

บทที่ 6

สรุปผลการวิจัย

ในบทนี้กล่าวถึง การสรุปผลการทดลองในการจำแนกเอกสารโดยใช้เทคนิคการลดมิติเอกสารเพื่อเป็นข้อมูลเข้าให้แก่โครงข่ายประสาท เนื้อหาประกอบด้วยบทสรุป ปัญหาและอุปสรรค และข้อเสนอแนะการทำงานวิจัย

6.1 บทสรุป

จากการเปรียบเทียบการลดมิติเอกสารทั้งสามวิธีประกอบด้วย การสร้างตัวแทนเอกสารด้วยวิธีแอลเอสไอ พีแอลเอส และการเลือกคำสำคัญบางคำมาใช้ในการสร้างดัชนีเอกสารด้วยวิธีเลแวนชีสเกอร์ เพื่อสร้างข้อมูลเข้าให้แก่โครงข่ายประสาท โดยใช้ข้อมูลชีวสารสนเทศเป็นข้อมูลตัวอย่างในการทดสอบ พบว่าการลดมิติเอกสารทั้งสามแบบช่วยปรับปรุงโครงข่ายประสาทให้สามารถจำแนกเอกสารได้ถูกต้องมากยิ่งขึ้น และช่วยลดเวลาการเรียนรู้ เมื่อเปรียบเทียบกับการสร้างดัชนีเอกสารในรูปของเมตริกซ์คำสำคัญ-เอกสาร เนื่องจาก การลดมิติทั้งสามวิธีสามารถลดปริมาณคำสำคัญที่ไม่แสดงความแตกต่างระหว่างกลุ่ม ซึ่งอาจส่งผลให้ประสิทธิภาพการจำแนกเอกสารลดลง

เมื่อพิจารณาการลดมิติทั้งสามวิธีที่ได้ทำการศึกษาเปรียบเทียบ พบว่ากลุ่มของการสร้างพื้นที่ความหมายในการลดมิติคือ พีแอลเอส และแอลเอสไอ สามารถช่วยปรับปรุงความถูกต้องในการจำแนกเอกสารได้ดีใกล้เคียงกัน และสูงกว่าวิธีการเลือกคำสำคัญด้วยวิธีรีเลแวนชีสเกอร์ เนื่องจากวิธีการเลือกคำสำคัญนั้นมีประสิทธิภาพที่ด้อยกว่า อาจเกิดจากปัญหาของการเลือกคำสำคัญ เช่น บางเอกสารอาจอยู่ในกลุ่มใดกลุ่มหนึ่งแต่ไม่ได้มีคำสำคัญที่เลือกมาที่บ่งบอกว่าเป็นสมาชิกของกลุ่มนั้นเลย จึงทำให้แบบจำลองในการจำแนกทำนายว่าไม่ได้เป็นสมาชิกของกลุ่มนั้น ซึ่งแตกต่างจากกลุ่มการลดมิติโดยการสร้างพื้นที่ความหมายทั้งสองวิธี ที่สร้างความสัมพันธ์ระหว่างคำสำคัญทั้งหมด เพื่อสร้างความแตกต่างระหว่างกลุ่ม โดยการลดมิติเอกสารทั้งสามวิธีมีปัญหาเหมือนกันคือ การเลือกจำนวนมิติที่เหมาะสม ในการหาความแตกต่างระหว่างกลุ่ม หากเลือกมิติน้อยเกินไป ก็จะแสดงความแตกต่างไม่ครบทุกกลุ่ม และหากเลือกมิตินมากเกินไป ส่งผลให้การจำแนกเอกสารมีความถูกต้องลดลงได้เช่นกัน มิติเอกสารที่ส่งผลให้การลดมิติเอกสารทำให้โครงข่ายประสาทสามารถจำแนกเอกสารได้ถูกต้องสูงขึ้นในแต่ละวิธีคือ พีแอลเอส ใช้ 23-25 มิติ แอลเอสไอ ใช้ 380-450 มิติ และรีเลแวนชีสเกอร์ใช้คำสำคัญ 400-500 คำ โดยที่พีแอลเอส ใช้จำนวน

มิติในการสร้างตัวแทนเอกสารน้อยกว่าวีแอลเอสไอ เนื่องจากพีแอลเอสไอการใช้ข้อมูลของกลุ่มของเอกสารเข้ามาใช้ในการสร้างพื้นที่ความหมายด้วย ซึ่งแตกต่างจากแอลเอสไอ โดยแอลเอสไอสร้างพื้นที่ความหมายโดยไม่ใช้ข้อมูลของกลุ่มเข้ามาเกี่ยวข้อง

ในงานของ Zeng (Zeng, Wang and Nie, 2007) ที่ใช้กระบวนการพีแอลเอสในการลดมิติในการจำแนกเอกสาร พบพีแอลเอสช่วยปรับปรุงความถูกต้องในการจำแนกเอกสารที่สูงกว่าแอลเอสไอ แต่ในงานวิจัยนี้พบว่าแอลเอสไอช่วยให้การจำแนกสูงกว่าพีแอลเอสเล็กน้อย แต่ไม่ได้สร้างความแตกต่างที่ชัดเจนนัก และการทดลองนี้ใช้จำนวนมิติของพีแอลเอสในการสร้างตัวแทนเอกสารที่น้อยกว่า ซึ่งอาจเกิดจากความแตกต่างของจำนวนกลุ่มข้อมูลที่ใช้ และ ข้อมูลที่ใช้

