บทที่ 4
ผลการวิจัย วิเคราะห์และสรุปผลการวิจัย

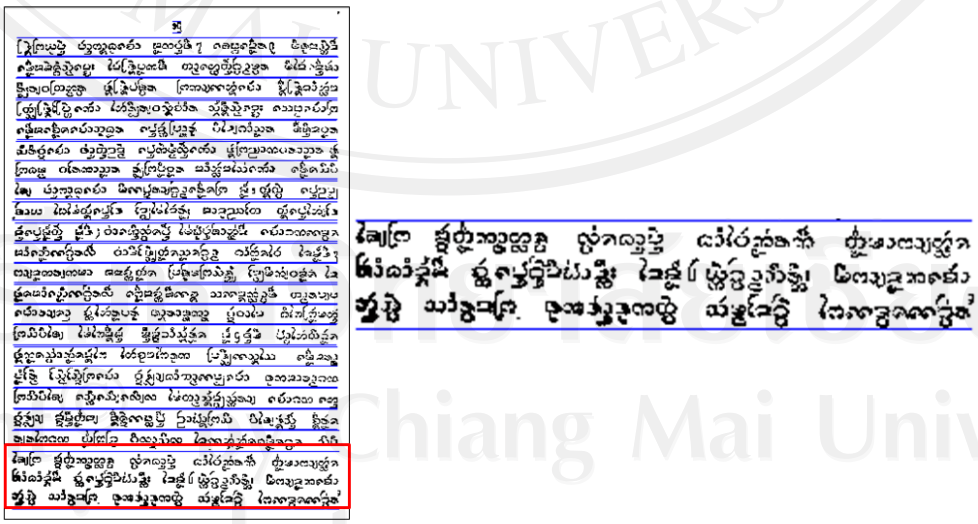
ในบทนี้จะกล่าวถึงรายละเอียดของการทดสอบประสิทธิภาพการรู้จำตัวพิมพ์อักษรธรรมล้านนา ผลการวิจัย รวมถึงการวิเคราะห์และสรุปผลการวิจัย โดยมีรายละเอียดดังต่อไปนี้

4.1 การทดสอบการตัดแบ่งบรรทัด

การทดสอบประสิทธิภาพการตัดแบ่งบรรทัด ใช้วิธีการร้อยละความถูกต้องของการตัดบรรทัด โดยสุ่มภาพเอกสาร จำนวน 20 ภาพ แต่ละภาพเอกสารถูกปรับปรุงคุณภาพ และตัดแบ่งเป็นบรรทัดตามกระบวนการที่ได้นำเสนอ แล้วพิจารณาผลการตัดบรรทัดของเอกสารทั้ง 20 ภาพ จากนั้นคำนวณหาความถูกต้องของการตัดแบ่งบรรทัดดังสมการ (4.1)

$$\text{ความถูกต้องของการตัดบรรทัด} = \frac{\text{จำนวนบรรทัดที่ตัดแบ่งได้ถูกต้อง}}{\text{จำนวนบรรทัดที่ปรากฏทั้งหมดในเอกสาร}} \times 100 \qquad (4.1)$$

ผลการทดสอบการตัดแบ่งบรรทัด มีอัตราการตัดแบ่งบรรทัดได้ถูกต้องสูงถึงร้อยละ 97.14 เนื่องจากได้มีการปรับความเอียงของหน้าเอกสารให้ตรง เพื่อให้ง่ายต่อการตัดแบ่งบรรทัดโดยใช้เส้นตรง แต่อย่างไรก็ตามในส่วนที่ตัดแบ่งบรรทัดผิดพลาดนั้น ส่วนใหญ่เกิดจากภาพอักษรเอียงบางบรรทัด ไม่ได้เอียงทั้งหน้าเอกสาร อีกทั้งมีตัวอักษรเหลื่อมกันระหว่างบรรทัด ทำให้ไม่สามารถแยกบรรทัดออกจากกันได้ จึงส่งผลให้การตัดแบ่งบรรทัดไม่ถูกต้อง ดังรูป 4.1



(ก) (ข)

รูป 4.1 ผลการตัดแบ่งบรรทัดที่ผิดพลาด

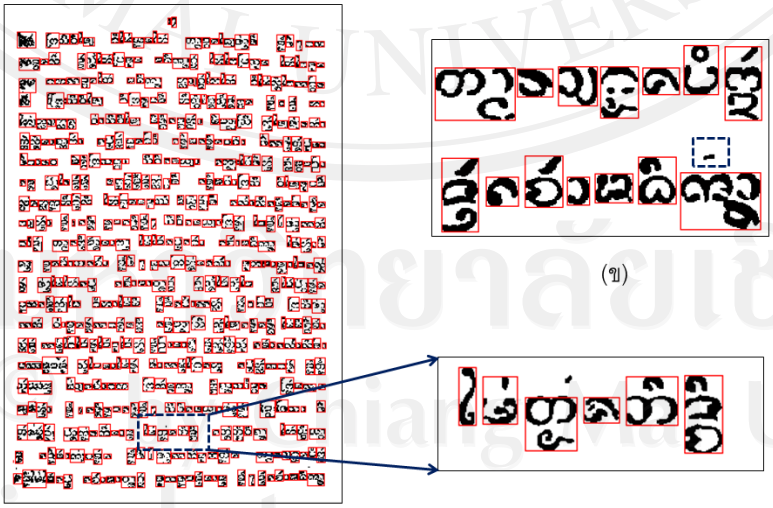
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

4.2 การทดสอบการแบ่งบล็อกตัวอักษร

การทดสอบตัดแบ่งบล็อกตัวอักษร ใช้วิธีการร้อยละของความถูกต้องของการตัดแบ่งบล็อกตัวอักษร โดยเลือกภาพบรรทัดเอกสารอย่างสุ่ม จำนวน 100 บรรทัด จากนั้นนำมาทดสอบตัดแบ่งบล็อก แล้วคำนวณหาความถูกต้องของการตัดแบ่งบล็อกตัวอักษรดังสมการ (4.2)

$$\text{ความถูกต้องของการแบ่งบล็อกตัวอักษร} = \frac{\text{จำนวนบล็อกที่ตัดแบ่งได้ถูกต้อง}}{\text{จำนวนบล็อกที่ปรากฏทั้งหมด}} \times 100 \qquad (4.2)$$

ผลการทดสอบการตัดแบ่งบล็อกตัวอักษร มีอัตราถูกต้องของการตัดแบ่งบล็อกตัวอักษรสูงถึงร้อยละ 98.63 แต่ยังมีเอกสารบางลักษณะที่ทำให้เกิดความผิดพลาดในการตัดแบ่งบล็อกตัวอักษร โดยส่วนใหญ่แล้วเกิดจากการที่รูปภาพนั้นมีลายเส้นขาดหาย กรณีที่เป็นอักษรระดับบนหรืออักษรที่หางลากยาวขึ้นไปด้านบน ที่มีบางส่วนของตัวอักษรแยกขาดจากตัวอักษรมากเกินไป ทำให้อักษรที่ขาดนั้นไม่ได้ถูกพิจารณาให้รวมอยู่ในบล็อก ดังรูป 4.2 (ก) ไม้ขอช้าง (ั) เหนือบล็อกอักษร ๐ มีเส้นขาดหายไป ทำให้เหลือชิ้นส่วนอักษรขนาดเล็กและอยู่แยกจากอักษรตัวอื่นมาก ทำให้ไม่ถูกรวมอยู่ภายในบล็อกอักษร ซึ่งสาเหตุการกำหนดให้ตัวอักษรที่อยู่แยกออกไปจากอักษรตัวอื่นมาก ไม่ถูกรวมให้อยู่ภายในบล็อก เนื่องจากชิ้นส่วนที่แยกจากอักษรตัวอื่นมาก อาจเป็นสัญญาณรบกวนที่มีขนาดใหญ่ ไม่สามารถกำจัดออกได้ในขั้นตอนการประมวลผลภาพเบื้องต้น ซึ่งอาจหลงเหลืออยู่ภายในภาพ ดังรูป 4.2 (ข) ชิ้นส่วนเหนือบล็อกตัวอักษร ๓๒ ไม่ถูกรวมให้อยู่ภายในบล็อกตัวอักษร เนื่องจากชิ้นส่วนนั้นเป็นสัญญาณรบกวน



