บทที่ 5
สรุปผลการวิจัย และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

1) นิวรอลเน็ตเวอร์คแบบ MLP มีความสามารถในการเรียนรู้สูงมาก สังเกตได้จากอัตราการรู้จำที่ได้เมื่อใช้ชุดฝึกฝนมาทดสอบ ไม่ว่าจำนวนผู้พูดในชุดฝึกฝนจะเพิ่มขึ้นมากก็ตาม ลักษณะเช่นนี้ทำให้นิวรอลเน็ตเวอร์คที่ผ่านการฝึกฝนมาแล้ว มีความสามารถในการรู้จำเสียงพูดที่หลากหลายได้ โดยมีความสามารถในการรู้จำแบบขึ้นต่อผู้พูด ได้ดีกว่าแบบไม่ขึ้นต่อผู้พูด และอัตราการรู้จำจะเพิ่มขึ้นเมื่อเพิ่มจำนวนผู้พูดที่ใช้ฝึกฝน

2) นิวรอลเน็ตเวอร์คแบบ MLP จะมีความสามารถในการรู้จำได้ดีขึ้นเมื่อจำนวนคำศัพท์ที่จะรู้จำมีจำนวนลดลง ดังนั้น ในกรณีที่มีคำศัพท์ที่จะรู้จำจำนวนมาก การแบ่งกลุ่มคำศัพท์เบื้องต้นเพื่อลดจำนวนคำศัพท์ที่จะรู้จำในแต่ละกลุ่ม จะช่วยเพิ่มอัตราการรู้จำให้สูงขึ้นได้มาก

3) เทคนิคแบบฟัซซีซึ่งจะครอบคลุมปัญหาเรื่องความกำกวมในการรู้จำ มีความเหมาะสมที่จะนำมาใช้กับการรู้จำเสียงซึ่งมีลักษณะกำกวมอยู่เสมอ ซึ่งจะให้อัตราการรู้จำสูงขึ้น สำหรับในงานวิจัยนี้ ใช้เทคนิคแบบฟัซซีในการปรับข้อมูลเข้า และข้อมูลออกที่ต้องการของนิวรอลเน็ตเวอร์ค

4) การใช้ข้อมูลออกที่ต้องการแบบค่าสมาชิกภาพของแต่ละคำศัพท์ กับข้อมูลเข้าที่ใช้ในงานวิจัยนี้ จะส่งผลให้อัตราการรู้จำตกลงอย่างมากเมื่อเทียบกับการใช้ข้อมูลออกที่ต้องการแบบเลขฐานสอง เนื่องจากจะสร้างความกำกวมในการฝึกฝนมากเกินไป รวมทั้งมีกรณีที่ฝึกฝนนิวรอลเน็ตเวอร์คแบบผิดๆ ทำให้การรู้จำผิดพลาดมาก

5) งานวิจัยนี้อาศัยเสียงพูดตัวอย่างในชุดฝึกฝนในการฝึกฝนระบบรู้จำ การทดสอบด้วยเสียงจากผู้พูด กลุ่มเดียวกับผู้พูดในชุดฝึกฝน ถือได้ว่าเป็นการการรู้จำแบบขึ้นกับผู้พูด ส่วนการทดสอบด้วยเสียงจากผู้พูดคนละกลุ่มกับชุดฝึกฝนจะถือได้ว่าเป็นการรู้จำแบบไม่ขึ้นกับผู้พูด ซึ่งเป็นจุดมุ่งหมายของงานวิจัย

6) การใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซี ร่วมกับข้อมูลออกที่ต้องการแบบเลขฐานสอง จะให้อัตราการรู้จำสูงกว่าการใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC และการใช้ฟังก์ชันสมาชิกภาพชนิด Pi จะให้อัตราการรู้จำสูงที่สุด รองลงมาคือ ชนิดสี่เหลี่ยมคางหมู และชนิดสามเหลี่ยมตามลำดับ โดยอัตราการรู้จำเฉลี่ยของชุดทดสอบแบบไม่ขึ้นต่อผู้พูดสำหรับกลุ่มคำศัพท์ตัวเลข 0-9 ในกรณีที่ใช้ฟังก์ชันสมาชิกภาพชนิด Pi, สี่เหลี่ยมคางหมู และสามเหลี่ยม จะสูงกว่ากรณีที่ใช้ค่าสัมประสิทธิ์ LPC 7.22, 6.74 และ 2.90 เปอร์เซ็นต์ ตามลำดับ

6) การใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC และใช้ผู้พูดจำนวน 50 คนในการฝึกฝน สำหรับกลุ่มคำศัพท์ที่ผ่านการแบ่งกลุ่มเบื้องต้นด้วยการตรวจสอบจำนวนพยางค์ จะให้อัตราการรู้จำเฉลี่ยของชุดทดสอบแบบขึ้นต่อผู้พูด และแบบไม่ขึ้นต่อผู้พูดเท่ากับ 83.5 และ 78.3 เปอร์เซ็นต์ ตามลำดับ แต่เมื่อทำการแบ่งกลุ่มมากขึ้นด้วยการตรวจสอบเสียงวรรณยุกต์ของพยางค์แรกของคำ จะได้อัตราการรู้จำเฉลี่ยของชุดทดสอบแบบขึ้นต่อผู้พูด และแบบไม่ขึ้นต่อผู้พูดเท่ากับ 91.1 และ 89.9 เปอร์เซ็นต์ ตามลำดับ

7) เมื่อใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด Pi ร่วมกับข้อมูลออกที่ต้องการแบบเลขฐานสอง และใช้ผู้พูดจำนวน 50 คนในการฝึกฝนสำหรับกลุ่มคำศัพท์ที่ผ่านการแบ่งกลุ่มเบื้องต้นด้วยการตรวจสอบจำนวนพยางค์ จะได้อัตราการรู้จำเฉลี่ยของชุดทดสอบแบบขึ้นต่อผู้พูด และแบบไม่ขึ้นต่อผู้พูดเท่ากับ 88.6 และ 86.2 เปอร์เซ็นต์ ซึ่งสูงขึ้นจากกรณีที่ใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC 5.1 และ 7.9 เปอร์เซ็นต์ ตามลำดับ แต่เมื่อทำการแบ่งกลุ่มมากขึ้นด้วยการตรวจสอบเสียงวรรณยุกต์ของพยางค์แรกของคำ จะได้อัตราการรู้จำเฉลี่ยของชุดทดสอบแบบขึ้นต่อผู้พูด และแบบไม่ขึ้นต่อผู้พูดเท่ากับ 94.4 และ 93.3 เปอร์เซ็นต์ ซึ่งสูงขึ้นจากกรณีที่ใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC 3.3 และ 3.4 เปอร์เซ็นต์ ตามลำดับ

5.2 ปัญหา และข้อเสนอแนะ

1) เนื่องจากยังไม่เคยมีข้อกำหนดของกรรมวิธีในการเก็บข้อมูลเสียง ทั้งอุปกรณ์ที่ใช้ ผู้พูด และหลักการพูดเสียงคำภาษาไทยในแบบต่างๆ ที่เป็นมาตรฐาน ทำให้การเก็บข้อมูลอยู่ในลักษณะที่จำกัดอยู่ในขอบเขตที่เป็นจุดมุ่งหมายของการวิจัย เพื่อให้ระบบสามารถรู้จำคำได้ดีที่สุด จำเป็นจะต้องมีการเก็บข้อมูลที่เป็นมาตรฐาน และครอบคลุมคำศัพท์มากพอเพียง เพื่อนำมาใช้ในการฝึกฝน

2) โปรแกรมที่ใช้ในการตัดหัวท้ายหน่วยที่ใช้ในการรู้จำ หากว่ามีการนำหลักทางภาษาศาสตร์เข้ามาช่วยพิจารณาช่วงของเสียงพูดที่ควรจะเป็น และสร้างโปรแกรมในการตัดหัวท้ายหน่วยที่ใช้ในการรู้จำได้ดีขึ้น รวมทั้งการเพิ่มความสามารถในการตัดแบ่งเสียง จะสามารถสร้างระบบการรู้จำที่ใช้หน่วยในการรู้จำที่เล็กลงไปได้ เช่น พยางค์ ซึ่งจะลดเวลาในการประมวลผล และช่วยเพิ่มอัตราการรู้จำได้