ด้านเวลาการเรียนรู้ของโครงข่ายประสาทพบว่าการลดมิติทั้งสามวิธีช่วยลดเวลาการเรียนรู้ได้เนื่องจากทั้งสามวิธีช่วยลดปริมาณข้อมูลเข้า และช่วยลดความซับซ้อนของข้อมูลทำให้รอบการเรียนรู้ลดลง โดยพีแอลเอสช่วยลดเวลาการเรียนรู้ดีที่สุด เนื่องจากใช้มิติในการสร้างตัวแทนเอกสารน้อยที่สุดและมีการใช้ข้อมูลของกลุ่มเอกสารสร้างพื้นที่ความหมาย รองลงมาคือแอลเอสไอ แม้ว่าแอลเอสไอจะใช้จำนวนมิติที่ใกล้เคียงกับรีเลแวนซ์สก็อร์ แต่แอลเอสไอมีประสิทธิภาพในการหาความแตกต่างระหว่างกลุ่มที่ดีกว่ารีเลแวนซ์สก็อร์ จึงช่วยลดเวลาการเรียนรู้ของโครงข่ายประสาทได้ดีกว่า สุดท้ายคือรีเลแวนซ์สก็อร์

ขั้นตอนวิธีการหาพีแอลเอสเดิม ในการปรับค่าของ ξ จากขั้นตอนวิธีเดิมใช้การปรับค่าไปเรื่อยๆจนค่าไม่เปลี่ยนแปลง ในการทดลองพบว่าการปรับค่าจนค่าไม่เปลี่ยนแปลงนั้นต้องใช้จำนวนรอบการปรับค่าที่มาก และถ้าข้อมูลมีขนาดใหญ่ขึ้นการปรับค่าก็จะต้องใช้จำนวนรอบเพิ่มขึ้นด้วย งานวิจัยนี้จึงได้เสนอการหาจำนวนรอบที่เหมาะสมโดยการทดลอง พบว่าในรอบต้นๆในการปรับค่าค่าของ ξ จะแตกต่างกันมาก แต่หลังจากนั้นจะมีความแตกต่างที่ลดลงมาก โดยพิจารณาจากค่าผิดพลาดเฉลี่ยกำลังสองพบว่าตั้งแต่รอบที่ 4 เป็นต้นไปค่าความผิดพลาดจะเข้าใกล้ค่า 0 และคงระดับนั้นไปเรื่อยๆ และจากการทดลองนี้พบว่ารอบที่ดีที่สุดในการปรับค่าคือ 7 รอบ จากการทดสอบการจำแนกโดยการสร้างตัวแทนเอกสาร 25 มิติ และทดสอบปรับค่า ξ ระหว่าง 1-100 รอบ

ในด้านโครงสร้างของโครงข่าย พบว่า ถ้ามีการเพิ่มจำนวนโหนดในชั้นกลางจะส่งผลให้การจำแนกเอกสารมีความถูกต้องลดลงในการลดมิติทุกแบบรวมทั้งการจำแนกบนดัชนีเอกสารโดยตรงด้วย ที่เป็นเช่นนี้อาจเกิดจากข้อมูลมีความซับซ้อนที่ต่ำ สามารถสังเกตได้จากการจำแนกบนดัชนีเอกสารโดยตรงให้ความถูกต้องค่อนข้างสูง ทั้งที่ยังไม่ได้มีการลดความซับซ้อนของข้อมูลเลย ดังนั้นการหาโครงสร้างของโครงข่ายที่เหมาะสม ขึ้นอยู่กับกลุ่มข้อมูลที่ใช้ในการจำแนก และรวมไปถึงวิธีการลดมิติและจำนวนมิติที่เลือกในการสร้างตัวแทนเอกสารสำหรับใช้ในการจำแนกด้วย

6.2 ปัญหาและอุปสรรค

จากการทำงานวิจัยนี้พบปัญหาและอุปสรรคคือ

- 1) การเลือกมิติในการสร้างตัวแทนเอกสารนั้นได้จากการทดลองเลือกมิติ ซึ่งต้องทำการเลือกมิติหลายครั้งจนกว่าจะได้มิติที่ให้ผลลัพธ์ที่ดี
- 2) การออกแบบโครงข่ายประสาทได้จากการทดลองซึ่งเป็นผลสืบเนื่องมาจากหัวข้อ 1) และค่าพารามิเตอร์ต่างๆที่กำหนดได้จากการประมาณค่าเพื่อใช้อ้างอิงในการเปรียบเทียบแบบจำลองแต่ละแบบ

6.3 ข้อเสนอแนะการทำงานวิจัย

กระบวนการการปรับค่า ζ_k ในพีแอลเอส พบว่ารอบการปรับค่าที่ดีที่สุดในรอบที่ 7 มีความใกล้เคียงกับค่าผิดพลาดเฉลี่ยกำลังสอง ที่ลดลงเข้าใกล้ 0 ในรอบที่ 4 ดังนั้นควรมีการทดลองเพิ่มเติมกับข้อมูลอื่นๆ ในการหาจุดสิ้นสุดของการปรับค่า ζ_k โดยใช้ค่าผิดพลาดเฉลี่ยกำลังสองที่เริ่มเข้าใกล้ 0 ถ้าสามารถใช้ค่าผิดพลาดเฉลี่ยกำลังสอง ในการหาจุดสิ้นสุดในการปรับค่าได้ ก็จะสามารถใช้เป็นจุดสิ้นสุดในการปรับค่าในข้อมูลอื่นๆ ซึ่งส่งผลดีกว่าการหาขอบการปรับค่าเองโดยผู้ใช้ และ นอกจากนี้ควรมีการเปรียบเทียบประสิทธิภาพการจำแนกเอกสาร โดยโครงข่ายประสาทกับแบบจำลองการจำแนกอื่นๆด้วย เพื่อทดสอบการลดมิติทั้งสามแบบที่ได้นำมาเปรียบเทียบกัน มีผลกระทบกับแบบจำลองการจำแนกแบบอื่นๆด้วยหรือไม่