(ก) (ข)

รูป 4.2 ผลการตัดแบ่งบล็อก

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

4.3 การทดสอบการรู้จำตัวอักษรธรรมล้านนา

การรู้จำตัวพิมพ์อักษรธรรมล้านนา ได้เลือกหนังสือเฉพาะหมวดคร่าวซอ เพื่อจำกัดขอบเขตในการสร้างชุดข้อมูลฝึกฝนสำหรับการจำแนกประเภทตัวอักษร

การทดสอบประสิทธิภาพการรู้จำตัวพิมพ์อักษรธรรมล้านนา วัดจากค่าความแม่นยำ (precision) ความถูกต้อง (recall) และค่า F-measure ดังสมการ (4.3) (4.4) และ (4.5) ตามลำดับ

$$\text{ความแม่นยำ} = \frac{\text{จำนวนตัวอักษรที่จำแนกได้ถูกต้อง}}{\text{จำนวนตัวอักษรที่จำแนกได้ทั้งหมด}} \quad (4.3)$$

$$\text{ความถูกต้อง} = \frac{\text{จำนวนตัวอักษรที่จำแนกได้ถูกต้อง}}{\text{จำนวนตัวอักษรทั้งหมดที่ปรากฏในหน้าเอกสาร}} \quad (4.4)$$

$$F-Measure = 2 \times \frac{\text{ความแม่นยำ} \times \text{ความถูกต้อง}}{\text{ความแม่นยำ} + \text{ความถูกต้อง}} \quad (4.5)$$

4.3.1 การทดสอบเพื่อเปรียบเทียบประสิทธิภาพการรู้จำระหว่างเค-เนียร์เรสเนเบอร์ และซัพพอร์ตเวกเตอร์แมชชีน

สำหรับการทดสอบการรู้จำตัวพิมพ์อักษรธรรมล้านนาด้วยเค-เนียร์เรสเนเบอร์ กำหนดให้ค่า k เท่ากับ 1 และการทดสอบการรู้จำตัวพิมพ์อักษรธรรมล้านนาด้วยซัพพอร์ตเวกเตอร์แมชชีน ใช้ไลบรารี LIBSVM โดยใช้เคอร์แนล เรเดียลเบสิสฟังก์ชัน กำหนดให้ Gamma เท่ากับ 1/จำนวนคุณลักษณะ และ Cost เท่ากับ 1

ชุดข้อมูลฝึกสอนทั้งหมด 1,108 ประเภท ดังตาราง 3.3 จำนวน 20,917 ตัวอย่าง สำหรับชุดข้อมูลทดสอบทั้งหมดที่ผ่านขั้นตอนประมวลผลภาพเบื้องต้นและการตัดแบ่ง ได้บล็อกจำนวน 4,722 บล็อก หรือ 9,656 ตัวอักษร และถูกเปลี่ยนขนาดให้มีขนาดเท่ากัน คือกว้าง 44 สูง 44 เพื่อสร้างเวกเตอร์คุณลักษณะที่มีความยาว 1,936 ซึ่งผลการทดสอบแสดงดังตาราง 4.1

ตาราง 4.1 ผลการทดสอบการรู้จำระหว่างเค-เนียร์เรสเนเบอร์ และซัพพอร์ตเวกเตอร์แมชชีน

วิธีการที่ใช้	ค่าเฉลี่ยความแม่นยำ	ค่าเฉลี่ยความถูกต้อง	F-Measure
เค-เนียร์เรสเนเบอร์	0.9190	0.9001	0.9095
ซัพพอร์ตเวกเตอร์แมชชีน	0.8006	0.7799	0.7901

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ตาราง 4.2 จำนวนบล็อกที่จำแนกผิดด้วยวิธีเค-เนียร์เรสเนเบอร์ที่ขึ้นกับระยะทางระหว่างบล็อกกับข้อมูลฝึกสอน

ระยะทางระหว่างบล็อกกับข้อมูลฝึกสอน (d)	จำนวนบล็อกที่จำแนกผิดพลาด	ระยะทางระหว่างบล็อกกับข้อมูลฝึกสอน (d)	จำนวนบล็อกที่จำแนกผิดพลาด
$d < 0.04$	0	$0.13 \leq d < 0.14$	4
$0.04 \leq d < 0.05$	1	$0.14 \leq d < 0.15$	3
$0.05 \leq d < 0.06$	13	$0.15 \leq d < 0.16$	7
$0.06 \leq d < 0.07$	32	$0.16 \leq d < 0.17$	9
$0.07 \leq d < 0.08$	13	$0.17 \leq d < 0.18$	9
$0.08 \leq d < 0.09$	8	$0.18 \leq d < 0.19$	14
$0.09 \leq d < 0.10$	7	$0.19 \leq d < 0.20$	9
$0.10 \leq d < 0.11$	13	$0.20 \leq d < 0.21$	13
$0.11 \leq d < 0.12$	4	$d \geq 0.21$	148
$0.12 \leq d < 0.13$	5		

จากการผลการทดสอบการเปรียบเทียบประสิทธิภาพการรู้จำระหว่างเค-เนียร์เรสเนเบอร์ และซัพพอร์ตเวกเตอร์แมชชีน ดังตาราง 4.1 เค-เนียร์เรสเนเบอร์มีประสิทธิภาพในการรู้จำตัวพิมพ์อักษรธรรมล้านนาสูงกว่าซัพพอร์ตเวกเตอร์แมชชีน โดยวัดได้จากค่า F-Measure ของการรู้จำด้วยเค-เนียร์เรสเนเบอร์ที่มีค่าเท่ากับ 0.9095 ซึ่งสูงกว่าการรู้จำด้วยซัพพอร์ตเวกเตอร์แมชชีนที่มีค่าเพียง 0.7901 ซึ่งงานวิจัยส่วนใหญ่มีผลการทดสอบการจำแนกด้วยซัพพอร์ตเวกเตอร์แมชชีนที่มีประสิทธิภาพดีกว่าเค-เนียร์เรสเนเบอร์ แต่ในงานวิจัยนี้ให้ผลตรงกันข้าม อาจมีสาเหตุมาจากประเภทตัวอักษรที่ใช้ในการจำแนกมีจำนวนมาก และคุณลักษณะที่ใช้อาจไม่เหมาะสม รวมทั้งชุดข้อมูลฝึกสอนบางประเภทมีเพียงหนึ่งตัวอย่าง ทำให้การใช้เค-เนียร์เรสเนเบอร์จำแนกตัวอักษรได้ถูกต้องกว่า

จากตาราง 4.2 เป็นตารางแสดงจำนวนบล็อกที่จำแนกผิดด้วยวิธีเค-เนียร์เรสเนเบอร์ที่ขึ้นกับระยะทางระหว่างบล็อกกับข้อมูลฝึกสอน ซึ่งในช่วง $0.06 \leq d < 0.07$ จำแนกบล็อกผิดพลาดมากที่สุด จำนวน 32 บล็อก แต่เมื่อพิจารณาแต่ละบล็อกที่จำแนกผิดพลาดแล้วพบว่า ตัวอักษรที่มักผิดพลาดนั้นเป็นอักษรไม้ซัดและไม้ขอช้าง ซึ่งยากที่จะแยกความแตกต่างด้วยเค-เนียร์เรสเนเบอร์ แต่สำหรับระยะห่างตั้งแต่ $d \geq 0.13$ ส่วนมากจำแนกผิดพลาด เนื่องจากไม่พบตัวอักษรในชุดข้อมูล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ฝึกฝน หรือตัวอักษรได้รับความเสียหายมาก เช่น เส้นขาดหาย หรือเส้นหนาติดกันหลายตัวอักษรในหนึ่งบล็อก

4.3.2 การทดสอบประสิทธิภาพการรู้จำตัวพิมพ์อักษรธรรมล้านนาด้วยวิธีเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิง

จากผลการทดสอบเพื่อเปรียบเทียบประสิทธิภาพการรู้จำระหว่างเค-เนียร์เรสเนเบอร์และซัพพอร์ตเวกเตอร์แมชชีนในหัวข้อ 4.3.1 ในขั้นตอนนี้จึงใช้เค-เนียร์เรสเนเบอร์ในการรู้จำตัวพิมพ์อักษรล้านนาร่วมกับเทมเพลตแมทชิง เพื่อแก้ปัญหาในกรณีที่ประเภทในชุดข้อมูลฝึกสอนอาจมีไม่ครอบคลุมข้อมูลทดสอบทั้งหมด เนื่องจากในงานวิจัยนี้รู้จำเป็นบล็อกตัวอักษรจึงมีประเภทของตัวอักษรจำนวนมาก ในการทดสอบนี้จึงมีวัตถุประสงค์เพื่อทดสอบประสิทธิภาพการรู้จำตัวพิมพ์อักษรธรรมล้านนาด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิง

กำหนดให้ dis_{th} เป็นค่าเทรชโฮลด์ของระยะห่างระหว่างบล็อกกับข้อมูลฝึกสอนมีค่าเท่ากับ 0.13 ถ้าบล็อกอักษรนั้นมีระยะห่างกับข้อมูลฝึกสอนมากเกินกว่าค่า 0.13 บล็อกอักษรนั้นจะถูกจำแนกซ้ำอีกครั้งด้วยเทมเพลตแมทชิง สำหรับขั้นตอนการจำแนกตัวพิมพ์อักษรธรรมล้านนาด้วยเค-เนียร์เรสเนเบอร์ กำหนดให้ค่า k เท่ากับ 1 ส่วนชุดข้อมูลฝึกสอนสำหรับเค-เนียร์เรสเนเบอร์มีทั้งหมด 1,108 ประเภท จำนวน 20,917 ตัวอย่าง ส่วนชุดข้อมูลฝึกสอนสำหรับเทมเพลตแมทชิงทั้งหมด 116 ประเภท จำนวน 1,896 ตัวอย่าง ชุดข้อมูลทดสอบทั้งหมดจำนวน 4,722 บล็อก หรือ 9,656 ตัวอักษร ซึ่งชุดข้อมูลฝึกสอนสำหรับเค-เนียร์เรสเนเบอร์ และชุดข้อมูลทดสอบเป็นชุดเดียวกับการทดสอบหัวข้อ 4.1.2 และผลการทดสอบแสดงดังตาราง 4.3

ตาราง 4.3 ผลการทดสอบการรู้จำด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิง

วิธีการที่ใช้	ค่าเฉลี่ยความแม่นยำ	ค่าเฉลี่ยความถูกต้อง	F-Measure
เค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิง	0.9223	0.9103	0.9163

จากผลการทดสอบพบว่า การรู้จำตัวพิมพ์อักษรธรรมล้านนาด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิงมีประสิทธิภาพดีกว่าการรู้จำด้วยเค-เนียร์เรสเนเบอร์ วัดได้จากค่า F-Measure ที่สูงกว่าการรู้จำด้วยเค-เนียร์เรสเนเบอร์ นั่นคือ 0.9163 เนื่องจากเทมเพลตแมทชิงช่วยในการจำแนกตัวอักษรที่คาดว่าจำแนกด้วยเค-เนียร์เรสเนเบอร์ผิดพลาด ซึ่งเป็นประเภทตัวอักษรที่ไม่อยู่ในชุดข้อมูลฝึกสอน หรือเป็นบล็อกตัวอักษรที่มีลักษณะไม่สมบูรณ์ ลายเส้นขาดจากกัน หรือ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ติดกันหลายตัวอักษร แต่อย่างไรก็ตามยังคงมีความผิดพลาดในการจำแนกตัวอักษรที่มีลายเส้นไม่สมบูรณ์ หรือขาดหายมาก รวมทั้งตัวอักษรที่มีลักษณะคล้ายคลึงกัน ซึ่งตัวอย่างอักษรที่มีลักษณะคล้ายคลึงกัน แสดงในรูป 4.3

รูป 4.3 ตัวอย่างกลุ่มอักษรที่มีลักษณะคล้ายกัน

4.3.3 การทดสอบประสิทธิภาพการรู้จำตัวพิมพ์อักษรธรรมล้านนาด้วยวิธีเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิงและคอนดิชันนัลแรนดอมฟิลด์

จากผลการทดสอบหัวข้อ 4.3.2 การรู้จำด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิงยังคงมีความผิดพลาดในการรู้จำอักษร ในขั้นตอนนี้จึงมีการใช้คอนดิชันนัลแรนดอมฟิลด์ร่วมกับการรู้จำด้วยเค-เนียร์เรสเนเบอร์และเทมเพลตแมทชิง เพื่อพิจารณาอักษรข้างเคียงร่วมกับอักษรที่ต้องการจำแนก โดยการจำแนกตัวอักษรด้วยคอนดิชันนัลแรนดอมฟิลด์จะดำเนินการเป็นขั้นตอนสุดท้ายหลังจากการจำแนกด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิง โดยชุดข้อมูลฝึกสอนสำหรับคอนดิชันนัลแรนดอมฟิลด์ใช้หนังสือจากหมวดคร่าวซอ 5 เล่ม ความยาวจำนวน 41,129 ตัวอักษร

ขั้นตอนการจำแนกตัวอักษรด้วยคอนดิชันนัลแรนดอมฟิลด์กำหนดให้มีรูปแบบการพิจารณาตัวอักษรข้างเคียง 4 รูปแบบ ดังนี้

1. พิจารณาอักษรด้านหน้าตัวอักษรที่กำลังพิจารณา 1 ตัว และอักษรด้านหลังตัวอักษรที่กำลังพิจารณา 1 ตัว
2. พิจารณาอักษรด้านหน้าตัวอักษรที่กำลังพิจารณา 2 ตัว และอักษรด้านหลังตัวอักษรที่กำลังพิจารณา 1 ตัว
3. พิจารณาอักษรด้านหน้าตัวอักษรที่กำลังพิจารณา 1 ตัว และอักษรด้านหลังตัวอักษรที่กำลังพิจารณา 2 ตัว
4. พิจารณาอักษรด้านหน้าตัวอักษรที่กำลังพิจารณา 2 ตัว และอักษรด้านหลังตัวอักษรที่กำลังพิจารณา 2 ตัว

โดยมีการเปรียบเทียบการใช้คุณลักษณะ 3 แบบ คือ แบบที่หนึ่งใช้ระดับของตัวอักษร 1 คุณลักษณะ แบบที่สองใช้หมวดของอักษรธรรมล้านนาจากการจำแนกด้วยเค-เนียร์เรสเนเบอร์และเทมเพลตแมทชิง 1 คุณลักษณะ แบบที่สามใช้ทั้ง 2 คุณลักษณะ คือหมวดของอักษรธรรมล้านนา

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

จากการจำแนกด้วยเค-เนียร์เรสเนเบอร์และเทมเพลตแมทชิง และระดับของตัวอักษร โดยทดสอบร่วมกับการใช้รูปแบบการพิจารณาตัวอักษรข้างเคียงทั้ง 4 ประเภท ผลการทดสอบแสดงดังตาราง 4.4

ตาราง 4.4 ผลการทดสอบการรู้จำด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิงและคอนดิชัน-นัลแรนดอมฟิลด์

คุณลักษณะ	รูปแบบการพิจารณา	ค่าเฉลี่ยความแม่นยำ	ค่าเฉลี่ยความถูกต้อง	F-Measure
คุณลักษณะแบบที่ 1	หน้า 1 หลัง 1	0.93967	0.92751	0.93355
	หน้า 2 หลัง 1	0.94061	0.92844	0.93445
	หน้า 1 หลัง 2	0.93956	0.92740	0.93344
	หน้า 2 หลัง 2	0.94428	0.93206	0.93813
คุณลักษณะแบบที่ 2	หน้า 1 หลัง 1	0.93526	0.92316	0.92917
	หน้า 2 หลัง 1	0.93830	0.92616	0.93219
	หน้า 1 หลัง 2	0.93746	0.92533	0.93136
	หน้า 2 หลัง 2	0.94134	0.92916	0.93521
คุณลักษณะแบบที่ 3	หน้า 1 หลัง 1	0.94386	0.93165	0.93771
	หน้า 2 หลัง 1	0.94208	0.92989	0.93594
	หน้า 1 หลัง 2	0.94124	0.92906	0.93511
	หน้า 2 หลัง 2	0.94544	0.93320	0.93928

จากผลการทดสอบพบว่าการรู้จำด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิงและคอนดิชันนัลแรนดอมฟิลด์ ทั้งคุณลักษณะแบบที่ 1 คุณลักษณะแบบที่ 2 และคุณลักษณะแบบที่ 3 เมื่อใช้ร่วมกับรูปแบบการพิจารณาตัวอักษรข้างเคียงแบบหน้า 2 หลัง 2 มีประสิทธิภาพดีที่สุด วัดได้จาก ค่า F-Measure ที่มีค่ามากที่สุดในคุณลักษณะแต่ละแบบ ส่วนคุณลักษณะแบบที่ 3 ซึ่งใช้ครบทั้งสองคุณลักษณะคือ หมวดของอักษรธรรมล้านนาจากการจำแนกด้วยเค-เนียร์เรสเนเบอร์และเทมเพลตแมทชิง และระดับของตัวอักษร เมื่อใช้ร่วมกับรูปแบบการพิจารณาตัวอักษรข้างเคียงแบบหน้า 2 หลัง 2 มีประสิทธิภาพดีที่สุด วัดได้จาก F-Measure ที่มีค่ามากที่สุด คือ 0.93928

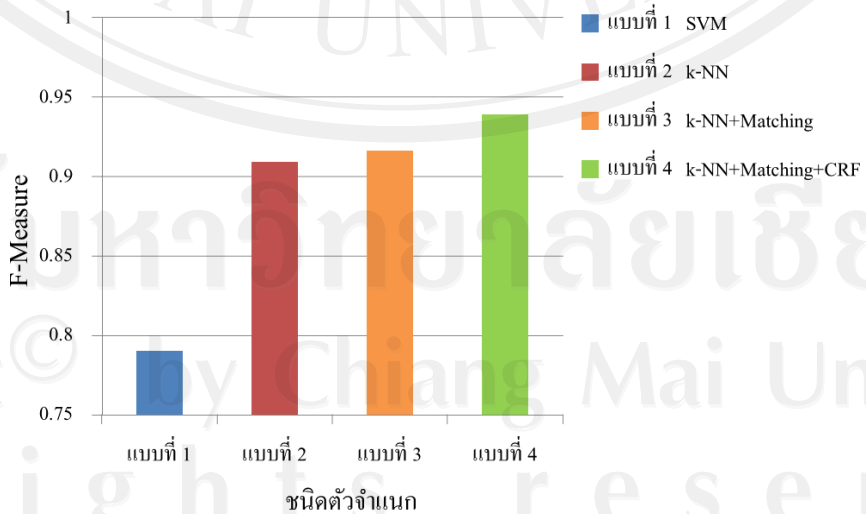
เมื่อพิจารณาผลการจำแนกด้วยคอนดิชันนัลแรนดอมฟิลด์ พบว่าตัวอักษรที่ผิดพลาดจากการจำแนกด้วยเค-เนียร์เรสเนเบอร์และเทมเพลตแมทชิง เมื่อจำแนกซ้ำและจำแนกได้ถูกต้อง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

มักเป็นอักษรที่ผิดพลาดและเกิดร่วมกับอักษรข้างเคียงตัวเดิมเป็นประจำ ซึ่งตัวอักษรที่ผิดสามารถจำแนกให้ถูกต้องมากที่สุด คือตัวอักษรไม้ซัด ( ) สามารถจำแนกให้เป็น ไม้ขอช้าง ( ) ได้ถูกต้อง นอกนั้นไม่สามารถจำแนกซ้ำให้ถูกต้องได้ อาจมีสาเหตุจากจำนวนชุดข้อมูลฝึกสอนสำหรับคอนดิชันนัลแรนดอมฟิลด์ยังมีจำนวนไม่มากพอ

4.4 วิเคราะห์และสรุปผลการวิจัย

กระบวนการรู้จำตัวพิมพ์อักษรธรรมล้านนา เมื่อแบ่งขั้นตอนการทดสอบออกเป็นสามส่วนแล้วพบว่า ขั้นตอนการตัดแบ่งบรรทัด สามารถตัดแบ่งบรรทัดได้ด้วยความถูกต้องสูงถึงร้อยละ 97.14 สำหรับขั้นตอนการตัดแบ่งบล็อก สามารถตัดแบ่งบล็อกได้ด้วยความถูกต้องสูงถึงร้อยละ98.63 และในขั้นตอนการรู้จำตัวอักษร การรู้จำด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิงและคอนดิชันนัลแรนดอมฟิลด์มีประสิทธิภาพดีที่สุดในการรู้จำตัวพิมพ์อักษรธรรมล้านนา กำหนดให้การรู้จำตัวพิมพ์อักษรธรรมล้านนาด้วยวิธีเค-เนียร์เรสเนเบอร์ ด้วยค่า k เท่ากับ 1 และ dis_{th} มีค่าเท่ากับ 0.13 ถ้าบล็อกอักษรนั้นมีความแตกต่างกับข้อมูลฝึกสอนมากเกินกว่าค่า 0.13 บล็อกอักษรนั้นจึงถูกจำแนกซ้ำอีกครั้งด้วยวิธีเทมเพลตแมทชิง จากนั้นจำแนกตัวอักษรด้วยคอนดิชันนัลแรนดอมฟิลด์จะดำเนินการเป็นขั้นตอนสุดท้าย โดยใช้คุณลักษณะแบบที่ 3 คือ หมวดของอักษรธรรมล้านนาจากการจำแนกด้วยเค-เนียร์เรสเนเบอร์และเทมเพลตแมทชิง และระดับของตัวอักษร และใช้รูปแบบการพิจารณาตัวอักษรข้างเคียงแบบหน้า 2 หลัง 2 โดยมีค่า F-Measure สูงที่สุด คือ 0.93928 เมื่อเปรียบเทียบกับการจำแนกด้วยซัพพอร์ตเวกเตอร์แมชชีน เค-เนียร์เรสเน-เบอร์ และเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทชิง ดังรูป 4.4 แสดงการเปรียบเทียบประสิทธิภาพวิธีการรู้จำอักษรธรรมล้านนา



รูป 4.4 การเปรียบเทียบประสิทธิภาพวิธีการรู้จำอักษรธรรมล้านนา

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved