

บทที่ 5

การจำแนกเอกสารด้วยแบบจำลองโครงข่ายประสาท

บทนี้กล่าวถึง การจำแนกเอกสาร โดยใช้โครงข่ายประสาท เนื้อหาประกอบด้วย หลักการ การจำแนกเอกสารด้วยแบบจำลองโครงข่ายประสาท การออกแบบการทดลอง ผลการทดลอง และ วิจารณ์และสรุปผล

5.1 หลักการการจำแนกเอกสารด้วยแบบจำลองโครงข่ายประสาท

โครงข่ายประสาทถูกใช้ในงานเรียนรู้รูปแบบ และงานจำแนกกลุ่มอย่างแพร่หลาย รวมไปถึงงานการจำแนกกลุ่มเอกสารด้วย โดยให้แบบจำลองเรียนรู้ข้อมูลสำหรับเรียนรู้ (Training Data) เพื่อใช้ในการกำหนดค่าน้ำหนัก (Weight) ของเส้นเชื่อมระหว่างชั้น ของชั้นข้อมูลเข้าและชั้นถัดไป ตัวอย่างงานวิจัยที่ใช้โครงข่ายประสาทในงานจำแนกเอกสาร เช่น งานของ Lam (Lam and Lee, 1999) ได้นำเสนอการเลือกลักษณะเด่นที่ใช้กับโครงข่ายประสาท Karras (Karras and Mertzios, 2002) ได้นำเสนอการคัดทอนลักษณะเด่นสำหรับโครงข่ายประสาท Mostafa (Mostafa and Lam, 2000) ได้นำเสนอโครงข่ายประสาทแบบฟีดฟอร์เวิร์ด (Feed forward) ในการจำแนกเอกสารทางการแพทย์

ในการจำแนกกลุ่มโดยใช้โครงข่ายประสาทแบบหนึ่งที่เป็นที่สนใจอย่างแพร่หลาย คือ โครงข่ายประสาทเทียมแบบแบคพรอพาทิกชัน ซึ่งโครงข่ายนี้ให้ความถูกต้องในการจำแนกสูง และมีการปรับค่าน้ำหนักที่มีประสิทธิภาพ เนื่องจากการนำระดับความผิดพลาด (Error rate) มาเป็นส่วนร่วมในการปรับค่าน้ำหนัก แต่ปัญหาของโครงข่ายนี้ คือความไม่แน่นอนของผลลัพธ์ และเวลาการเรียนรู้นาน ดังนั้นงานวิจัยนี้จึงหาวิธีการในการปรับผลลัพธ์ในการจำแนกให้มีความถูกต้องที่สูงขึ้น และใช้เวลาการเรียนรู้ที่น้อยลง ด้วยวิธีการ การสร้างตัวแทนเอกสาร โดยการลดจำนวนของข้อมูล เข้าให้กับโครงข่ายประสาท จากเดิมข้อมูลเข้ามีจำนวนเท่ากับจำนวนคำสำคัญทั้งหมด ให้เหลือเท่ากับจำนวนการลดมิติของเอกสารด้วยวิธีที่เหมาะสม และ การหาโครงสร้างของโครงข่ายที่เหมาะสมแก่การจำแนกเอกสารชีวสารสนเทศที่นำมาทดสอบ

5.2 การออกแบบการทดลอง

ในการทดลองสำหรับงานวิจัยนี้ ได้ใช้โปรแกรมแมทแล็บ (Matlab) เวอร์ชัน 2007a โดยฟังก์ชันที่ใช้ทดลองคือ ฟังก์ชันฟีดฟอร์เวิร์ดแบคพรอพาทิกชัน ในการทดสอบการจำแนกข้อมูล

ชีวสารสนเทศที่ได้จากการสร้างดัชนีเอกสาร ในบทที่ 3 เพื่อใช้เป็นมาตรฐานในการวัดผล ประกอบด้วย ความถูกต้องในการจำแนกกลุ่มและเวลาการเรียนรู้ของแบบจำลองในการจำแนกเอกสารแบบฟีดฟอร์เวิร์ดแบคพรอพาทเกชัน และการลดมิติข้อมูลแต่ละแบบในบทที่ 4 เพื่อในการทดสอบ

ขั้นตอนในการทดลองประกอบด้วยขั้นตอนดังต่อไปนี้

1) ทำการกำหนดโครงสร้างของโครงข่ายประสาท ประกอบด้วย การกำหนดจำนวนโหนดในชั้นข้อมูลเข้า จำนวนโหนดในชั้นภายใน และจำนวนโหนดของชั้นข้อมูลออก ในที่นี้จำนวนโหนดของชั้นข้อมูลออกมีค่าเท่ากับจำนวนกลุ่มของเอกสารที่ใช้ในการทดลอง มีค่าเท่ากับ 7 กลุ่มเอกสาร ประกอบด้วย Biotechnology, Cancer, Evolutionary Biology, Family Practice, Infection Diseases, Medical Genetics และ Musculoskeletal Disorders ซึ่งมีรายละเอียดสำหรับข้อมูลเข้าแต่ละรูปแบบดังนี้

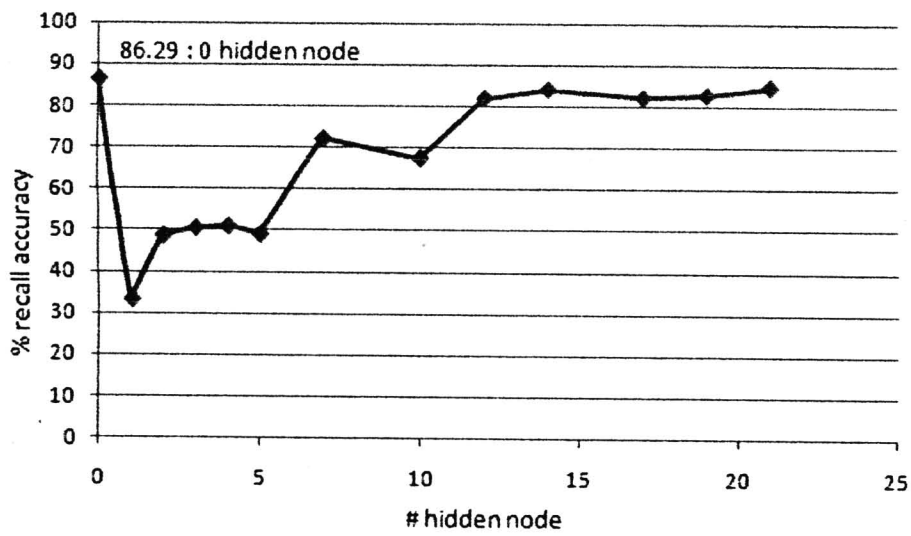
1) สำหรับการสร้างตัวแทนเอกสารด้วยดัชนีเอกสารในรูปเมตริกซ์คำสำคัญ-เอกสาร NN+term-doc ประกอบด้วย จำนวนโหนดชั้นข้อมูลเข้า มีจำนวนเท่ากับ 3586 และ จำนวนโหนดชั้นข้อมูลออกมีจำนวนเท่ากับ 7

2) สำหรับการสร้างตัวแทนเอกสารด้วยการเลือกลักษณะเด่นด้วยวิธีรีเลแวนซ์สก็อร์ NN+Re มีการเลือกคำสำคัญจำนวน k คำสำคัญ โดย k มีค่าระหว่าง 1-500 คำสำคัญ โหนดชั้นข้อมูลเข้า มีจำนวนเท่ากับ k และ จำนวนโหนดชั้นข้อมูลออก มีจำนวนเท่ากับ 7

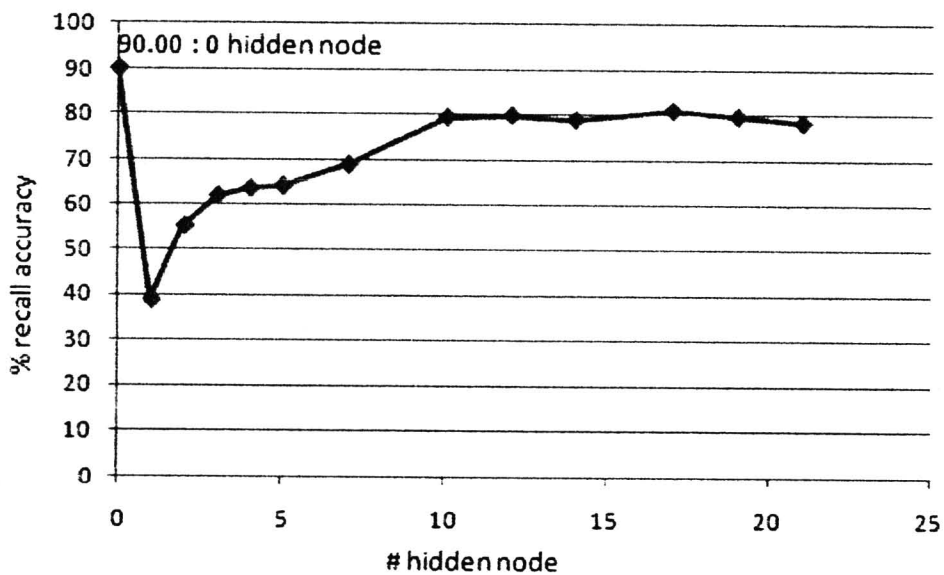
3) สำหรับการสร้างตัวแทนเอกสารด้วยการสกัดลักษณะเด่นด้วยวิธีแอลเอสไอ NN+LSI มีการเลือกมิติในการสร้างตัวแทนเอกสาร จำนวน k มิติ โดย k มีค่าระหว่าง 1-500 คำสำคัญ โหนดชั้นข้อมูลเข้า มีจำนวนเท่ากับ k และ จำนวนโหนดชั้นข้อมูลออก มีจำนวนเท่ากับ 7

4) สำหรับการสร้างตัวแทนเอกสารด้วยการสกัดลักษณะเด่นด้วยวิธีพีแอลเอส NN+PLS มีการเลือกมิติในการสร้างตัวแทนเอกสาร จำนวน k มิติ โดย k มีค่าระหว่าง 1-500 คำสำคัญ โหนดชั้นข้อมูลเข้า มีจำนวนเท่ากับ k และ จำนวนโหนดชั้นข้อมูลออก มีจำนวนเท่ากับ 7

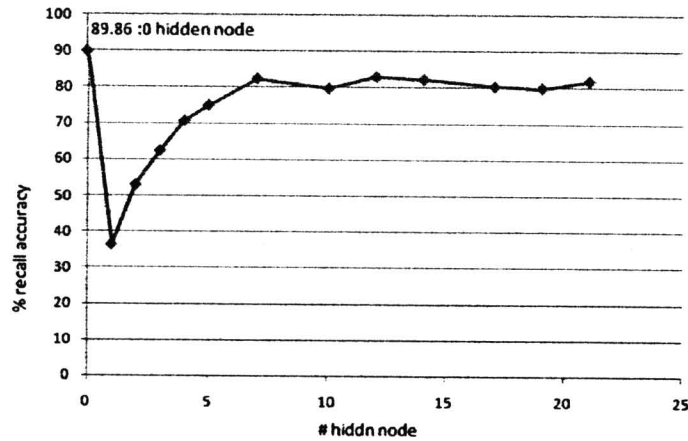
จากการกำหนดโครงสร้างของโครงข่ายประสาทของงานวิจัยนี้ ใช้โครงข่ายประสาทแบบชั้นเดียว (Single layer) เนื่องจาก ได้ทดลองการใช้โหนดของชั้นกลางระหว่าง 0 ถึง 21 โหนดกับกระบวนการสร้างตัวแทนเอกสารแบบต่างๆ พบว่าการเพิ่มจำนวนโหนดในชั้นกลางส่งผลให้โครงข่ายประสาทใช้เวลาการเรียนรู้เพิ่มขึ้น และ ทำให้ความถูกต้องในการจำแนกน้อยกว่าโครงข่ายประสาทแบบชั้นเดียว ดังรูป 5.1, 5.2, 5.3 และ 5.4



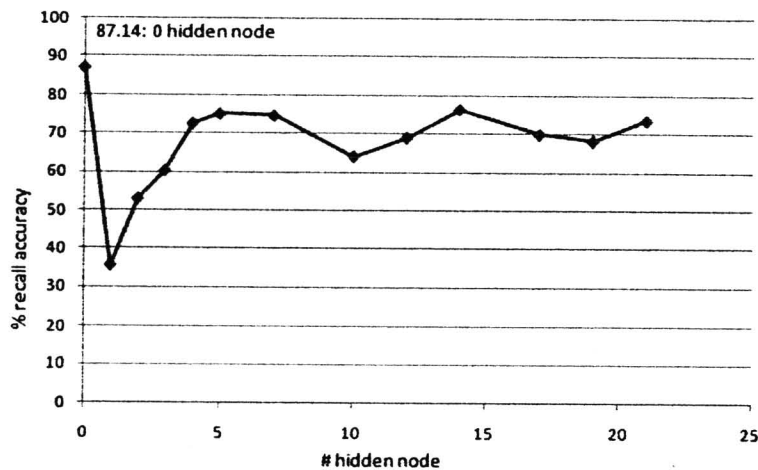
รูป 5.1 แสดงเปอร์เซ็นต์ความถูกต้องของการจำแนกจากการใช้โหนดในชั้นกลาง โดยมีข้อมูลเข้าคือ เมตริกซ์ค่าสำคัญ-เอกสาร



รูป 5.2 แสดงเปอร์เซ็นต์ความถูกต้องของการจำแนกจากการใช้โหนดในชั้นกลาง โดยมีข้อมูลเข้าคือ ตัวแทนเอกสารสร้างจากแอลเอสไอ 400 มิติ



รูป 5.3 แสดงเปอร์เซ็นต์ความถูกต้องของการจำแนกจากการใช้โหนดในชั้นกลาง โดยมีข้อมูลเข้าคือ ตัวแทนเอกสารสร้างจากพีแอลเอส 25 มิติ



รูป 5.4 แสดงเปอร์เซ็นต์ความถูกต้องของการจำแนกจากการใช้โหนดในชั้นกลาง โดยมีข้อมูลเข้าคือ เมตริกซ์ค่าสำคัญ-เอกสาร ได้จากการเลือกค่าสำคัญด้วยวิธีรีเลแวนซ์ที่สกออร์ 450 คำ

2) กำหนดค่าพารามิเตอร์ต่างๆ ของโครงข่ายประสาท ประกอบด้วยเงื่อนไขในการหยุดเรียนรู้ ดังนี้ ค่าผิดพลาดเฉลี่ยกำลังสอง (Mean square error) โดยกำหนดให้มีค่าเท่ากับ 0.0001 และ จำนวนรอบการเรียนรู้ (Epoch) สูงสุดคือ 30000 รอบ ค่ากราเดียนต์ต่ำสุด (Minimum gradient) เท่ากับ 1×10^{-6}

3) ทำการกำหนดกลุ่มของแต่ละเอกสาร โดยการสร้างเมตริกซ์กลุ่ม-เอกสาร Y โดยที่ m แทนกลุ่มและ r แทนเอกสาร จะได้ว่า

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1r} \\ y_{21} & y_{22} & \cdots & y_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mr} \end{bmatrix}$$

โดยกำหนด y_{ij} มีค่าเป็น 1 ก็ต่อเมื่อเอกสาร X_i เป็นสมาชิกของกลุ่มนั้น ๆ และมีค่าเป็น 0 ก็ต่อเมื่อเอกสาร X_i ไม่เป็นสมาชิกของกลุ่มนั้น

- 4) นำข้อมูลสำหรับเรียนรู้ไปให้แบบจำลองโครงข่ายแบบต่างๆที่ได้กำหนดไว้ในขั้นตอนที่ 1)
- 5) ทำการบันทึกเวลาการเรียนรู้ของแต่ละแบบจำลอง
- 6) นำแบบจำลองที่ทำการเรียนรู้แล้วแต่ละแบบจำลอง ไปจำแนกกลุ่มข้อมูลสำหรับทดสอบ
- 7) ทำการบันทึกผลการจำแนกของแต่ละแบบจำลอง

5.3 การวัดประสิทธิภาพ

การวัดประสิทธิภาพของการจำแนกเอกสาร ในด้านความถูกต้องของการจำแนก โดยใช้ การวัดแบบ $\text{Micro}_{\text{avg}}\text{F1}$ (Micro Average F1 Measure) มีสมการดังนี้

$$\text{Micro}_{\text{avg}}\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$$

โดยที่การหาค่า Precision และ Recall หาได้จาก สมการดังต่อไปนี้

$$\text{Precision} = \frac{\text{TP}}{\text{Tp} + \text{Fp}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{Tp} + \text{Fn}}$$

โดยที่ Tp คือ จำนวนที่แบบจำลองทำนายถูกต้องตามกลุ่มที่กำลังทดสอบ Fp คือ จำนวนที่แบบจำลองทำนายผิด โดยข้อมูลที่ทำนายไม่อยู่ในกลุ่มแต่ทำนายว่าเป็นสมาชิกของกลุ่ม และ Fn คือ จำนวนที่แบบจำลองทำนายผิด โดย ข้อมูลที่ทำนายเป็นสมาชิกในกลุ่มที่กำลังทดสอบแต่ทำนายว่าไม่ได้เป็นสมาชิกของกลุ่ม (Caropreso, Matwin and Sebastiani, 2001) และขั้นตอนการวัดความถูกต้อง เริ่มจาก สุ่มเอกสารในกลุ่มที่ต้องการทดสอบ 100 เอกสาร และสุ่มเอกสารที่อยู่นอกกลุ่มที่ต้องการทดสอบ 50 เอกสารจำนวน 7 กลุ่ม จากนั้นนำข้อมูลแต่ละชุดไปทดสอบจำแนกเอกสารทำการนับค่า Tp Fp และ Fn จากนั้นทำการคำนวณค่า Precision และ Recall จากนั้น ทำการหาค่า $\text{Micro}_{\text{avg}}\text{F1}$ สุดท้ายนำผลที่ได้มาใช้ในการเปรียบเทียบค่าความถูกต้อง

5.4 ผลการทดลอง

จากการทดลองเพื่อวัดความถูกต้องของการจำแนกกลุ่มของเอกสารและ การใช้เวลาการเรียนรู้ของโครงข่ายประสาทแบบฟีดฟอร์เวิร์ดแบคพรอพาเกชัน ที่มีข้อมูลเข้าคือตัวแทนเอกสารที่สร้างจากกระบวนการพีแอลเอส NN+PLS เมตริกซ์คำสำคัญสำคัญ-เอกสาร NN+term-doc การสร้างตัวแทนเอกสารด้วยวิธีการของแอลเอสไอ NN+LSI และการสร้างเมตริกซ์คำสำคัญ-เอกสาร โดยใช้

วิธีการเลือกคำสำคัญบางคำจากวิธีเลือกลักษณะเด่นด้วยวิธีการ รีเลแวนซ์สกออร์ NN+RE โดยการจำแนกแต่ละแบบยกเว้น NN+term-doc ใช้มิติเอกสารอยู่ในช่วง 1-500 มิติ ซึ่งช่วงความถี่ที่ใช้ทดสอบแต่ละแบบอาจจะไม่เท่ากัน โดยบางช่วงข้อมูลของแต่ละแบบที่ให้ความถูกต้องสูง อาจมีช่วงทดสอบที่ถี่ขึ้น เพื่อหาช่วงมิติที่เหมาะสมของแต่ละแบบ

5.4.1 ความถูกต้องของการจำแนกเอกสาร

การทดสอบความถูกต้องในการจำแนกเอกสารของโครงข่ายประสาทที่มีข้อมูลเข้าจากการสร้างตัวแทนเอกสารแต่ละแบบมีดังนี้

- 1) การจำแนกเอกสารที่มีข้อมูลเข้าได้จากการสร้างเมตริกซ์คำสำคัญ-เอกสาร

ตาราง 5.1 ผลการจำแนกเอกสารของ NN+term-doc

จำนวนคำสำคัญ	% Precision	% Recall	% Micro _{avg} F1
3586	97.89	86.29	91.72

- 2) การจำแนกเอกสารที่มีข้อมูลเข้าได้จากการสร้างตัวแทนเอกสารด้วยวิธีพีแอลเอส

ตาราง 5.2 ผลการจำแนกเอกสารของ NN+PLS

จำนวนมิติ	% Precision	% Recall	% Micro _{avg} F1
1	80.28	24.43	37.46
2	95.90	36.71	53.10
3	94.88	39.71	55.99
4	96.98	50.43	66.35
5	95.04	54.71	69.45
6	97.52	67.43	79.73
7	97.76	81.00	88.59
8	98.01	84.29	90.63
9	98.05	86.00	91.63
10	99.02	86.29	92.21
15	99.04	88.00	93.19
20	98.73	88.57	93.37
23	98.90	89.57	94.00

ตาราง 5.2 ผลการจำแนกเอกสารของ NN+PLS (ต่อ)

จำนวนมิติ	% Precision	% Recall	% Micro _{avg} F1
24	98.90	89.86	94.16
25	99.06	89.86	94.23
26	98.89	88.86	93.60
27	98.89	89.29	93.84
30	98.89	88.86	93.60
50	99.03	87.14	92.71
100	97.82	83.14	89.88
150	97.59	81.14	88.61
200	96.89	80.00	87.64
250	97.28	81.86	88.91
300	97.10	81.43	88.58
350	96.90	67.00	79.22
400	97.55	79.71	87.74
450	96.92	81.00	88.25
500	97.43	81.14	88.54

3) การจำแนกเอกสารที่มีข้อมูลเข้าได้จากการสร้างตัวแทนเอกสารด้วยวิธีแอลเอสไอ

ตาราง 5.3 ผลการจำแนกเอกสารของ NN+LSI

จำนวนมิติ	% Precision	% Recall	% Micro _{avg} F1
1	81.55	24.00	37.09
2	93.04	30.57	46.02
3	94.74	36.00	52.17
5	97.31	41.29	57.97
10	97.45	60.00	74.27
15	98.14	75.43	85.30
20	98.21	78.57	87.30
25	98.41	79.43	87.91

ตาราง 5.3 ผลการจำแนกเอกสารของ NN+LSI (ต่อ)

จำนวนมิติ	% Precision	% Recall	% Micro _{avg} F1
50	98.66	83.86	90.66
100	98.87	87.29	92.72
150	98.89	88.71	93.52
200	99.21	89.71	94.22
250	99.21	90.14	94.46
300	98.90	89.86	94.16
350	98.90	76.71	86.40
380	99.06	89.86	94.23
400	99.21	90.00	94.38
420	99.05	89.57	94.07
450	99.21	89.43	94.06
500	98.89	89.29	93.84

- 4) การจำแนกเอกสารที่มีข้อมูลเข้าได้จากการสร้างเมตริกซ์คำสำคัญ-เอกสาร โดยการเลือกคำสำคัญด้วยวิธีเรเวนซ์สเกอร์

ตาราง 5.4 ผลการจำแนกเอกสารของ NN+RE

จำนวนคำสำคัญ	% Precision	% Recall	% Micro _{avg} F1
1	76.42	23.14	35.53
2	76.96	25.29	38.06
3	84.91	38.57	53.05
4	86.45	41.00	55.62
5	80.14	31.71	45.45
6	91.73	52.29	66.61
7	92.14	53.57	67.75
8	92.62	55.57	69.46

ตาราง 5.4 ผลการจำแนกเอกสารของ NN+RE (ต่อ)

จำนวนคำสำคัญ	% Precision	% Recall	% Micro _{avg} F1
9	93.60	56.43	70.41
10	94.09	61.43	74.33
15	98.60	70.57	82.26
20	98.59	70.14	81.97
25	97.46	71.14	82.25
50	98.04	78.43	87.14
100	98.96	81.71	89.51
150	98.83	84.57	91.15
200	98.51	85.29	91.42
250	98.52	85.43	91.51
300	98.35	85.14	91.27
350	98.11	74.14	84.46
400	98.22	86.86	92.19
450	98.55	87.14	92.49
500	98.70	86.86	92.40

5.4.2 ด้านเวลาการเรียนรู้

จากการทดลอง ได้มีการจับเวลาเรียนรู้ของโครงข่ายประสาทที่มีข้อมูลเข้าจากการสร้างตัวแทนเอกสารแต่ละแบบ ดังนี้

- 1) เวลาการเรียนรู้ ของการจำแนกเอกสารที่มีข้อมูลเข้าได้จากการสร้างเมตริกซ์ คำสำคัญ-เอกสาร

ตาราง 5.5 เวลาการเรียนรู้ข้อมูลของ NN+term-doc

จำนวนคำสำคัญ	เวลาการเรียนรู้
3586	7868.59

2) เวลาการเรียนรู้ ของการจำแนกเอกสารที่มีข้อมูลเข้าได้จากการสร้างตัวแทนเอกสารด้วยวิธีพีแอลเอส

ตาราง 5.6 เวลาการเรียนรู้ข้อมูลของ NN+PLS

จำนวนมิติ	เวลาการสร้างพีแอลเอส	เวลาการเรียนรู้ NN+PLS	จำนวนมิติ	เวลาการสร้างพีแอลเอส	เวลาการเรียนรู้ NN+PLS
1	0.28	3.19	25	6.86	2.26
2	0.56	1.05	26	7.23	2.31
3	0.84	1.00	27	7.40	2.39
4	1.13	1.14	30	8.30	3.63
5	1.41	1.04	50	13.71	4.57
6	1.74	0.94	100	26.92	13.56
7	1.94	1.12	150	42.01	31.47
8	2.23	1.01	200	56.69	80.02
9	2.49	1.12	250	69.36	15.70
10	2.76	1.29	300	88.27	15.77
15	4.11	1.66	350	103.35	17.80
20	5.46	1.85	400	119.33	21.39
23	6.29	2.05	450	135.40	19.59
24	6.58	2.15	500	153.77	25.03

3) เวลาการเรียนรู้ ของการจำแนกเอกสารที่มีข้อมูลเข้าได้จากการสร้างตัวแทนเอกสารด้วยวิธีแอลเอสไอ

ตาราง 5.7 เวลาการเรียนรู้ข้อมูลของ NN+LSI

จำนวนมิติ	เวลาการสร้างแอลเอสไอ	เวลาการเรียนรู้ NN+LSI	จำนวนมิติ	เวลาการสร้างแอลเอสไอ	เวลาการเรียนรู้ NN+LSI
1	87.04	3.47	150	87.04	18.84
2	87.04	0.95	200	87.04	31.82

ตาราง 5.7 เวลาการเรียนรู้ข้อมูลของ NN+LSI (ต่อ)

จำนวนมิติ	เวลาการสร้าง แอลเอสไอ	เวลาการเรียนรู้ NN+LSI	จำนวนมิติ	เวลาการสร้าง แอลเอสไอ	เวลาการเรียนรู้ NN+LSI
3	87.04	1.69	250	87.04	31.03
5	87.04	2.14	300	87.04	49.48
10	87.04	2.28	350	87.04	66.78
15	87.04	2.48	380	87.04	61.58
20	87.04	4.70	400	87.04	60.81
25	87.04	5.67	420	87.04	68.36
50	87.04	11.34	450	87.04	90.33
100	87.04	12.42	500	87.04	140.73

- 4) เวลาการเรียนรู้ ของการจำแนกเอกสารที่มีข้อมูลเข้าได้จากการสร้างเมตริกซ์ค่า
สำคัญ-เอกสาร โดยการเลือกค่าสำคัญด้วยวิธีรีเลแวนซ์สกออร์

ตาราง 5.8 เวลาการเรียนรู้ข้อมูลของ NN+RE

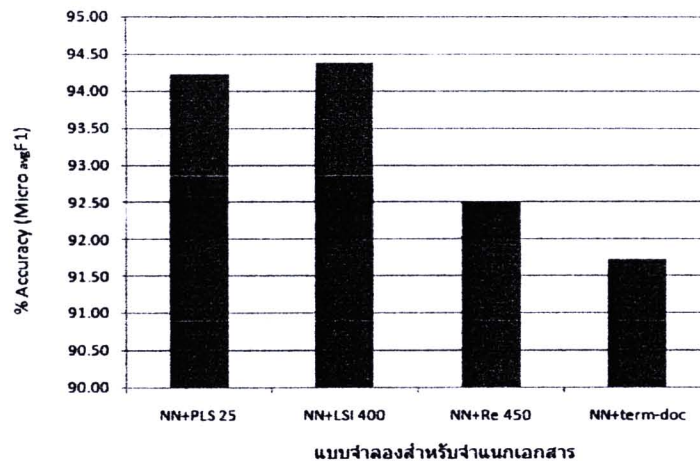
จำนวนมิติ	เวลาการเรียนรู้ NN+RE	จำนวนมิติ	เวลาการเรียนรู้ NN+RE
1	3.42	25	3.09
2	3.03	50	19.38
3	1.13	100	37.43
4	1.75	150	62.50
5	3.30	200	87.94
6	1.83	250	142.81
7	2.18	300	167.80
8	2.09	350	261.16
9	1.93	400	313.30
10	2.16	450	399.27

ตาราง 5.8 เวลาการเรียนรู้ข้อมูลของ NN+RE (ต่อ)

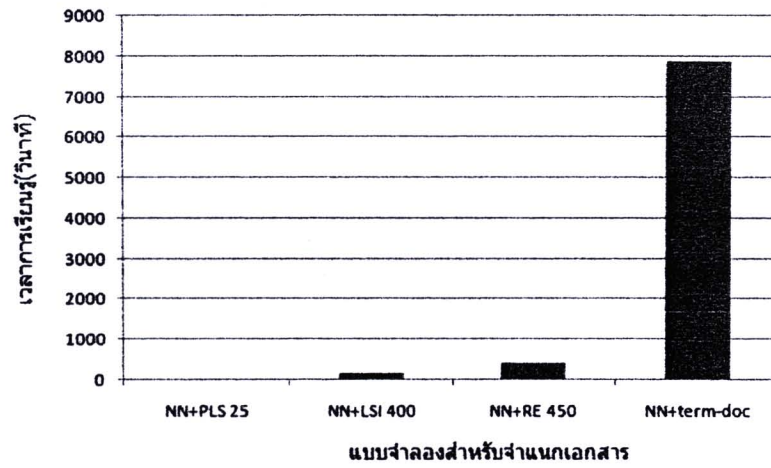
จำนวนมิติ	เวลาการเรียนรู้ NN+RE	จำนวนมิติ	เวลาการเรียนรู้ NN+RE
15	2.66	500	404.78
20	3.05		

5.5 วิจารณ์และสรุปผล

จากการทดลองการสร้างแบบจำลองการจำแนกเอกสารด้วยโครงข่ายประสาทที่มีข้อมูลเข้าสร้างจากการสร้างตัวแทนเอกสารแบบต่างๆ พบว่าการสร้างตัวแทนเอกสารในกลุ่มของ การสร้างพื้นที่ความหมายประกอบด้วย แอลเอสไอ (NN+LSI) โดยมีมิติที่ใช้สร้างตัวแทนเอกสารอยู่ในช่วงระหว่าง 380-450 มิติ และ พีแอลเอส (NN+PLS) โดยมีมิติที่ใช้สร้างตัวแทนเอกสารอยู่ในช่วงระหว่าง 23-25 มิติ ให้ความถูกต้องสูงสุดและใช้เวลาการเรียนรู้ที่น้อยที่สุด รองลงมาคือการคัดเลือกลักษณะเด่นโดยวิธีรีเลแวนซ์สกออร์ โดยเลือกคำสำคัญที่มีคะแนนสูงสุดในช่วงระหว่าง 400-500 คำ และสุดท้ายคือการสร้างเมตริกซ์คำสำคัญ-เอกสารที่ได้จากการใช้คำสำคัญของเอกสารสำหรับเรียนรู้ทั้งหมด โดยมีจำนวนคำสำคัญ 3586 คำ



รูป 5.5 แสดงการเปรียบเทียบความถูกต้องสูงสุดของการจำแนกเอกสารแต่ละแบบ



รูป 5.6 แสดงการเปรียบเทียบเวลาการเรียนรู้ของการจำแนกเอกสารแต่ละแบบ

จากรูป 5.5 และ 5.6 ตัวเลขที่ค่อท้ายแบบจำลองสำหรับจำแนกเอกสารคือจำนวนมิติของตัวแทนเอกสารเช่น NN+PLS 25 คือ แบบจำลองโครงข่ายประสาทที่มีข้อมูลเข้าสร้างจากวิธีพีแอลเอส 25 มิติ