3) การนอร์มอลไลซ์ทางเวลาทำให้รูปคลื่นผิดเพี้ยนไป ดังนั้นควรหาวิธีการที่ไม่ต้องมีการนอร์มอลไลซ์ หรือมีการนอร์มอลไลซ์ทางเวลาด้วยจำนวนจุดข้อมูลที่ใกล้เคียงกับจำนวนจุดที่แท้จริงของคำนั้นๆ ให้มากที่สุด วิธีการหนึ่งคือ การแบ่งกลุ่มคำศัพท์โดยพิจารณาระยะเวลาของเสียงร่วมด้วย จะทำให้ได้กลุ่มคำศัพท์ที่ประกอบด้วยคำศัพท์ที่มีช่วงเวลาใกล้เคียงกัน และทำการนอร์มอลไลซ์ทางเวลาแยกแต่ละกลุ่ม

4) ขั้นตอนก่อนการฝึกฝน จะต้องมีการเตรียมข้อมูลเข้า และข้อมูลออกเป็นคู่ๆ ซึ่งเสียเวลาในการเตรียม จึงควรมีการเขียนโปรแกรมให้สามารถเรียกข้อมูลเข้า และสร้างข้อมูลออกที่ต้องการได้อย่างอัตโนมัติ นอกจากนี้ การฝึกฝนนิวรอลเน็ตเวอร์กต้องใช้เวลานานมากถ้าใช้จำนวนผู้พูดในการฝึกฝนจำนวนมาก โดยเฉพาะอย่างยิ่งในกรณีที่ใช้เทคนิคแบบฟัซซีร่วมด้วย เนื่องจากเป็นข้อมูลที่พยายามครอบคลุมปัญหาเรื่องความกำกวม และยังมีจำนวนข้อมูลเข้าหลายเท่าของข้อมูลเข้าปกติ การแก้ไขวิธีหนึ่งคือ การพัฒนาระบบรู้จำให้มีความสามารถในการปรับเพิ่มคำศัพท์ หรือผู้พูดภายหลังได้ หรืออาจจะหาวิธีการปรับปรุงการนำเทคนิคแบบฟัซซีที่นำมาใช้

6) บางกรณีที่การฝึกฝนนิวรอลเน็ตเวอร์คนานมากเกินไป หรือฝึกฝนไม่สำเร็จ ควรมีการพัฒนาโปรแกรมเพื่อแสดงค่าพารามิเตอร์ที่เป็นเงื่อนไขในการหยุดการฝึกฝน หรือให้แสดงเส้นทางในการปรับค่าน้ำหนักการเชื่อมต่อ เพื่อใช้ในการพิจารณาแนวทางการฝึกฝนว่าจะสำเร็จหรือไม่

7) การปรับปรุงวิธีการสกัดลักษณะเด่นของสัญญาณเสียงเป็นขั้นตอนที่สำคัญมาก หากสามารถสกัดลักษณะเด่นที่ช่วยแยกแยะได้ดีกว่านี้ และมีจำนวนน้อยกว่านี้ได้ จะช่วยเพิ่มอัตราการรู้จำ และน่าจะใช้เทคนิคในการใช้ข้อมูลออกที่ต้องการแบบค่าสมาชิกภาพของแต่ละคำศัพท์ เข้ามาช่วยได้ดีขึ้น ซึ่งวิธีการสกัดลักษณะเด่นที่เหมาะสมสำหรับคำภาษาไทยอยู่ในขั้นตอนการวิจัย

8) การแบ่งกลุ่มเบื้องต้นเป็นขั้นตอนที่สำคัญเช่นกัน เพราะถ้ามีการแบ่งกลุ่มผิดพลาดจะทำให้รู้จำผิดไปทันที ดังนั้นวิธีการที่ใช้ในแบ่งกลุ่มเบื้องต้นจึงต้องมีการวิจัยเพื่อให้ได้ประสิทธิภาพสูงที่สุด ซึ่งขณะนี้ ยังอยู่ในขั้นตอนของการวิจัย

9) ระบบที่พัฒนาขึ้นนี้ เป็นการใช้คำทั้งคำเป็นตัวทาบเทียบ ซึ่งเหมาะสมกับจำนวนคำศัพท์ไม่มากนัก หากต้องการจะเพิ่มจำนวนคำศัพท์ให้มากขึ้น ควรจะขยายไปใช้หน่วยเสียงที่ย่อยกว่านี้เป็นตัวทาบเทียบ ซึ่งยังคงสามารถใช้เทคนิคที่พัฒนาขึ้นในงานวิจัยนี้ ไปช่วยในการรู้จำได้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย