บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึง ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับอักษรธรรมล้านนา การประมวลผลภาพเบื้องต้น และการรู้จำตัวอักษร โดยมีรายละเอียดดังต่อไปนี้

2.1 อักษรธรรมล้านนา

การศึกษาค้นคว้าความเป็นมาของอักษรธรรมล้านนานั้นยังไม่มีข้อสรุป หรือหลักฐานที่สามารถยืนยันได้แน่ชัดว่ามีต้นกำเนิดมาจากที่ใด แต่มีผู้รู้หลายท่านได้ศึกษาค้นคว้าและตั้งข้อสันนิษฐานไว้ จากการศึกษาของนัยนา โปร่งธุระ พบว่าหลักฐานประเภทจารึกที่เก่าแก่ที่สุด คือ จารึกลานทอง พ.ศ. 1919 ที่ใช้อักษรธรรมคู่กับอักษรไทยสุโขทัย โดยพบที่ฐานพระประธาน วัดมหาธาตุ จังหวัดสุโขทัย ส่วนการกำเนิดอักษรธรรมล้านนา อาจมีสาเหตุสำคัญมาจากทางศาสนา เพื่อใช้เขียนพระธรรมคัมภีร์ให้เป็นสื่อแก่คนในวงกว้าง อักษรธรรมล้านนาจึงเริ่มจากการใช้เขียนภาษาบาลี และนิยมใช้จารลงในคัมภีร์ใบลาน เพราะถือว่าเป็นอักษรที่เป็นตัวแทนของพระธรรมคัมภีร์ทางพระพุทธศาสนา ดังมีคำกล่าวที่ว่า “อักษรธรรมแต่ละตัว เปรียบได้กับพระพุทธเจ้าหนึ่งพระองค์” อันแสดงให้เห็นฐานะอันสูงส่งของอักษรธรรมอย่างชัดเจน ในด้านบทบาทของอักษรธรรม ถือได้ว่าเป็นอักษรที่รับใช้ในทางพระพุทธศาสนาเป็นหลัก และนอกจากนี้ยังเป็นตัวอักษรที่ใช้บันทึกภาษาพูดของกลุ่มไทยวนในระยะต่อมาด้วย (อภิวันท์ พันธ์สุข, 2551)

สำหรับโรงพิมพ์อักษรธรรมล้านนา เกิดขึ้นครั้งแรกเมื่อ พ.ศ. 2435 โดยศาสนาจารย์ Daniel McGilvary ซึ่งเป็นมิชชันนารีชาวอเมริกัน ได้ตั้งโรงพิมพ์ขึ้นที่ตำบลวังสิงห์คำ และได้ออกแบบตัวพิมพ์อักษรธรรมล้านนา เพื่อใช้ในการประกาศศาสนาและเผยแพร่ข่าวสารของคณะมิชชันนารี รวมทั้งเอกสารอื่นๆ ต่อมาคณะมิชชันนารีเห็นว่านโยบายของประเทศที่ให้ใช้ภาษาไทยกลางเป็นภาษาราชการ จึงขายโรงพิมพ์ให้นายเมืองใจ ไชยนิลพันธุ์ ใน พ.ศ. 2469 และได้ตั้งชื่อโรงพิมพ์ว่า “โรงพิมพ์เจริญเมือง” โดยรับพิมพ์เอกสารต่างๆ ทั้งด้วยอักษรไทยกลางและอักษรธรรมล้านนา โรงพิมพ์ดำเนินการจนถึง พ.ศ. 2493 จึงปิดกิจการ นอกจากนี้ยังมีโรงพิมพ์ของชาวล้านนาที่เกิดขึ้นตามมาหลายโรงพิมพ์ และได้ดำเนินงานมาจนถึง พ.ศ. 2496 แล้วจึงค่อยๆ เลิกราไป (จีรยุทธ ไชยจารุวณิช และคณะ, 2552)

2.1.1 ประเภทตัวอักษรธรรมล้านนา

ในงานวิจัยนี้ได้นำเสนอประเภทของตัวอักษรธรรม เพื่อสะดวกและง่ายแก่การทำความ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

เข้าใจเบื้องต้นทั้งหมด 7 ประเภท คือ พยัญชนะ พยัญชนะตัวเชิง สระเดี่ยว สระลอย เครื่องหมาย ตัวเลขโหรา และตัวเลขในธรรม (นัยนา โปร่งธุระ, 2527)

(1) พยัญชนะ

พยัญชนะในอักษรธรรมล้านนาที่ปรากฏในฐานะพยัญชนะต้น แม้จะมีความกว้างแตกต่างกันไปบ้าง แต่ความสูงของพยัญชนะเหล่านั้นจะต้องเท่ากันเสมอ โดยพยัญชนะทั้งหมดมี 42 รูป แสดงในตาราง 2.1

ตาราง 2.1 พยัญชนะในอักษรธรรมล้านนา

พยัญชนะ			พยัญชนะ			พยัญชนะ		
ลำดับ	อักษรธรรม	อักษรไทย	ลำดับ	อักษรธรรม	อักษรไทย	ลำดับ	อักษรธรรม	อักษรไทย
1	๓	ก	15	๕	ฏ,ฑ,ด,	29	๒๖	ภ
2	๖	ข	16	๒๗	ฒ	30	๒๘	ม
3	๑	ค	17	๓	ณ	31	๒๙	ย
4	๗	ฅ	18	๐	ต	32	๒๗	อย
5	๒๘	ฆ	19	๘	ถ	33	๒๑	ร
6	๑	ง	20	๑๑	ท	34	๒๐	ล
7	๑๐	จ	21	๑	ธ	35	๐	ว
8	๒๓	ฉ	22	๔	น	36	๒๔	ศ
9	๕	ช	23	๒	บ	37	๒๘	ษ
10	๕	ซ	24	๒	ป	38	๒๓	ส
11	๑๒	ฌ	25	๒๓	ผ	39	๒๒	ห
12	๒	ญ	26	๕	ฝ	40	๒๕	ฬ
13	๒	ฏ	27	๑	พ	41	๒๘	อ
14	๑๒	ฐ	28	๗	ฟ	42	๒๒	ฮ

(2) พยัญชนะตัวเชิง

พยัญชนะตัวเชิง มีจำนวนทั้งหมด 33 รูป โดยใช้พยัญชนะตัวเชิง เมื่อพยัญชนะ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ไม่ได้ปรากฏในฐานะของพยัญชนะต้น แสดงดังตาราง 2.2

ตาราง 2.2 พยัญชนะตัวเชิง ในอักษรธรรมล้านนา

พยัญชนะ			พยัญชนะ			พยัญชนะ		
ลำดับ	ต้น	ตัวเชิง	ลำดับ	ต้น	ตัวเชิง	ลำดับ	ต้น	ตัวเชิง
1	ᨣ	ᩣ	15	ᨯ	ᩯ	29	ᨹ	᩹
2	ᨦ	ᩦ	16	ᨰ	ᩰ	30	ᨺ	᩺
3	ᨧ	ᩧ	17	ᨱ	ᩱ	31	ᨻ	᩻
4	ᨩ	-	18	ᨲ	ᩲ	32	ᨼ	-
5	ᨫ	-	19	ᨵ	᩵	33	ᨽ	᩽
6	ᨬ	ᩬ	20	ᨶ	᩶	34	ᨾ	᩾
7	ᨭ	ᩭ	21	ᨷ	᩷	35	ᨿ	᩿
8	ᨮ	ᩮ	22	ᨸ	᩸	36	ᩀ	ᩀ
9	ᨳ	ᩳ	23	ᨹ	-	37	ᩁ	ᩁ
10	ᨵ	-	24	ᨻ	᩻	38	ᩂ	ᩂ
11	ᨶ	᩶	25	ᨼ	᩼	39	ᩃ	ᩃ
12	ᨷ	᩷	26	ᨽ	-	40	ᩄ	-
13	ᨸ	᩸	27	ᨾ	ᩌ	41	ᩅ	ᩍ
14	ᨹ	ᩌ	28	ᨿ	-	42	ᩆ	-

(3) สระเดี่ยว

สระเดี่ยว เป็นสระที่ใช้ผสมกับพยัญชนะ มี 14 รูป รายละเอียดแสดงดังตาราง 2.3

(4) สระลอย

สระลอยเป็นสระที่มาจากภาษาบาลี ซึ่งเขียนอยู่ลำพังโดยไม่อาศัยรูปพยัญชนะเป็นหลัก ปรากฏทั้งสิ้น 8 รูป รายละเอียดแสดงดังตาราง 2.4

(5) เครื่องหมาย

อักษรธรรมล้านนามีเครื่องหมายที่ใช้เขียนกำกับคำศัพท์ เพื่อช่วยให้สามารถบันทึกเสียงและภาษาล้านนาให้เป็นไปอย่างชัดเจนและสมบูรณ์ มีจำนวน 9 รูป แสดงดังตาราง 2.5

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ตาราง 2.3 สระเดี่ยวในอักษรธรรมล้านนา

สระเดี่ยว			สระเดี่ยว			สระเดี่ยว		
ลำดับ	อักษรธรรม	อักษรไทย	ลำดับ	อักษรธรรม	อักษรไทย	ลำดับ	อักษรธรรม	อักษรไทย
1	-ะ	-ะ	7	-ุ	-ุ	13	เ-	เ-
2	-า	-า	8	-ู	-ู	14	เ-	ไ-
3	-ิ	-ิ	9	เ-	เ-			
4	-ี	-ี	10	แ-	โ-			
5	-ึ	-ึ	11	-ำ	-อ			
6	-ื	-ื	12	-ั	-ัว			

ตาราง 2.4 สระลอยในอักษรธรรมล้านนา

สระลอย			สระลอย			สระลอย		
ลำดับ	อักษรธรรม	เสียง	ลำดับ	อักษรธรรม	เสียง	ลำดับ	อักษรธรรม	เสียง
1	ฤ	อิ	4	ฦ	อู	7	ๅ	ฺรึ
2	ฤๅ	อี	5	ฦๅ	เอ	8	ๆ	ฺลึ
3	ฦ	อุ	6	ใ-	โอ			

ตาราง 2.5 เครื่องหมายในอักษรธรรมล้านนา

เครื่องหมาย			เครื่องหมาย			เครื่องหมาย		
ลำดับ	อักษรธรรม	ชื่อ	ลำดับ	อักษรธรรม	ชื่อ	ลำดับ	อักษรธรรม	ชื่อ
1	◌๋	ไม้หยัก	4	◌๎	ระห้าม	7	◌๊	ไม้เก๋าห่อหนึ่ง
2	◌๋	ไม้ขอช้าง	5	◌๏	ไม้กั๋งมน	8	◌๎	ไม้ซ้ำ
3	◌๋	ไม้ซัด	6	◌๎	ไม้กั๋งไหล	9	◌๊	ไม้ก๋ง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

(6) ตัวเลขโหรา

ตัวเลขโหราใช้ในการเขียนบอกจำนวนในเอกสารต่างๆ มีทั้งหมด 10 รูป ดังแสดงในตาราง 2.6

(7) ตัวเลขในธรรม

ตัวเลขในธรรมใช้ในการเขียนเรื่องราวทางศาสนา มักเป็นเอกสารประเภทใบลาน มีทั้งหมด 10 รูป ดังแสดงในตาราง 2.6

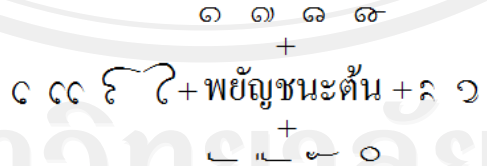
(นัยนา โปร่งธุระ, 2527)

