี่ นอกจากนี้ยังรวมถึงการนำเอาเทคนิคแบ่งกลุ่มข้อมูลไม่ว่าจะเป็นการแบ่งกลุ่มข้อมูลแยกตามแถวของคู่อักขระ และค่าเฉลี่ยเคกลุ่ม อีกทั้งยังเพิ่มเทคนิคหน้าตาต่างเลื้อนที่จะจัดการปรับค่าข้อมูลอินพุต เพื่อหาระยะเวลาในการพิมพ์ตัวอักขระ และหาค่าอัตราการเรียนรู้ของผู้ใช้ ซึ่งนำค่าดังกล่าวไปจัดทำกราฟการเรียนรู้การพิมพ์ของคู่อักขระที่บ่งบอกถึงพัฒนาการในการพิมพ์ของผู้ใช้แต่ละคน

3.5 สมมติฐานของการวิจัย

1. การทดลองสร้างแบบจำลอง เพื่อเปรียบเทียบประสิทธิภาพของเครื่องมือสร้างแบบจำลองการทำงานของสมองขณะพิมพ์ด้วย การถดถอยเชิงเส้น โครงข่ายประสาทเทียม และแบบจำลองต้นไม้เอมไพร์ฟี่
2. การทดลองสร้างแบบจำลอง เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองที่ผ่านการแบ่งกลุ่มข้อมูลกับข้อมูลที่ไม่ผ่านการแบ่งกลุ่ม
3. การทดลองสร้างแบบจำลอง เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองที่ผ่านการแบ่งกลุ่มข้อมูลด้วยวิธีแยกตามแถวของคู่อักขระ กับค่าเฉลี่ยเคกลุ่ม

4. การทดลองสร้างแบบจำลอง เพื่อวัดประสิทธิภาพของแบบจำลองด้วยชุดข้อมูลที่
ผ่านการจัดการข้อมูลเปลี่ยนแปลงตามลำดับเวลาด้วยเทคนิคหน้าต่างเลื่อน