ตาราง 2.6 ตัวเลขในอักษรธรรมล้านนา

ตัวเลขโหรา	ง	ง	๒	๖	๖	๖	๗	๑	๔	๐
ตัวเลขในธรรม	๑	๒	๓	๔	๕	๖	๗	๘	๙	๐
ไทย	๑	๒	๓	๔	๕	๖	๗	๘	๙	๐

2.1.2 โครงสร้างคำของตัวอักษรธรรมล้านนา

การเขียนอักษรธรรมล้านนา สระต่างๆ จะมีตำแหน่งตายตัว โดยจะปรากฏอยู่ในด้านต่างๆ ของพยัญชนะต้น คือ ด้านหน้า ด้านหลัง ด้านบน และด้านล่าง ซึ่งเป็นตำแหน่งคล้ายกับภาษาไทย ยกเว้นไม้ก๋อ (◌ั) และไม้กั๋ว (◌ั) จะวางไว้ด้านล่างของพยัญชนะต้น ซึ่งตำแหน่งที่รูปสระปรากฏอยู่นี้จะมีส่วนสัมพันธ์กับตำแหน่งของตัวสะกดอีกด้วย ซึ่งตัวสะกดโดยทั่วไปจะอยู่ด้านล่างของพยัญชนะต้น และนิยมใช้อักษรตัวเชิงเป็นตัวสะกด ยกเว้นว่ามีรูปสระอยู่ด้านล่างของพยัญชนะต้นอยู่แล้ว ตัวสะกดนั้นจะปรากฏอยู่ด้านหลังของพยัญชนะต้นและปรากฏในรูปอักษรปกติ ตัวอย่างโครงสร้างคำในภาษาล้านนา แสดงดังรูป 2.1 (เกษม ศิริรัตน์พิริยะ, 2548)



รูป 2.1 โครงสร้างคำในภาษาล้านนา

2.1.3 รูปแบบการพิมพ์อักษรธรรมล้านนา

ในการพิมพ์อักษรธรรมล้านนา มีโครงสร้างการวางตัวอักษรธรรมล้านนาที่ประกอบด้วย 3 ระดับหลัก คือ ระดับกลาง ระดับบน และระดับล่าง โดยระดับบนและระดับล่างถูก

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

แบ่งเป็นระดับย่อยอีกหนึ่งระดับ ดังรูป 2.2 โครงสร้างการวางตัวอักษรธรรมล้านนา ซึ่งระดับที่ 1 หรือระดับกลาง สำหรับพยัญชนะต้น ตัวสะกดที่ไม่ใช่พยัญชนะซ้อน อักษรพิเศษระดับกลาง สระระดับกลาง หรือตัวเลข ระดับที่ 2 สำหรับสระระดับบน วรรณยุกต์ หรือตัวอักษรพิเศษระดับบน ระดับที่ 3 สำหรับวรรณยุกต์ เมื่อมีสระอยู่เหนือพยัญชนะต้น ระดับที่ 4 สำหรับพยัญชนะซ้อน หรือสระระดับล่าง และระดับที่ 5 สำหรับสระที่อยู่ใต้พยัญชนะซ้อน ซึ่งเรียกรวมระดับที่ 2 และ 3 ว่า ระดับบน และเรียกรวมระดับที่ 4 และ 5 ว่า ระดับล่าง

ในการพิมพ์ตัวอักษรบนเครื่องคอมพิวเตอร์ ด้วยแบบอักษร TILOK คล้ายลำดับการวางตัวอักษร ซึ่งขั้นตอนการพิมพ์มีลำดับดังต่อไปนี้

(1) พิมพ์พยัญชนะก่อน

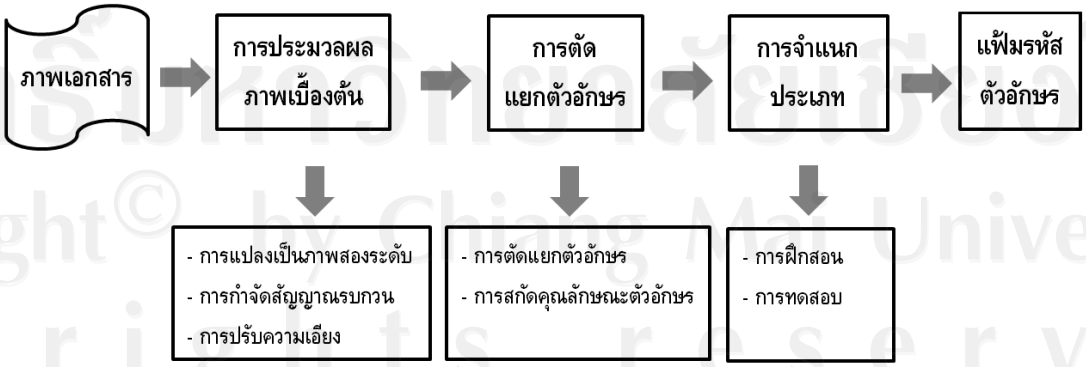
(2) พิมพ์สระด้านบน หรือวรรณยุกต์ด้านบน

(3) พิมพ์ตัวควบกล้ำ หรือตัวสะกดที่เป็นตัวซ้อน ตามด้วยไปยาลน้อย (พิชัย แสงบุญ, 2552)

รูป 2.2 โครงสร้างการวางตัวอักษรธรรมล้านนา

2.2 กระบวนการรู้จำตัวอักษร

กระบวนการรู้จำตัวอักษรประกอบด้วยการทำงาน 3 ขั้นตอน คือ การประมวลผลภาพเบื้องต้น การตัดแยกตัวอักษร และการจำแนกประเภท ดังรูป 2.3 โดยในแต่ละขั้นตอนมีการทำงานข้อมูลนำเข้า และข้อมูลส่งออก ดังนี้



รูป 2.3 กระบวนการรู้จำตัวอักษร

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

(1) การประมวลผลภาพเบื้องต้น (Preprocessing)

ขั้นตอนนี้รูปภาพเอกสารจะถูกปรับปรุงคุณลักษณะต่างๆ เช่น การแปลงเป็นภาพสองระดับ การกำจัดสัญญาณรบกวน การปรับเอียงภาพเอกสาร เป็นต้น เพื่อให้ได้รูปภาพเอกสารที่เหมาะสมในการตัดแยก และรู้จำตัวอักษร

- ข้อมูลนำเข้า : รูปภาพของเอกสารต้นฉบับ
- ข้อมูลส่งออก : รูปภาพของเอกสารที่เหมาะสมในการตัดแยกตัวอักษร

(2) การตัดแยกตัวอักษร (Segmentation)

การตัดแยกตัวอักษร เป็นการนำรูปภาพเอกสารที่ได้จากขั้นตอนการประมวลผลภาพเบื้องต้นมาสกัดหาส่วนที่เป็นตัวอักษร แล้วทำการตัดรูปภาพนั้นให้เป็นรูปภาพตัวอักษรแต่ละตัว โดยทั่วไปจะใช้พื้นที่สีขาวรอบรูปเป็นตัวกำหนดขอบเขตในการตัด ในขั้นตอนนี้มักประสบปัญหาที่ส่งผลกระทบต่ออัตราความถูกต้องของการรู้จำสองปัญหา ปัญหาแรก คือ ปัญหาตัวติด เกิดจากรูปของตัวอักษรตั้งแต่สองตัวขึ้นไปมีส่วนที่เชื่อมติดกัน ทำให้ไม่สามารถแยกตัวอักษรออกจากกันโดยใช้พื้นที่สีขาวรอบตัวอักษรได้ ส่วนปัญหาที่สองในทางตรงกันข้าม คือ เป็นปัญหาตัวขาดที่รูปตัวอักษรหนึ่งถูกแยกออกเป็นส่วน ทำให้ตัดตัวอักษรจากตัวเดียวเป็นหลายตัวอักษร ซึ่งต้องหาวิธีการเฉพาะสำหรับจัดการปัญหาดังกล่าว จากนั้นนำรูปภาพตัวอักษรแต่ละตัวมาสกัดคุณลักษณะส่วนสำคัญออกจากข้อมูลดิบ เนื่องจากการนำข้อมูลดิบไปใช้ในการรู้จำโดยตรงนั้นอาจไม่เหมาะสมทั้งเรื่องของหน่วยความจำ ความถูกต้อง และระยะเวลาที่ใช้ในการฝึกสอน ดังนั้นรูปภาพของตัวอักษรจึงถูกนำเสนอในรูปแบบของคุณลักษณะ เพื่อเข้าสู่ขั้นตอนการจำแนกประเภทต่อไป

- ข้อมูลนำเข้า : รูปภาพของเอกสารที่ได้จากขั้นตอนการประมวลผลภาพเบื้องต้น
- ข้อมูลส่งออก : เวกเตอร์คุณลักษณะที่สกัดจากรูปภาพตัวอักษรแต่ละตัว

(3) การจำแนกประเภท (Classification)

เป็นการนำเวกเตอร์คุณลักษณะของรูปภาพตัวอักษรแต่ละตัวของชุดข้อมูลฝึกสอนมาใช้ฝึกสอนตัวแบบ เพื่อจำแนกเวกเตอร์คุณลักษณะชุดข้อมูลทดสอบเป็นประเภทต่างๆ ซึ่งมีหลายวิธีการ เช่น วิธีการเทียบคู่รูปแบบ วิธีการทางสถิติ วิธีการวิเคราะห์ทางโครงสร้าง และวิธีทางโครงข่ายประสาทเทียม เป็นต้น

- ข้อมูลนำเข้า : เวกเตอร์คุณลักษณะ
- ข้อมูลส่งออก : ลำดับของรหัสตัวอักษรที่ได้จากการรู้จำ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.3 การประมวลผลภาพเบื้องต้น

ทฤษฎีที่เกี่ยวข้องกับการประมวลผลภาพเบื้องต้นสำหรับการรู้จำอักษรธรรมล้านนา มีดังต่อไปนี้

2.3.1 ภาพดิจิทัล (Digital image)

ภาพดิจิทัล คือฟังก์ชัน f(*x*,*y*) สองมิติของความเข้มของแสง (Intensity) โดยที่ *x* และ *y* คือค่าแสดงตำแหน่งในระบบพิกัดฉาก ค่าของฟังก์ชัน f ณ ตำแหน่ง (*x*,*y*) ใดๆ เป็นสัดส่วนกับความสว่างของแสง ณ ตำแหน่งนั้น ภาพดิจิทัลสามารถแบ่งได้หลายประเภท ในงานวิจัยนี้เกี่ยวข้องกับภาพดิจิทัล 2 ประเภท คือ ภาพระดับสีเทา และภาพสองระดับ มีรายละเอียดดังนี้

(1) ภาพระดับสีเทา (Gray scale image)

ภาพระดับสีเทา คือ ภาพที่มีค่าของแต่ละพิกเซลเป็นค่าความเข้มแสง โดยมีค่าความเข้มแสงที่น้อยที่สุด คือสีดำ ไปจนถึงค่าความเข้มแสงที่มากที่สุด คือสีขาว ภาพระดับสีเทามีได้หลายระดับ เช่น 4 ระดับ 16 ระดับ หรือ 256 ระดับ ในภาพหนึ่งถ้าต้องการแบ่งระดับความเข้มแสงหรือระดับสีเทาให้มีหลายค่านั้นต้องเพิ่มจำนวนบิตสำหรับแสดงค่า เช่น ถ้าต้องการภาพระดับสีเทา 4 ระดับ ต้องแทนด้วยเลขฐานสองจำนวน 2 บิต ถ้าต้องการภาพที่มีระดับสีเทา 16 ระดับ ต้องแทนด้วยเลขฐานสองจำนวน 4 บิต เป็นต้น

(2) ภาพสองระดับ (Binary image)

ภาพสองระดับมีค่าความเข้มแสงเพียง 2 ค่าเท่านั้น คือ 0 และ 1 หากพิกเซลใดที่มีค่าเป็น 0 หมายถึงพิกเซลนั้นแสดงสีดำ และพิกเซลใดที่มีค่าเป็น 1 หมายถึงพิกเซลนั้นแสดงสีขาว (โอฬาริก สุรินต๊ะ, 2553)

2.3.2 การแปลงเป็นภาพสองระดับ (Binarization)

ภาพเอกสารที่นำมาประมวลผลเป็นภาพระดับสีเทาแบบ 256 ระดับ ข้อมูลสีจึงมีค่าที่เป็นไปได้ตั้งแต่ 0 ถึง 255 ดังนั้นเพื่อให้ง่ายต่อการแยกตัวอักษรจากสีพื้นหลัง และหาขอบเขตของวัตถุที่สนใจ จึงต้องแปลงภาพให้เป็นภาพสองระดับก่อน โดยการเปรียบเทียบค่าความเข้มแสงของภาพระดับสีเทาของพิกเซลที่กำลังพิจารณากับค่าคงที่ค่าหนึ่งที่เรียกว่า ค่าเทรชโฮลด์ (Threshold) กำหนดให้พิกเซลที่มีค่ามากกว่าหรือเท่ากับค่าเทรชโฮลด์ มีค่าเป็น 1 คือสีขาว ส่วนพิกเซลที่น้อยกว่าค่าเทรชโฮลด์ มีค่าเป็น 0 คือสีดำ ซึ่งสามารถเขียนได้ดังสมการ (2.1)

$$I_{bw}(x,y) = \begin{cases} 1 \; if \; I(x,y) \geq Threshold \\ 0 \; if \; I(x,y) < Threshold \end{cases} \quad (2.1)$$

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

การกำหนดค่าเทรชโฮลด์ พิจารณาจากฮิสโตแกรมค่าความเข้มแสงของทุกพิกเซลในภาพ สำหรับภาพที่มีความแตกต่างระหว่างพื้นหลังกับตัวอักษรมาก ฮิสโตแกรมจะมี 2 ยอด ซึ่งระดับความเข้มแสงที่อยู่ระหว่างสองยอดนั้นจะถูกเลือกเป็นค่าเทรชโฮลด์ ดังรูป 2.4

ในปัจจุบันมีวิธีการเลือกค่าเทรชโฮลด์ที่มีสมมุติฐานว่าฮิสโตแกรมมี 2 ยอดหลายวิธี แต่วิธีที่เป็นที่นิยมมากที่สุด คือ วิธีการของโอสึ (Otsu's thresholding method) ซึ่งอัลกอริทึมของโอสึ มีหลักการดังต่อไปนี้

(1) แบ่งช่วงข้อมูลของฮิสโตแกรมออกเป็นสองกลุ่มโดยใช้ค่า k (ค่าระดับเทรชโฮลด์) เป็นค่าคั่นกลางระหว่างกลุ่ม เริ่มจากช่วงแรก กลุ่มแรกอยู่ในช่วงระดับ 0 ถึง 1 กลุ่มที่สองอยู่ในช่วงระดับ 1 ถึง 255 จากนั้นเพิ่มช่วงของกลุ่มแรก และลดช่วงของกลุ่มที่สองไปเรื่อยๆ จนกระทั่งกลุ่มแรกอยู่ในช่วง 0 ถึง 254 และกลุ่มที่สองอยู่ในช่วง 254 ถึง 255

(2) หลังจากที่แบ่งกลุ่มแต่ละรอบแล้ว คำนวณหาค่าความแปรปรวนระหว่างกลุ่ม จากสมการ (2.2)

$$\sigma_B^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2 \qquad (2.2)$$

โดยที่

$$P_i = \frac{n_i}{N},\ P_i \geq 0,\quad \sum_{i=1}^{L} P_i = 1 \qquad \mu_1(t) = \frac{\mu(t)}{\omega(t)}$$

$$\omega_1(t) = \sum_{i=1}^{t} P_i \qquad \mu_2(t) = \sum_{i=t+1}^{L} \frac{iP_i}{\omega_1}$$

$$\omega_2(t) = \sum_{i=t+1}^{L} P_i$$

กำหนดให้

N คือ ผลรวมของจำนวนพิกเซลในแต่ละระดับ $N = n_1 + n_2 + \cdots + n_L$

n_i คือ จำนวนพิกเซล ณ ระดับที่ i

P_i คือ จำนวนพิกเซลที่ทำให้เป็นบรรทัดฐานที่มีค่าอยู่ระหว่าง 0 ถึง 1

k คือ ค่าเทรชโฮลด์ที่ใช้แบ่งแยกระหว่างกลุ่มสองกลุ่ม

ω_1 คือ ค่าความน่าจะเป็นของกลุ่มที่ 1

ω_2 คือ ค่าความน่าจะเป็นของกลุ่มที่ 2

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

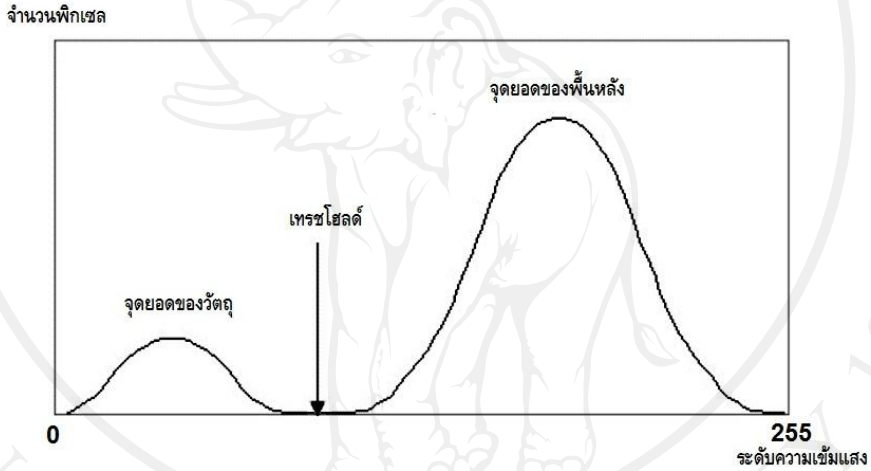
μ_1 คือ ค่าเฉลี่ยของกลุ่มที่ 1

μ_2 คือ ค่าเฉลี่ยของกลุ่มที่ 2

(3) คำนวณหาค่าที่มากที่สุดของค่าความแตกต่างระหว่างกลุ่ม $\sigma^2(k^*)$ ตามสมการ (2.3)

$$\sigma^2(k^*) = \max_{1<k<L} \sigma_B^2(k) \tag{2.3}$$

ดังนั้นพิจารณาการแบ่งกลุ่มช่วงที่มีค่าความแปรปรวนระหว่างกลุ่มมากที่สุด และกำหนดให้ค่า k ที่ใช้แบ่งกลุ่มนั้นจะถูกเลือกเป็นเทรชโฮลด์ (โอพาริก สุรินต๊ะ, 2553)



รูป 2.4 ฮิสโตแกรมความหนาแน่นของค่าความเข้มแสง

2.3.3 การดำเนินการเชิงสัณฐานวิทยา (Morphological operation)

การดำเนินการเชิงสัณฐานวิทยาเป็นการจัดการกับภาพสองระดับ โดยอาศัยหลักการทางคณิตศาสตร์เกี่ยวกับรูปแบบ โครงสร้างวัตถุ และทฤษฎีเซต ซึ่งการดำเนินการประกอบด้วย 3 ส่วน คือ ภาพที่ต้องการดำเนินการ ตัวดำเนินการเชิงสัณฐานวิทยา และสตรักเจอร์อิลิเมนต์ (Structure element) ซึ่งเป็นเมตริกซ์ที่ถูกกำหนดขึ้นให้เป็นรูปร่างและขนาดหนึ่ง ประกอบด้วยค่า 0 (สีดำ) หรือ 1 (สีขาว) การดำเนินการทำได้โดยนำสตรักเจอร์อิลิเมนต์ไปวางแต่ละตำแหน่งบนภาพ แล้วใช้ตัวดำเนินการระหว่างสตรักเจอร์อิลิเมนต์กับภาพที่อยู่ใต้สตรักเจอร์อิลิเมนต์ สุดท้ายจึงนำผลลัพธ์มาปรับค่าที่ตำแหน่งนั้น การดำเนินการเชิงสัณฐานวิทยามีตัวดำเนินการพื้นฐาน 4 แบบ คือ ไดเลชัน (Dilation) อีรอชัน (Erosion) โคลสซิง (Closing) และโอเพนนิง (Opening) (เศรษฐา โพธา, 2552) ตัวอย่างภาพ A และสตรักเจอร์อิลิเมนต์ B ก่อนการดำเนินการเชิงสัณฐานวิทยา แสดงดังรูป 2.5

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

(1) ไดเลชัน

มีเครื่องหมายบวก (+) เป็นสัญลักษณ์ทางคณิตศาสตร์ เพื่อเพิ่มจำนวนขอบของวัตถุในภาพ โดยพิจารณาร่วมกันระหว่างเซตของภาพ A และสตรักเจอร์อิลิเมนต์ B เมื่อนำเซตทั้งสองมาดำเนินการด้วยไดเลชันจะเป็นไปตามความสัมพันธ์ของนิยาม (2.4)

$$A + B = \{x|(\hat{B})_x \cap A \neq \emptyset\} \tag{2.4}$$

ผลลัพธ์ที่ได้ คือ กลุ่มพิกเซลสีดำของเซต A ถูกขยายให้มีขนาดกว้างขึ้น ส่งผลให้ช่องว่างระหว่างพิกเซลสีดำมีขนาดแคบลง หรืออาจทำให้พิกเซลสีดำสองกลุ่มที่อยู่ใกล้กันเชื่อมติดกันได้

(2) อีรอชัน

มีเครื่องหมายลบ (–) เป็นสัญลักษณ์ทางคณิตศาสตร์ เพื่อลดจำนวนขอบของวัตถุในภาพ โดยพิจารณาร่วมกันระหว่างเซตของภาพ A และสตรักเจอร์อิลิเมนต์ B เมื่อนำเซตทั้งสองมาดำเนินการด้วยอีรอชันจะเป็นไปตามความสัมพันธ์ของนิยาม (2.5)

$$A - B = \{x|(B)_x \subseteq A\} \tag{2.5}$$

ผลลัพธ์ที่ได้ คือ กลุ่มพิกเซลสีดำของเซต A จะถูกลดจำนวนลง และทำให้ขนาดช่องว่างระหว่างสองกลุ่มพิกเซลสีดำมีขนาดกว้างมากขึ้น และแยกออกจากกันอย่างชัดเจน ถ้ากลุ่มพิกเซลสีดำมีขนาดเล็กมากอาจถูกขจัดออกจากภาพได้

(3) โคลสซิง

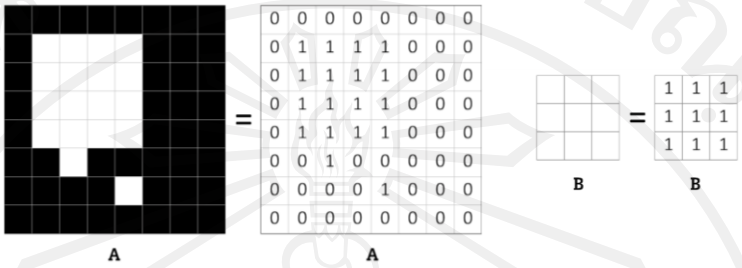
เป็นการแปลงการดำเนินการทางสัณฐานวิทยา โดยเริ่มจากการดำเนินการไดเลชันแล้วตามด้วยการดำเนินการอีรอชันเท่ากับจำนวนครั้งของการดำเนินการไดเลชัน ซึ่งเป็นการเพิ่มจำนวนพิกเซลของขอบวัตถุก่อน จากนั้นค่อยกัดเซาะขอบของวัตถุออกไป ทำให้ขอบของวัตถุเรียบมากขึ้น ถ้ากำหนดจำนวนรอบในการวนทำซ้ำมาก อาจทำให้กลุ่มจุดของวัตถุสองกลุ่ม ที่อยู่ใกล้กันเชื่อมเข้าเป็นกลุ่มวัตถุเดียวกันได้ นอกจากนี้ยังทำให้พิกเซลที่เป็นสัญญาณรบกวนขยายขนาดขึ้นจนไม่สามารถกำจัดออกได้ เนื่องจากยากต่อการแยกแยะว่าเป็นสัญญาณรบกวนหรือไม่ ตัวอย่างการดำเนินการโคลสซิงแสดงดังรูป 2.6 (ก)

(4) โอเพนนิง

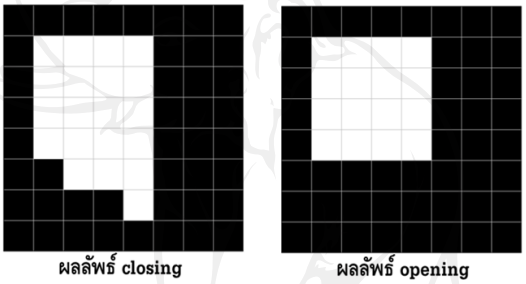
เป็นการแปลงการดำเนินการทางสัณฐานวิทยาที่ดำเนินการตรงข้ามกับโคลสซิง กล่าวคือ การดำเนินการเริ่มจากการดำเนินการอีรอชันเพื่อลดจำนวนจุดของขอบวัตถุก่อน แล้วจึง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ดำเนินการไดเลชันเพื่อเพิ่มขอบของวัตถุ การดำเนินการมีผลทำให้วัตถุเล็กๆ ถูกกำจัด แยกวัตถุที่เชื่อมต่อกันด้วยส่วนบางๆ ออกจากกัน และทำให้วัตถุขนาดใหญ่มีขอบเรียบขึ้น ดังรูป 2.6 (ข)



รูป 2.5 ภาพตัวอย่างก่อนการดำเนินการเชิงสัณฐานวิทยา



(ก) (ข)

รูป 2.6 ผลลัพธ์การดำเนินการโคลสซิง และโอเพนนิง

2.3.4 วิธีการระบุป้ายให้องค์ประกอบที่เชื่อมต่อ (Connected component labeling)

กระบวนการระบุป้ายให้องค์ประกอบที่เชื่อมต่อ คือการจัดให้พิกเซลที่เชื่อมต่อกันและมีค่าความเข้มแสงเดียวกันเป็นพิกเซลของวัตถุชิ้นเดียวกัน กระบวนการดังกล่าวจะกำหนดหมายเลขให้พิกเซลแต่ละตำแหน่ง และเก็บผลลัพธ์ในลาเบลเมตริกซ์ (Label matrix) ถ้าภาพที่ต้องการระบุป้ายให้องค์ประกอบที่เชื่อมต่อเป็นภาพสองระดับ ลาเบลเมตริกซ์ที่เป็นผลลัพธ์ของกระบวนการจะมีขนาดเท่ากัน ทั้งจำนวนแถวและจำนวนหลัก สมาชิกที่อยู่ในภาพนั้นจะมีค่า 0 และ 1 เท่านั้น แต่ข้อมูลที่อยู่ในลาเบลเมตริกซ์นั้นเป็นจำนวนเต็มหรือจำนวนนับตั้งแต่ 1 เรื่อยไป ซึ่งค่าที่อยู่ในแต่ละตำแหน่งของลาเบลเมตริกซ์นั้นจะเท่ากับหมายเลขวัตถุที่พิกเซลของภาพสองระดับในตำแหน่งเดียวกันจัดเป็นสมาชิกอยู่ และจากผลที่ได้สามารถนำไปคำนวณหาคุณสมบัติของวัตถุแต่ละชิ้นในขั้นตอนอื่นๆ ต่อไป

ลักษณะการเชื่อมต่อของพิกเซล โดยทั่วแบ่งออกเป็น 2 ลักษณะ ดังนี้

(1) การเชื่อมต่อแบบ 4 ทิศทาง (4 Connectivity)

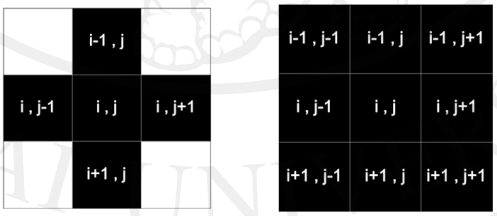
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

หากพิกเซลที่กำลังพิจารณานั้นอยู่ที่ตำแหน่ง i และ j และมีพิกเซลที่มีการเชื่อมต่อแบบ 4 ทิศทางกับพิกเซลดังกล่าว ซึ่งพิกเซลที่มีการต่อเชื่อมอยู่ทั้ง 4 ทิศ คือ ทิศเหนือ ทิศใต้ ทิศตะวันออก และทิศตะวันตก ดังรูป 2.7 (ก)

(2) การเชื่อมต่อแบบ 8 ทิศทาง (8 Connectivity)

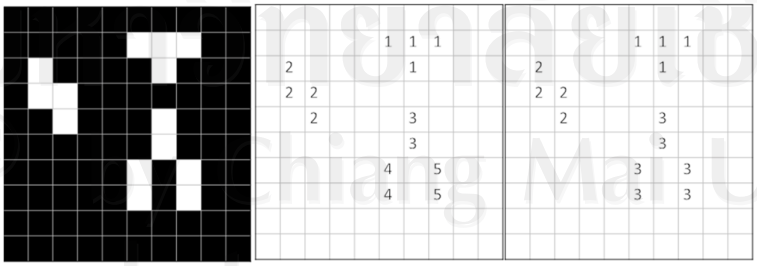
หากพิกเซลที่กำลังพิจารณานั้นอยู่ที่ตำแหน่ง i และ j และมีพิกเซลที่มีการเชื่อมต่อแบบ 8 ทิศทางกับพิกเซลดังกล่าว ซึ่งพิกเซลที่มีการต่อเชื่อมอยู่ทั้ง 8 ทิศ คือ ทิศเหนือ ทิศใต้ ทิศตะวันออก ทิศตะวันตก ทิศตะวันออกเฉียงเหนือ ทิศตะวันตกเฉียงเหนือ ทิศตะวันออกเฉียงใต้ และทิศตะวันตกเฉียงใต้ ดังรูป 2.7 (ข)

ซึ่งการเลือกรูปแบบการเชื่อมต่อแบบ 4 ทิศทางหรือ 8 ทิศทางนั้นจะส่งผลต่อจำนวนวัตถุที่อยู่ในภาพ ตัวอย่างภาพที่ต้องการระบุป้ายให้องค์ประกอบที่เชื่อมต่อ แสดงดังรูป 2.8 (ก) รูป 2.8 (ข) เป็นผลที่ได้จากกระบวนการกำหนดอันดับการเชื่อมต่อพิกเซลที่เลือกใช้การเชื่อมต่อแบบ 4 ทิศทาง พบว่ามีวัตถุอยู่ในภาพถึง 5 ชิ้น ซึ่งตรงกันข้ามกับการเลือกใช้การเชื่อมต่อแบบ 8 ทิศทาง ดังรูป 2.8 (ค) ที่มีจำนวนวัตถุที่อยู่ในภาพนั้นเพียง 3 ชิ้นเท่านั้น อย่างไรก็ตามเวลาที่ใช้สำหรับพิกเซลที่มีการเชื่อมต่อแบบ 8 ทิศทางนั้นจะยาวนานกว่าพิกเซลที่มีการเชื่อมต่อแบบ 4 ทิศทาง ซึ่งการเลือกใช้การเชื่อมต่อแบบใดขึ้นอยู่กับความต้องการของงานแต่ละประเภท (นพณัฐ มันตลักษ์ และวัชระ โต๊ะเงิน, 2553)



(ก) (ข)

รูป 2.7 พิกเซลที่มีการเชื่อมต่อแบบ 4 ทิศทาง และแบบ 8 ทิศทาง



(ก) (ข) (ค)

รูป 2.8 ตัวอย่างการเชื่อมต่อพิกเซลในภาพสองระดับ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.3.5 การปรับให้เรียบด้วยการนับความยาว (Run-Length smoothing algorithm)

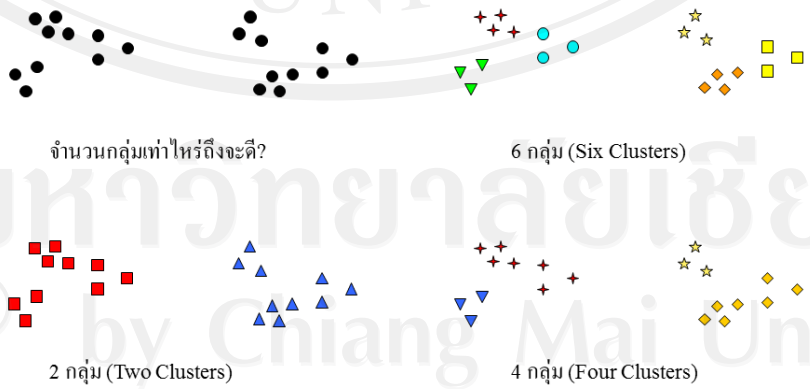
การปรับให้เรียบด้วยการนับความยาวเป็นเทคนิคที่ดำเนินการกับภาพสองระดับ โดยพิจารณาพิกเซลแต่ละแถวจากซ้ายไปขวา และจากบนลงล่าง กำหนดให้ x คือลำดับพิกเซลของภาพสองระดับ และ y คือลำดับพิกเซลของภาพสองระดับที่ถูกกำจัดค่า 0 เมื่อมีพิกเซลติดกันน้อยกว่าค่า C ที่กำหนดไว้ เงื่อนไขการปรับให้เรียบด้วยการนับความยาว ดังสมการ (2.6)

$$I_{run-length}(x,y) = \begin{cases} 1 & if\ I_{bw}(x,y) = 1, \\ & I_{bw}(x,y) = 0\ and\ L \le C \\ 0 & if\ I_{bw}(x,y) = 0\ and\ L > C \end{cases} \quad (2.6)$$

โดยค่า 0 ในลำดับ x จะเปลี่ยนเป็น 1 ในลำดับ y ถ้าพิกเซลที่มีค่า 0 และอยู่ติดกันมีจำนวนน้อยกว่าหรือเท่ากับค่า C ที่กำหนด ส่วนค่า 1 ในลำดับ x จะไม่เปลี่ยนค่าในลำดับ y ตัวอย่างลำดับ x คือ 0001000001010000100000001 1000 และกำหนดให้ C มีค่าเท่ากับ 3 ดังนั้นผลลัพธ์ในลำดับ y คือ 1111000001110000100000001 1111 (K. Y. Wong et al., 1982)

2.4 การจัดกลุ่ม

การจัดกลุ่มแตกต่างจากการจำแนกประเภทข้อมูล โดยการจัดกลุ่มแบ่งกลุ่มข้อมูลจากความคล้าย โดยไม่มีการกำหนดฉลากประเภทข้อมูลไว้ก่อน หรือไม่ทราบจำนวนกลุ่มล่วงหน้า เป็นการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) การจัดวัตถุต่างๆ ให้อยู่ในกลุ่มที่เหมาะสม ดังรูป 2.9 ซึ่งมีคุณสมบัติ คือ วัตถุที่อยู่ในกลุ่มเดียวกันจะคล้ายกัน และมีความแตกต่างจากวัตถุในกลุ่มอื่น การจัดกลุ่มมีหลายวิธีการ เช่น การจัดกลุ่มแบบลำดับขั้น (Hierarchical clustering) การจัดกลุ่มแบบเคมีน (K-means clustering) เป็นต้น



รูป 2.9 ตัวอย่างการจัดกลุ่ม

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ในงานวิจัยนี้ใช้การจัดกลุ่มแบบเคมีน เพื่อสร้างชุดข้อมูลฝึกสอน โดยการแบ่งกลุ่มข้อมูลของวัตถุทั้งหมดออกตามจำนวนกลุ่มที่ต้องการจำนวน K กลุ่ม และค่า K ต้องน้อยกว่าจำนวนของวัตถุทั้งหมด (N) ซึ่งจำนวนกลุ่มต้องเป็นเลขจำนวนเต็มบวก และการจัดกลุ่มวัตถุต้องอาศัยความเหมือนของวัตถุ โดยวัดจากระยะห่างที่น้อยที่สุดระหว่างวัตถุกับจุดศูนย์กลางของกลุ่มทั้งหมด เพื่อจัดวัตถุเข้าสู่กลุ่มต่างๆ ตามจำนวนกลุ่มที่ต้องการ โดยทั่วไปนิยมใช้การวัดระยะแบบยูคลิเดียน (Euclidean distance) และจุดศูนย์กลางเหล่านั้นเป็นตัวแทนของกลุ่มต่างๆ สามารถคำนวณได้จากค่าเฉลี่ยของวัตถุที่อยู่ภายในกลุ่มเดียวกัน ซึ่งวัตถุเหล่านั้นมีคุณลักษณะเหมือนกัน ส่วนวัตถุที่อยู่ต่างกลุ่มก็มีคุณลักษณะที่แตกต่างกัน การจัดกลุ่มแบบเคมีนมีขั้นตอนดังนี้

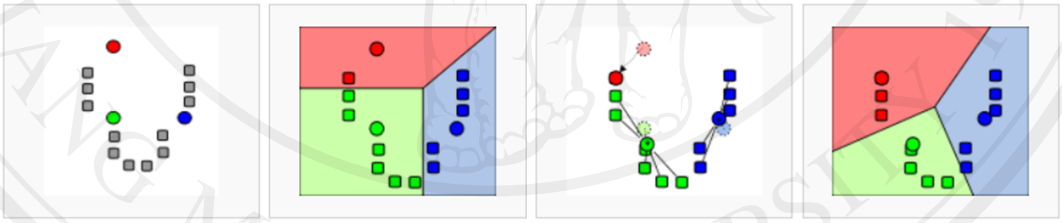
(1) แบ่งกลุ่มข้อมูลเป็น K กลุ่ม ดังรูป 2.10 (ก)

(2) คำนวณจุดกึ่งกลาง (Centroid) ของกลุ่ม โดยใช้ค่าเฉลี่ยเลขคณิต ดังรูป 2.10 (ข)

(3) สำหรับแต่ละข้อมูล นำข้อมูลเทียบกับจุดกึ่งกลาง เพื่อกำหนดกลุ่มให้กับข้อมูล โดยเลือกระยะจากข้อมูลไปจุดกึ่งกลางที่ใกล้ที่สุด ดังรูป 2.10 (ค)

(4) วนซ้ำจนกระทั่งไม่มีการเปลี่ยนกลุ่มของข้อมูล หรือครบจำนวนรอบสูงสุดที่กำหนดไว้ ดังรูป 2.10 (ง)

(นเรศ ผ่องสวัสดิ์กุล และจีรพร วีระพันธุ์, 2552)



(ก) (ข) (ค) (ง)

รูป 2.10 การจัดกลุ่มแบบเคมีน

2.5 การจำแนกประเภท

การจำแนกประเภท เมื่อแบ่งตามชนิดการเรียนรู้สามารถแบ่งได้เป็น 2 ประเภท คือ การเรียนรู้แบบมีผู้สอน และการเรียนรู้แบบไม่มีผู้สอน ซึ่งการจำแนกประเภทอักษรธรรมล้านนานั้นเป็นการเรียนรู้แบบมีผู้สอน โดยเป็นการหาความสัมพันธ์ของข้อมูลโดยอาศัยชุดข้อมูลฝึกสอนที่ประกอบด้วยข้อมูลตัวอย่างหลายๆ ชุด ซึ่งตัวอย่างแต่ละตัวได้รับการกำหนดประเภทของข้อมูลแล้ว เพื่อสอนตัวจำแนกว่าตัวอย่างอยู่ในกลุ่มใด ดังนั้นตัวจำแนกจะอ้างอิงข้อมูลดังกล่าวในกระบวนการจำแนกชุดข้อมูลทดสอบ ซึ่งวิธีการจำแนกประเภทที่เป็นการเรียนรู้แบบมีผู้สอนมีหลายกระบวนวิธี ดังต่อไปนี้

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.5.1 เค-เนียร์เรสเนเบอร์ (k-Nearest neighbor : k-NN)

เค-เนียร์เรสเนเบอร์เป็นกระบวนวิธีจำแนกประเภทที่ถูกใช้งานอย่างแพร่หลาย เนื่องจากเป็นวิธีง่ายและมีความรวดเร็ว โดยใช้ในการรู้จำตัวอักษรในหลายภาษา เช่น อังกฤษ อารบิก สันสกฤต รวมทั้งอักษรธรรมล้านนา

(1) หลักการของเค-เนียร์เรสเนเบอร์

ในการจำแนกประเภทด้วยเค-เนียร์เรสเนเบอร์ ชุดข้อมูลฝึกสอน คือ ชุดของข้อมูลที่อยู่ในปริภูมิ f มิติ ซึ่ง f คือ จำนวนของคุณลักษณะ ส่วนข้อมูลทดสอบ คือ ข้อมูลที่ยังไม่ทราบประเภท และอยู่ในปริภูมิ f มิติเช่นเดียวกับชุดข้อมูลฝึกสอน

กำหนดให้จำแนกประเภทเวกเตอร์ข้อมูลทดสอบ x ด้วยชุดข้อมูลฝึกสอน $Td(x_y, c_y)$ โดยที่ c_y เป็นประเภทของเวกเตอร์ข้อมูลฝึกสอน x_y และ $y = 1,2,\dots,M$ จากนั้นข้อมูลฝึกสอนแต่ละตัวจะถูกคำนวณหาระยะห่างจากข้อมูลทดสอบ จากนั้นพิจารณาเวกเตอร์ข้อมูลฝึกสอนที่ระยะห่างจากข้อมูลทดสอบน้อยที่สุด k อันดับ และตรวจสอบว่าใน k อันดับดังกล่าวเป็นข้อมูลประเภทใดมากที่สุด แล้วให้เวกเตอร์ข้อมูลทดสอบ x จัดอยู่ในประเภทนั้น (A. N. Holambe et al., 2010)

ซึ่งวิธีการวัดระยะทางนั้นมีหลายวิธีการ อาจเป็นการวัดระยะทางแบบยูคลิเดียนเป็นตัววัดที่นิยมใช้มาก และหาได้จากสมการที่ (2.7)

$$dis(x, x_y) = \sqrt{\sum_{i=1}^{f} \left(x_i - x_{y_i}\right)^2} \tag{2.7}$$

(2) งานวิจัยที่เกี่ยวข้องกับเค-เนียร์เรสเนเบอร์

J. Alkhateeb และคณะ ได้ใช้เค-เนียร์เรสเนเบอร์ในการรู้จำคำที่เป็นตัวเขียนอักษรอารบิก โดยเริ่มจากการปรับปรุงคุณภาพของภาพ และตัดแบ่งเอกสารออกเป็นคำ จากนั้นเปลี่ยนขนาดภาพคำให้มีขนาด 45×629 ต่อมาคำจะถูกตัดแยกเป็นบล็อกขนาด 12 พิกเซล แต่ละบล็อกถูกคำนวนหาค่าเฉลี่ยสัมบูรณ์เพื่อใช้เป็นเวกเตอร์คุณลักษณะในการจำแนกประเภท สุดท้ายจำแนกประเภทของคำด้วยเค-เนียร์เรสเนเบอร์ โดยวัดระยะทางแบบยูคลิเดียน ประเภทของคำมีทั้งสิ้น 937 ประเภท มีชุดข้อมูลฝึกสอน 32,492 คำ จากผู้เขียน 1,000 คน เมื่อทดสอบแบบข้ามชุดข้อมูล มีอัตราความถูกต้องในการรู้จำร้อยละ 76.0421 (J. Alkhateeb et al., 2009)

เชาวน์ ปอแก้ว และคณะ เสนอการรู้จำตัวพิมพ์อักษรธรรมล้านนาโดยใช้เค-เนียร์เรสเนเบอร์ด้วยการวัดระยะทางแบบยูคลิเดียน ร่วมกับการจำกัดปัจจัยของข้อมูล เช่น ความไม่คงที่

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

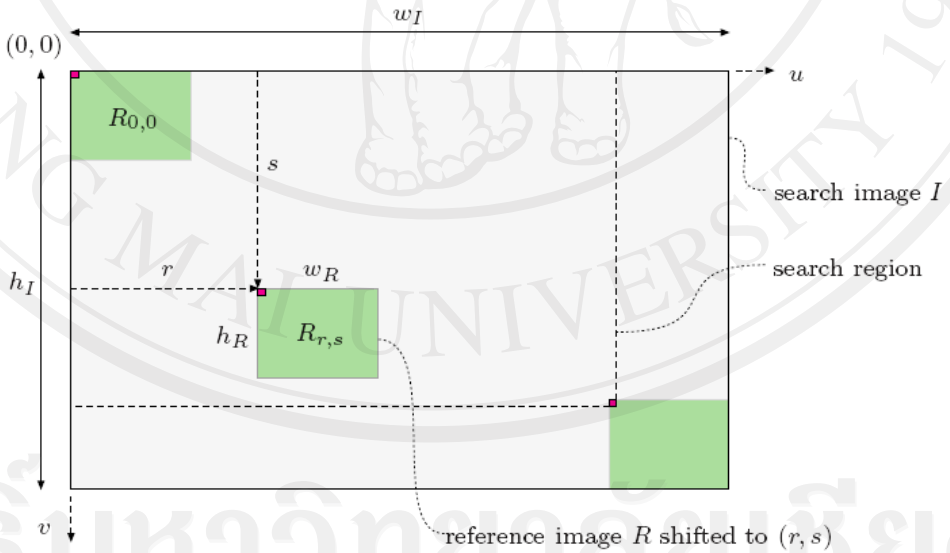
ของการสแกนรูป สัญญาณรบกวน ความไม่สมบูรณ์ของเอกสาร โดยสุ่มข้อมูลตัวอักษรประเภทละ 50 ตัว ทั้งหมด 112 ประเภท กำหนดให้ k มีค่าอยู่ในช่วง 1 ถึง 10 และเวกเตอร์คุณลักษณะคือ อัตราส่วนความกว้างและความสูงของอักษร และโซนนิ่ง ผลการวิจัยพบว่าเมื่อค่า k เท่ากับ 3 มีอัตราการรู้จำของการทดสอบข้ามชุดข้อมูล 5 ชุด ร้อยละ 95 แต่งานวิจัยนี้ยังมีข้อจำกัด เนื่องจากยังไม่สามารถรู้จำอักษรที่มีลายเส้นขาด หรือตัวอักษรที่เส้นเชื่อมติดหลายตัวอักษรได้อย่างถูกต้อง (เชาวน์ ปอแก้ว และคณะ, 2554)

2.5.2 เทมเพลตแมทชิง (Template matching)

เทมเพลตแมทชิงถูกใช้งานอย่างกว้างขวางในงานหลายประเภท เช่น การรู้จำ การตรวจจับวัตถุ การตรวจจับเหตุการณ์ เนื่องจากเป็นวิธีที่ง่ายและมีประสิทธิภาพ

(1) หลักการของเทมเพลตแมทชิง

หลักการของเทมเพลตแมทชิง คือการย้ายตำแหน่งของเทมเพลตวางบนรูปภาพที่ต้องการค้นหา แล้ววัดความคล้ายระหว่างเทมเพลตกับภาพย่อย ณ ตำแหน่งที่อยู่ใต้เทมเพลตนั้น ตำแหน่งที่มีค่าความคล้ายมากที่สุด หมายถึงตำแหน่งที่เทมเพลตและรูปภาพย่อยนั้นคล้ายกันที่สุด ดังรูป 2.11



รูป 2.11 หลักการของเทมเพลตแมทชิง

กำหนดให้ภาพ I คือ ภาพที่ต้องการค้นหา R คือ เทมเพลต และเป้าหมาย คือ การค้นหาตำแหน่งที่เทมเพลต R คล้ายกับภาพย่อยในภาพ I มากที่สุด โดยใช้เทมเพลต R ย้ายตำแหน่งในการวัดความคล้ายจากซ้ายไปขวาและจากบนลงล่าง ดังสมการ (2.8)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

$$R_{r,s}(u,v) = R(u-r, v-s) \tag{2.8}$$

การวัดความคล้ายด้วยการวัดระยะทางระหว่างเทมเพลต R กับภาพย่อย I แต่ละตำแหน่งมีตัววัดหลายแบบ ในที่นี้นำเสนอ 2 แบบ คือ ตัววัดพื้นฐาน และตัววัดสหสัมพันธ์ มีรายละเอียดดังต่อไปนี้

1) ตัววัดพื้นฐาน

การวัดความคล้ายด้วยการวัดระยะทาง โดยใช้ตัววัดพื้นฐานมี 3 ชนิด ดังนี้

- ตัววัดแบบผลรวมของผลต่างสัมบูรณ์ (Sum of absolute differences)

$$d_A(r,s) = \sum_{(i,j)\in R} |I(r+i, s+j) - R(i,j)| \tag{2.9}$$

- ตัววัดแบบผลต่างที่มากที่สุด (Maximum difference)

$$d_M(r,s) = \max_{(i,j)\in R} |I(r+i, s+j) - R(i,j)| \tag{2.10}$$

- ตัววัดแบบผลรวมของรากที่สองของผลต่าง (Sum of squared differences)

$$d_E(r,s) = \left[\sum_{(i,j)\in R} (I(r+i, s+j) - R(i,j))^2 \right]^{1/2} \tag{2.11}$$

2) ตัววัดสหสัมพันธ์

นอกจากนี้ตัววัดชนิดอื่นที่นอกเหนือจากการวัดระยะทาง เช่น สหสัมพันธ์ ซึ่งเป็นตัววัดอีกประเภทหนึ่งที่นิยมใช้ ซึ่งแบ่งได้เป็นหลายประเภท เช่น

- ตัววัดแบบสหสัมพันธ์ไขว้ปรับปรกติ (Normalized cross correlation)

$$C_N(r,s) = \frac{\sum_{(i,j)\in R} I(r+i, s+j) \cdot R(i,j)}{\left[\sum_{(i,j)\in R} I^2(r+i, s+j)\right]^{1/2} \cdot \left[\sum_{(i,j)\in R} R^2(i,j)\right]^{1/2}} \tag{2.12}$$

ค่าตัวเลขของตัววัดแบบสหสัมพันธ์ไขว้ปรับปรกติจะเป็นค่ามาตรฐาน มีการลดความซับซ้อนของข้อมูลด้วยการใช้รากที่สองของพลังงานที่เป็นพื้นที่หลัก ถ้าค่าของภาพ I และเทมเพลต R เป็นบวกทั้งหมด ผลลัพธ์จะอยู่ในช่วง [0,1] กรณีที่ผลลัพธ์เท่ากับ 1 หมายถึง เทมเพลต R มีความคล้ายกับภาพย่อย I มากที่สุด ณ ตำแหน่งนั้น ขณะที่ผลลัพธ์เท่ากับ 0 หมายถึง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

เทมเพลต R มีความคล้ายกับภาพย่อย I น้อยที่สุด ซึ่งตัววัดนี้มีข้อเสียเนื่องจากมีความไวต่อการเปลี่ยนแปลงของค่าแสงสว่าง

- ตัววัดแบบสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient)

การแก้ปัญหาความไวต่อแสง ส่วนใหญ่จะใช้วิธีสัมประสิทธิ์สหสัมพันธ์ ซึ่งคงทนต่อการเปลี่ยนแปลงของแสงมากกว่าตัววัดแบบสหสัมพันธ์ไขว้ปรับปรกติ ดังสมการ (2.13)

$$C_L(r,s) = \frac{\sum_{(i,j)\in R}(I(r+i,s+j) - \bar{I}(r,s)) \cdot (R(i,j) - \bar{R})}{\left[\sum_{(i,j)\in R}(I(r+i,s+j) - \bar{I}_{r,s})^2\right]^{1/2} \cdot \left[\sum_{(i,j)\in R}(R(i,j) - \bar{R})^2\right]^{1/2}} \quad (2.13)$$

$$\bar{I}(r,s) = \frac{1}{K}\sum_{(i,j)\in R} I(r+i,s+j) \quad (2.14)$$

$$\bar{R} = \frac{1}{K}\sum_{(i,j)\in R} R(i,j) \quad (2.15)$$

เมื่อ $\bar{I}(r,s)$ และ $\bar{R}$ แทนค่าเฉลี่ยของความเข้มแสงของภาพ $I(r,s)$ และเทมเพลต R ตามลำดับ ซึ่ง K เป็นจำนวนพิกเซลของภาพเทมเพลต R ผลลัพธ์ของ $C_L(r,s)$ จะอยู่ในช่วง [-1,1] ถ้าผลลัพธ์มีค่าเท่ากับ 1 หมายถึงทั้งสองภาพคล้ายกันมากที่สุด และเป็นพื้นที่ที่อยู่มุมบนซ้าย ณ ตำแหน่งที่ r, s ของภาพที่ค้นหาได้ ส่วนค่า -1 มีความหมายตรงกันข้าม คือมีความคล้ายน้อยที่สุด (W. Burger and M. J. Burge, 2009)

(2) งานวิจัยที่เกี่ยวข้องกับเทมเพลตแมทชิง

C. Kimpan และคณะ ใช้วิธีการแมตชิงในการรู้จำตัวพิมพ์อักษรไทย โดยเริ่มจากการทำให้ตัวอักษรเบลอ เพื่อกำจัดรูปร่างที่ไม่จำเป็นและสัญญาณรบกวน ด้วยการค้นหาข้อมูลที่มีค่าเป็น 1 ในเมตริกซ์ทีละแถว จากแถวบนสุดไปล่างสุด ถ้าพบข้อมูลที่มีค่าเท่ากับ 1 จะเปลี่ยนข้อมูลที่อยู่รอบตำแหน่งนั้นให้มีค่าเป็น 1 ด้วย อักษรที่ถูกเบลอจะถูกแบ่งเป็นกลุ่ม โดยวิธีวัดค่าสัมประสิทธิ์สหสัมพันธ์ และใช้การกระจายคาร์ฮูเนนโลบหารูปแบบมาตรฐาน เพื่อใช้เป็นตัวแทนของแต่ละกลุ่ม ในขั้นตอนการจำแนกใช้วิธีซ้อนทับด้วยเทมเพลตย่อย ผลที่ได้มีอัตราความถูกต้องร้อยละ 98.2 (C. Kimpan et al., 1983)

M. A. Qatran เสนอการรู้จำอักษรมุสนัด ซึ่งเป็นอักษรที่ใช้ในแถบอาระเบียนตอนใต้ โดยใช้วิธีเทมเพลตแมทชิงด้วยการวัดความคล้ายแบบสัมประสิทธิ์สหสัมพันธ์ เริ่มจากการกำจัดขอบส่วนเกินด้านข้างทั้งสี่ด้านของตัวอักษรที่ใช้เป็นเทมเพลต จากนั้นเทมเพลตจะถูกเปลี่ยน

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ให้มีขนาด 42×24 และถูกจัดเก็บในฐานข้อมูล ส่วนภาพสำหรับทดสอบถูกเปลี่ยนให้มีขนาดเดียวกัน และนำไปวัดความคล้ายกับทุกเทมเพลตในฐานข้อมูล โดยได้ทดลองกับความละเอียดภาพ 3 แบบ คือ 512×512 256×256 และ 128×128 โดยสองแบบแรกมีอัตราความถูกต้องในการรู้จำร้อยละ 100 และแบบที่สาม มีอัตราความถูกต้องในการรู้จำร้อยละ 90 (M. A. Qatran, 2011)

2.5.3 คอนดิชันนัลแรนดอมฟิลด์ (Conditional random fields : CRFs)

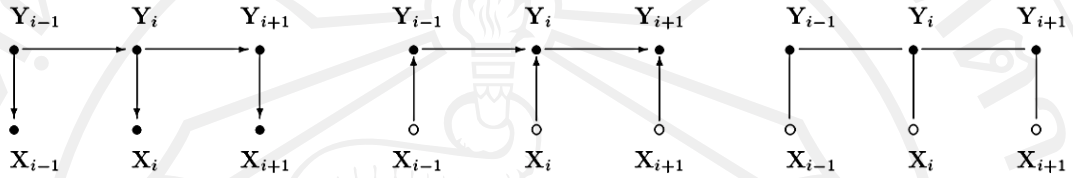
คอนดิชันนัลแรนดอมฟิลด์ เป็นตัวแบบดิสคริมิเนทีฟ (Discriminative model) ที่อาศัยค่าความน่าจะเป็นแบบเงื่อนไขของลำดับของผลลัพธ์ (Label Sequence) แบบต่างๆ เมื่อพบลำดับของข้อมูลที่พบ (Observation sequence) ทำให้จับความสัมพันธ์ระหว่างคุณสมบัติต่างๆ ที่เกี่ยวข้องกันในข้อมูลที่พบได้ (J. Lafferty et al., 2001) ซึ่งเป็นวิธีหนึ่งของการเรียนรู้ของเครื่อง (Machine learning)

(1) หลักการของคอนดิชันนัลแรนดอมฟิลด์

ตัวแบบทางสถิติคอนดิชันนัลแรนดอมฟิลด์ เป็นตัวแบบที่ยอมรับกันในปัจจุบันว่ามีประสิทธิภาพ เนื่องจากตัวแบบฮิดเดนมาร์คอฟ (Hidden markov models : HMMs) ซึ่งมีลักษณะเป็นตัวแบบเจเนอเรทีฟ (Generative model) ที่อาศัยค่าความน่าจะเป็นร่วมระหว่างข้อมูลเข้ากับผลลัพธ์หรือเลเบลที่ออกมา จึงมีปัญหาว่าไม่สามารถจับความสัมพันธ์ระหว่างคุณสมบัติต่างๆ ที่เกี่ยวข้องกันในข้อมูลเข้าได้เพราะคุณสมบัติต่างๆ เป็นอิสระต่อกัน ตัวแบบที่ลดปัญหาดังกล่าวได้คือตัวแบบแมกซิมัมเอนโทรปีมาร์คอฟ (Maximum entropy markov models : MEMMs) ที่มีลักษณะเป็นตัวแบบดิสคริมิเนทิฟที่อาศัยค่าความน่าจะเป็นแบบเงื่อนไขของผลหรือสายของเลเบลแบบต่างๆ เมื่อพบสายข้อมูลเข้า ซึ่งทำให้จับความสัมพันธ์ระหว่างคุณสมบัติต่างๆ ในข้อมูลที่พบได้ แต่ตัวแบบแมกซิมัมเอนโทรปีมาร์คอฟยังพบปัญหา เนื่องจากการตัดสินผล ณ สภาวะใดขึ้นกับสภาวะปัจจุบันและสายข้อมูลที่พบเท่านั้น สภาวะอื่นๆ ทั้งหมดในตัวแบบไม่มีผลต่อการคำนวณค่าความน่าจะเป็น ตัวแบบคอนดิชันนัลแรนดอมฟิลด์จึงเป็นตัวแบบที่ลดปัญหาดังกล่าว เมื่อนำตัวแบบทั้งสามแสดงในรูปกราฟ เพื่อเปรียบเทียบได้ดังรูป 2.12 โดยรูป 2.12 (ก) คือตัวแบบฮิดเดนมาร์คอฟ รูป 2.12 (ข) คือตัวแบบแมกซิมัมเอนโทรปีมาร์คอฟ และรูป 2.12 (ค) คือตัวแบบคอนดิชันนัลแรนดอมฟิลด์ โดยที่ X คือ สายข้อมูลที่พบ และ Y คือ สายของผลลัพธ์ โดยกราฟฮิดเดนมาร์คอฟและแมกซิมัมเอนโทรปีมาร์คอฟ เป็นกราฟระบุทิศทาง (Directed graphical model) โดยฮิดเดนมาร์คอฟ มีลักษณะเป็นตัวแบบเจเนอเรทีฟ คือแสดงการแจกแจงความน่าจะเป็นร่วม P(X, Y) แต่แมกซิมัมเอนโทรปีมาร์คอฟและคอนดิชันนัลแรนดอมฟิลด์ เป็นตัวแบบดิสคริมิเนทีฟ มีลักษณะการแจกแจงความน่าจะเป็นแบบมีเงื่อนไข P(Y|X) อีกทั้งตัวแบบคอนดิชันนัลแรนดอม-

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ฟิลด์ มีลักษณะเป็นตัวแบบกราฟไม่ระบุทิศทาง (Undirected graphical model) จึงต่างจากแมกซิมัมเอนโทรปีมาร์คอฟ เนื่องจากคอนดิชันนัลแรนดอมฟิลด์หาค่าความน่าจะเป็นของเลเบลถัดไปโดยนำเลเบลก่อนหน้าทั้งหมดที่มีลำดับเหตุการณ์เป็นเงื่อนไขมาคำนวณด้วย ทำให้น้ำหนักของคุณสมบัติต่างๆ จากสภาวะที่ต่างกันจึงมีการปรับสมดุลให้ค่าไม่เอนเอียงไปสภาวะใดสภาวะหนึ่ง (C. Sutton and A. Mccallum, 2006)



รูป 2.12 กราฟแสดงการเปรียบเทียบตัวแบบฮิดเดนมาร์คอฟ ตัวแบบแมกซิมัมเอนโทรปีมาร์คอฟ และตัวแบบคอนดิชันนัลแรนดอมฟิลด์

สมการแจกแจงความน่าจะเป็นแบบมีเงื่อนไขของตัวแบบคอนดิชันนัลแรนดอม-ฟิลด์แบบห่วงโซ่ตรง (Linear-chain) อธิบายได้ดังนี้

$$p_\theta = p(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{t=1}^{T} F(y, x, t)\right) \tag{2.16}$$

$$Z(x) = \sum_{y'} exp\left(\sum_{t=1}^{T} F(y', x, t)\right) \tag{2.17}$$

โดยที่

θ คือ พารามิเตอร์ของโมเดล

y คือ ลำดับของลาเบลที่เป็นผลลัพธ์

x คือ ลำดับของข้อมูลเข้าหรือเหตุการณ์

t คือ ตำแหน่งของลำดับของสภาวะที่ต่อเนื่องตั้งแต่ $t_1, t_2, \dots, t_T$

$Z(x)$ คือการปรับข้อมูลให้เข้ากับบรรทัดฐาน (Normalization)

$F(y, x, t)$ คือฟังก์ชันคุณสมบัติ หาได้จากสมการ (2.18)

$$F(y, x, t) = \sum_{l', l, j} \lambda_{l', l}^{j} f_{l', l}^{j}(y_{t-1}, y_t, x) + \sum_{l, j} \beta_l^j g_l^j(y_t, x) \tag{2.18}$$

โดยที่

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

$f^j_{l',l}(y_{t-1}, y_t, x)$ คือ ฟังก์ชันคุณลักษณะแทรนซิชัน (Transition function)

$g^j_l(y_t, x)$ คือ ฟังก์ชันคุณลักษณะสถานะ (State function)

$\lambda^j_{l',l}$ และ β^j_l คือ น้ำหนักคุณลักษณะที่สัมพันธ์กับ $f^j_{l',l}$ และ g^j_l ประมาณได้จากชุดข้อมูลฝึกสอน และกำหนดให้

$$z_j(x,t) = \begin{cases} 1 & \text{if } x \text{ at position } t-k \text{ to } t+k \text{ is equal to } (n,s) \\ 0 & \text{otherwise} \end{cases} \quad (2.19)$$

$$f^j_{l',l}(y_{t-1}, y_t, x) = \begin{cases} z_j(x,t) & \text{if } y_{t-1} = l' \text{ and } y_t = l \\ 0 & \text{otherwise} \end{cases} \quad (2.20)$$

$$g^j_l(y_t, x) = \begin{cases} z_j(x,t) & \text{if } y_t = l \\ 0 & \text{otherwise} \end{cases} \quad (2.21)$$

การจำแนกด้วยคอนดิชันนัลแรนดอมฟิลด์ทำได้โดยหาค่าความน่าจะเป็นของลำดับ y^* จากชุดข้อมูลฝึกสอน ดังสมการ (2.22)

$$\begin{aligned} y^* &= arg\ max_y p_\theta(y|x) \\ &= arg\ max_y \exp\left(\sum_{t=1}^{T} F(y,x,t)\right) \end{aligned} \quad (2.22)$$

เมื่อนำวิธีคอนดิชันนัลแรนดอมฟิลด์ประยุกต์ใช้กับการรู้จำตัวอักษร ทำได้โดยการให้เครื่องเรียนรู้ลักษณะการปรากฏร่วมกันของตัวอักษร โดยต้องกำหนดคุณสมบัติต่างๆ เพื่อช่วยในการเรียนรู้ ซึ่งวิธีนี้ไม่จำเป็นต้องมีผู้เชี่ยวชาญในการวิเคราะห์กฎของภาษา แต่ต้องใช้คลังข้อมูลขนาดใหญ่ในการฝึกสอนแทน (นัชชา ถิระสาโรช, 2553)

(2) งานวิจัยที่เกี่ยวข้องกับคอนดิชันนัลแรนดอมฟิลด์

S. Feng และคณะ ได้เสนอการใช้ตัวแบบคอนดิชันนัลแรนดอมฟิลด์ และฮิดเดน-มาร์คอฟสำหรับการรู้จำเอกสารตัวเขียนโบราณ โดยข้อมูลที่ใช้ คือเอกสารโบราณ จำนวน 20 หน้า ซึ่งมีลักษณะเป็นลายมืออักษรโรมันที่ตัวอักษรติดกันเป็นคำ ทำให้ยากที่จะแยกเป็นตัวอักษรแต่ละตัวได้ จึงใช้ตัวแบบคอนดิชันนัลแรนดอมฟิลด์เพื่อทำนายเป็นคำ โดยจะพิจาณาจากตัวอักษรรอบข้างของคำนั้นด้วย ซึ่งใช้คุณลักษณะแบบต่อเนื่อง (Continuous feature) ที่สกัดจากรูปภาพแต่ละคำมีทั้งสิ้น 27 คุณลักษณะ เช่นความสูง ความกว้าง อัตราส่วนของความกว้างและความสูง พื้นที่ การแปลงฟูเรียร์เต็มหน่วย (Discrete fourier transform) เป็นต้น แล้วแปลงให้เป็นคุณลักษณะแบบไม่ต่อเนื่อง (Discrete feature)ได้ 52 คุณลักษณะ เมื่อทดสอบร่วมกับการค้นหาแบบเคแอลไดเวอร์

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

เจนส์ (K-L Divergence) มีค่าความถูกต้องร้อยละ 42.8 ซึ่งได้ผลลัพธ์ที่มีประสิทธิภาพสูงกว่าแมกซิมัมเอนโทรปี และฮิดเดนมาร์คอฟที่มีค่าความถูกต้องร้อยละ 41.6 และ 33.6 ตามลำดับ โดยที่ใช้คุณลักษณะและชุดข้อมูลเดียวกัน (S. Feng et al., 2006)

นัชชา ถิระสาโรช ได้เสนอการใช้ตัวแบบคอนดิชันนัลแรนดอมฟิลด์ในการรู้จำชื่อเฉพาะภาษาไทย เพื่อเปรียบเทียบประสิทธิภาพของระบบการรู้จำชื่อเฉพาะภาษาไทยระหว่างตัวแบบที่รับข้อมูลเข้าเป็นพยางค์กับที่รับข้อมูลเข้าเป็นคำ โดยใช้คลังข้อมูลข่าวขนาด 367,673 คำ ประกอบด้วยชื่อเฉพาะทั้งหมด 16,179 ชื่อ ซึ่งใช้คุณลักษณะดังนี้ คุณสมบัติรายการชื่อเฉพาะ คุณสมบัติคำย่อ คุณสมบัติคำบริบท คุณสมบัติคำทั่วไป คุณสมบัติค่าทางสถิติ คุณสมบัติยูนิแกรม (Unigram) และไบแกรม (Bigram) การเรียนรู้ของระบบเป็นแบบมีการให้คำตอบในคลังข้อมูลสำหรับฝึกสอน คำตอบที่ใช้มีทั้งหมด 5 แบบ โดยแบบที่ 1 มีข้อมูลขอบเขตของชื่อเฉพาะน้อยที่สุด และแบบที่ 5 มีข้อมูลขอบเขตของชื่อเฉพาะมากที่สุด ทดสอบโดยใช้โปรแกรมซีอาร์เอฟพลัสพลัส พบว่า แบบคำตอบที่ให้ข้อมูลมากช่วยให้ประสิทธิภาพของทั้งสองระบบดีกว่าแบบคำตอบที่ให้ข้อมูลน้อย และประสิทธิภาพของระบบที่ใช้ข้อมูลตัดคำและตัดพยางค์ไม่ต่างกัน โดยมีค่าความถูกต้องเท่ากัน คือร้อยละ 81.30 (นัชชา ถิระสาโรช, 2553)

2.5.4 ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine : SVM)

ซัพพอร์ตเวกเตอร์แมชชีนเป็นเทคนิคหนึ่งที่ได้รับความนิยมอย่างแพร่หลายในงานที่เกี่ยวข้องกับการจดจำรูปแบบ ตลอดจนการแก้ปัญหาการจัดกลุ่ม

(1) หลักการของซัพพอร์ตเวกเตอร์แมชชีน

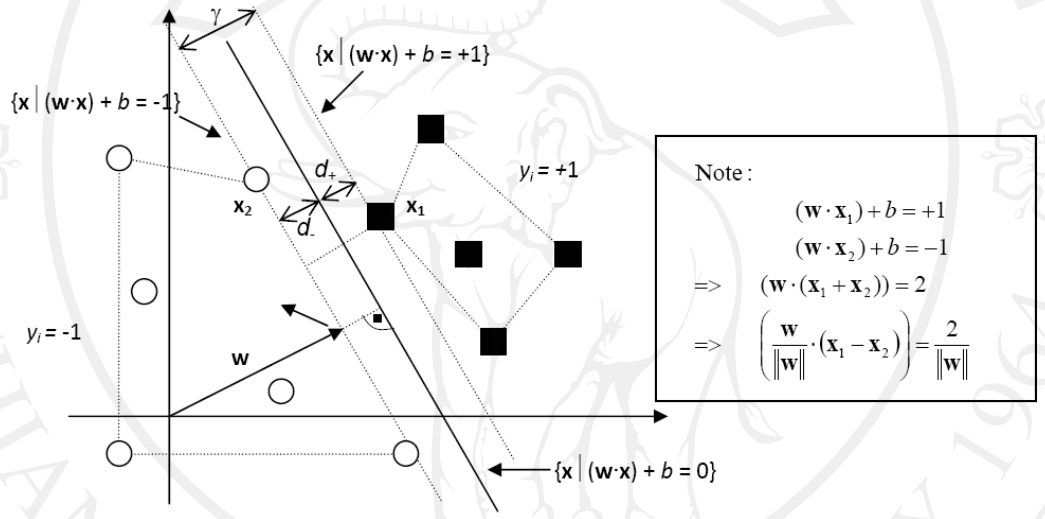
หลักการของซัพพอร์ตเวกเตอร์แมชชีน คือการสร้างไฮเปอร์เพลน (Hyperplane) ที่เหมาะสมบนระนาบของข้อมูลฝึกสอน γ เพื่อแบ่งแยกกลุ่มข้อมูลที่แตกต่างกัน ในการสร้างไฮเปอร์เพลนที่เหมาะสมจะนิยามขอบจากระยะทางระหว่างจุดของข้อมูลที่อยู่ใกล้กับไฮเปอร์เพลนมากที่สุดทั้งสองด้าน คือ d_+ และ d_- โดยระยะขอบ (Margin) γ เกิดจากผลรวมของระยะ d_+ และ d_- ซึ่งไฮเปอร์เพลนที่เหมาะสม คือไฮเปอร์เพลนที่มีค่าขอบ γ กว้างที่สุด แสดงดังรูป 2.13 โดยข้อมูลฝึกสอนที่อยู่บนขอบของ γ จะถูกเรียกว่า ซัพพอร์ตเวกเตอร์

จากรูป 2.13 เป็นการแบ่งกลุ่มข้อมูลแบบ 2 กลุ่ม กำหนดให้กลุ่มข้อมูลที่ใช้ในการฝึกสอน ประกอบด้วย l ตัวอย่าง ซึ่งสามารถแสดงอยู่ในรูป $\{x_k, y_k\}, k = 1, \ldots l$ และ $x_k \in \Re^n, y_k \in \{-1, +1\}$ โดย x_k เป็นเวกเตอร์นำเข้า ในขณะที่ y_k เป็นประเภทของข้อมูล หลักการของซัพพอร์ตเวกเตอร์แมชชีน คือการสร้างไฮเปอร์เพลนที่เหมาะสมบนระนาบของข้อมูลตัวอย่าง ซึ่งไฮเปอร์เพลนดังกล่าวจะถูกกำหนดโดยพารามิเตอร์(w,b) โดย w เป็นเวกเตอร์ที่ตั้งฉาก

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

กับไฮเปอร์-เพลน และ b เป็นค่าคงที่ ซึ่งกำหนดตำแหน่งของเวกเตอร์ที่สัมพันธ์กับตำแหน่งดั้งเดิมในปริภูมิข้อมูลนำเข้า (Input space) โดยสมการของไฮเปอร์เพลนแบบเชิงเส้นถูกกำหนดโดยสมการ $(\mathrm{w}\cdot\mathrm{x})+b=0$ และเพื่อลดปัญหาในเรื่องของสเกล w และ b จะถูกกำหนดด้วยสมการ $|\mathrm{w}\cdot\mathrm{x}|+b=1$ สำหรับจุดที่อยู่ใกล้ไฮเปอร์เพลนมากที่สุด ดังนั้นสามารถแสดงสมการของไฮเปอร์เพลนได้ดังสมการ 2.23

$$y_i[(\mathrm{w}\cdot\mathrm{x}_i)+b]\geq 1 \tag{2.23}$$



รูป 2.13 การแบ่งกลุ่มข้อมูลแบบ 2 กลุ่มด้วยซัพพอร์ตเวกเตอร์แมชชีน

จากที่กล่าวมาข้างต้นเป็นการแบ่งกลุ่มข้อมูลด้วยไฮเปอร์เพลนแบบเชิงเส้นเท่านั้น ดังนั้นเพื่อให้อัลกอริทึมดังกล่าวสามารถแบ่งแยกกลุ่มข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น (Nonlinear dataset) จำเป็นต้องแปลงกลุ่มข้อมูลตัวอย่างไปสู่ปริภูมิมิติที่สูงขึ้น (Higher dimensional space) ซึ่งเรียกว่าปริภูมิคุณลักษณะ (Feature space) โดยการแปลงดังกล่าวจะกระทำผ่านฟังก์ชันที่ไม่เป็นเชิงเส้น (Nonlinear function) โดยสามารถแสดงสมการที่ใช้คำนวณค่าไฮเปอร์เพลน เพื่อแบ่งกลุ่มข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น ได้ดังสมการ 2.24

$$W(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathrm{x}_i,\mathrm{x}_j) \tag{2.24}$$

$$\sum_{i=1}^{l}\alpha_i y_i=0$$

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

$$0 \leq \alpha_i \leq C \quad \forall i$$

โดยตัวแปร $\alpha_i \leq 0$ ถูกเรียกว่า Positive Lagrange Multipliers ซึ่ง $K(\mathrm{x}_i, \mathrm{x}_j)$ คือ ฟังก์ชันเคอร์แนล (Kernel function) ที่ใช้ในการแปลงข้อมูลไปสู่ปริภูมิสูงขึ้น และ C เป็นค่าคงที่เพื่อใช้ในการปรับหรือชดเชยระหว่างค่าผิดพลาดของการฝึกสอนและความซับซ้อนของแบบจำลอง จากสมการ 2.24 สามารถแสดงฟังก์ชันเคอร์แนลที่นิยมใช้กันโดยทั่วไปได้ดังสมการที่ 2.25 ถึง 2.27 โดย γ β และ d คือพารามิเตอร์ของเคอร์แนล

- โพลิโนเมียลดีกรี d (Polynomial of degree d)

$$K(\mathrm{x}, \mathrm{y}) = (\gamma(\mathrm{x} \cdot \mathrm{y}) + \beta)^d \tag{2.25}$$

- เรเดียลเบสิสฟังก์ชัน (Radial basis function : RBF)

$$K(\mathrm{x}, \mathrm{y}) = \exp(-\gamma \|\mathrm{x} - \mathrm{y}\|^2) \tag{2.26}$$

- ซิกมอยด์ฟังก์ชัน (Sigmoid function)

$$K(\mathrm{x}, \mathrm{y}) = \tanh(\gamma(\mathrm{x} \cdot \mathrm{y} + \beta)) \tag{2.27}$$

(สุพจน์ จันทร์วิพัฒน์, 2554)

(2) งานวิจัยที่เกี่ยวข้องกับซัพพอร์ตเวกเตอร์แมชชีน

A. N. Holambe และคณะ ได้เปรียบเทียบการรู้จำตัวอักษรและตัวเลขระหว่างการใช้เค-เนียร์เรสเนเบอร์ และซัพพอร์ตเวกเตอร์แมชชีน สำหรับตัวพิมพ์และตัวเขียนภาษาเทวนาครี ซึ่งเป็นภาษาที่ใช้ในประเทศอินเดีย เมื่อนำภาพผ่านกระบวนการปรับปรุงภาพและตัดแบ่งเป็นตัวอักษร ได้ประเภทชุดข้อมูลฝึกสอนทั้งหมด 55 ประเภท และชุดข้อมูลทดสอบทั้งหมด 1,500 ตัวอักษร และ 10,000 ตัวเลข ผลการทดสอบพบว่า การรู้จำด้วยซัพพอร์ตเวกเตอร์แมชชีนมีอัตราความถูกต้องในการรู้จำสูงกว่าเค-เนียร์เรสเนเบอร์ทั้งอักษรตัวพิมพ์และตัวเขียน โดยมีอัตราเฉลี่ยสูงกว่าร้อยละ 94 (A. N. Holambe et al., 2010)

สุพจน์ จันทร์วิพัฒน์ ได้ประยุกต์ใช้งานซัพพอร์ตเวกเตอร์แมชชีนในการแก้ปัญหาการจัดกลุ่มข้อมูลผ่านการพัฒนาโปรแกรมจดจำลายมือเขียนพยัญชนะภาษาไทย โดยเริ่มจากการจัดเตรียมข้อมูล การฝึกสอน การสร้างแบบจำลอง และการทดสอบแบบจำลองเพื่อหาประสิทธิภาพในการจำแนกกลุ่มข้อมูล ข้อมูลตัวอักษรที่ใช้ในการทดลอง คือพยัญชนะลายมือเขียน โดยบุคคล

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ต่างๆ ตัวอักษรละ 40 ตัวอย่าง รวมเป็นพยัญชนะที่ใช้ในการฝึกสอนทั้งสิ้น 1,760 ตัวอักษร ข้อมูลที่ใช้ในการทดสอบมาจากการเก็บข้อมูลลายมือเขียนจากกลุ่มผู้ทดสอบที่แตกต่างกัน โดยเก็บจำนวนตัวอักษรละ 20 ตัวอย่าง รวมจำนวนพยัญชนะที่ใช้ในการทดสอบจำนวนทั้งสิ้น 880 ตัวอักษร ซึ่งผลการทดลองพบว่า ซัพพอร์ตเวกเตอร์แมชชีนสามารถจำแนกกลุ่มข้อมูลพยัญชนะในแต่ละกลุ่มได้อย่างถูกต้องโดยมีอัตราเฉลี่ยสูงกว่าร้อยละ 86 (สุพจน์ จันทร์วิพัฒน์, 2554)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved