



รายงานการวิจัย

เรื่อง

การจำแนกความหมายของภาพโดยใช้โครงสร้างสเกตริตรอน

Semantic Images Classification with Structure Skeleton

โดย

นศัพ์ชาณัณ ชินปัญชัณนะ

มหาวิทยาลัยธุรกิจบัณฑิตย์

รายงานผลการวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2552

## กิตติกรรมประกาศ

งานวิจัยชิ้นนี้สำเร็จลุล่วงได้ทั้งนี้เพราะได้รับความอนุเคราะห์ การสนับสนุน และผลักดัน รวมถึงติดตามงานของผู้วิจัยอย่างต่อเนื่อง ได้แก่ รองศาสตราจารย์ ดร.สรชัย พิศาลบุตร และ รองศาสตราจารย์ ดร.นุชรี เปรมชัยสวัสดิ์ รวมทั้งอาจารย์อีกหลายท่านที่ให้ความความรู้ และคำแนะนำ ทางด้านเนื้อหาในงานวิจัย ได้แก่ รองศาสตราจารย์ ดร.โกสินทร์ จ่านงไทย ผู้ช่วยศาสตราจารย์ ดร. บัณฑิต ทิพากร ผู้ช่วยศาสตราจารย์ ดร.ทรงฤทธิ์ มณีวงศ์วัฒนา และงานวิจัยนี้จะไม่สมบูรณ์ได้ หากไม่ได้ข้อเสนอแนะที่เป็นประโยชน์จากท่านอาจารย์ผู้ทรงคุณวุฒิ ผู้ซึ่งให้คำแนะนำที่ดีและมุมมองที่ ผู้วิจัยได้นำมาปรับปรุงให้สมบูรณ์ยิ่งขึ้น

นอกจากนี้ใคร่ขอขอบคุณมหาวิทยาลัยธุรกิจบัณฑิตฯ ผู้ให้ทุนสนับสนุนงานวิจัยชิ้นนี้ และ ขอขอบคุณคำแนะนำในการทำงานวิจัยจากคณาจารย์ในคณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจ บัณฑิตฯ หากงานวิจัยเล่มนี้มีข้อผิดพลาด ประการใดขออภัยไว้ ณ ที่นี้ด้วยค่ะ

นศพัชฌัน ชินปัญญาชนะ

กุมภาพันธ์ 2553

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	(1)
สารบัญตาราง	(3)
สารบัญรูปภาพ	(4)
<b>บทที่ 1 บทนำ</b>	<b>1</b>
1.1 ความเป็นมาของปัญหา	1
1.2 วัตถุประสงค์	2
1.3 สมมุติฐาน	2
1.4 ขอบเขตของการวิจัย	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ	3
1.6 นิยามคำศัพท์	4
<b>บทที่ 2 ทฤษฎีและงานวิจัย</b>	<b>5</b>
2.1 ทฤษฎีเบื้องต้น	5
2.2 การประมวลผลภาพดิจิทัล	8
2.3 การคัดเลือกข้อมูล	9
2.4 การจำแนกข้อมูลภาพ	11
2.5 การใช้โมเดลในการสร้างการตัดสินใจ	24
2.6 การวัดประสิทธิภาพ	25
<b>บทที่ 3 ขั้นตอนวิธีการดำเนินงานวิจัย</b>	<b>27</b>
3.1 ขั้นตอนการเตรียมข้อมูล	27
3.2 ขั้นตอนการประมวลผล	30
3.3 ขั้นตอนการวัดประสิทธิภาพ	32

บทที่ 4 ผลการทดลอง	34
4.1 ผลการคัดเลือกข้อมูล	34
4.2 ผลการจำแนกความหมายของข้อมูลภาพ	37
4.2.1 ผลการจำแนกข้อมูลด้วยวัตถุ	38
4.2.2 ผลการจำแนกข้อมูลภาพด้วยการรวมกันของพีเจอร์	41
บทที่ 5 สรุปผลการทดลอง	48
5.1 สรุปและอภิปรายผลการทดลอง	48
5.2 ข้อเสนอแนะ	49
บรรณานุกรม	51
ประวัติผู้วิจัย	56

## สารบัญตาราง

ตารางที่	หน้า
2.1 ตารางแสดงการวัดประสิทธิภาพ	26
4.1 แสดงผลค่าความถูกต้องในการจำแนกข้อมูลภาพด้วยการเรียงลำดับพีเจอร์ของวัตถุ ( $O_i$ )	35
4.2 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ในชุดข้อมูลแบบ 7 พีเจอร์	39
4.3 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ในชุดข้อมูลแบบ 7 พีเจอร์	39
4.4 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ในชุดข้อมูลแบบ 7 พีเจอร์	39
4.5 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ในชุดข้อมูลแบบ 9 พีเจอร์	40
4.6 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ในชุดข้อมูลแบบ 9 พีเจอร์	40
4.7 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ในชุดข้อมูลแบบ 9 พีเจอร์	40
4.8 ตารางแสดงผลค่าความถูกต้องของพีเจอร์ต่าง ๆ	41
4.9 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ด้วยชุดข้อมูล $O_i^7$ และ $p_i$	43
4.10 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ด้วยชุดข้อมูล $O_i^7$ และ $p_i$	43
4.11 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ด้วยชุดข้อมูล $O_i^7$ และ $p_i$	43
4.12 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ด้วยชุดข้อมูล $O_i^7$ และ $s_i$	44

- 4.13 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ด้วยชุดข้อมูล  $O_i^7$  และ  $s_i$  44
- 4.14 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ด้วยชุดข้อมูล  $O_i^7$  และ  $s_i$  44
- 4.15 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ด้วยชุดข้อมูล  $O_i^7$ ,  $p_i$  และ  $s_i$  45
- 4.16 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ด้วยชุดข้อมูล  $O_i^7$ ,  $p_i$  และ  $s_i$  45
- 4.17 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ด้วยชุดข้อมูล  $O_i^7$ ,  $p_i$  และ  $s_i$  45

## สารบัญรูปภาพ

รูปภาพที่		หน้า
2.1	แสดงคุณลักษณะพื้นฐานที่มีความแตกต่างในภาพตัวอย่างหมวดหมู่ชายทะเล	6
2.2	เซลล์ระบบประสาทของสิ่งมีชีวิตและโครงข่ายประสาทเทียม	14
2.3	แสดงตัวแบบของ Neuron ในคอมพิวเตอร์	15
2.4	แสดงสถาปัตยกรรมของ feedforward network	16
2.5	แสดงสถาปัตยกรรมของ feedback network	16
2.6	แสดงโครงสร้างแบบป้อนไปข้างหน้าหลายชั้น	17
2.7	แสดงการเรียนรู้แบบมีการสอน	18
2.8	แสดงการเรียนรู้แบบไม่มีการสอน	19
2.9	แสดงหลักการทำงานของซัพพอร์ตเวกเตอร์แมชชีน	21
2.10	แสดงเส้นแบ่งข้อมูลด้วยซัพพอร์ตเวกเตอร์แมชชีน	23
3.1	ขั้นตอนการจัดกลุ่มความหมายภาพ	27
3.2	แสดงขั้นตอนการให้ความหมายกับภาพ	29
3.3	แสดงโครงสร้างสเกตริตรอน	29
3.4	แสดงภาพที่ถูก Mapping ด้วยโครงสร้างสเกตริตรอน	30
4.1	แสดงผลค่าความถูกต้องในการจำแนกข้อมูลภาพด้วยการเรียงลำดับพีเจอร์ของ วัตถุ ( $O_i$ )	36
4.2	ผลลัพธ์ของการจำแนกความหมายภาพในแต่ละกลุ่ม	47

ชื่อเรื่อง : การจัดกลุ่มความหมายของภาพโดยใช้โครงสร้างสเกตรี้ตรอน  
ผู้วิจัย : นศัพธ์ชาณัณ ชินปัญชรณะ สถาบัน : มหาวิทยาลัยธุรกิจบัณฑิตย์  
ปีที่พิมพ์ : พุทธศักราช 2553 สถานที่พิมพ์ : มหาวิทยาลัยธุรกิจบัณฑิตย์  
แหล่งที่เก็บรายงานการวิจัยฉบับสมบูรณ์ : ศูนย์วิจัยมหาวิทยาลัยธุรกิจบัณฑิตย์  
จำนวนหน้าวิจัย : 59 หน้า ลิขสิทธิ์ : มหาวิทยาลัยธุรกิจบัณฑิตย์  
คำสำคัญ : เทคโนโลยีสารสนเทศธุรกิจ, คอมพิวเตอร์ธุรกิจ

## บทคัดย่อ

วิวัฒนาการของเครื่องมือและอุปกรณ์ถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็วจนทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้น ปัญหาที่ตามมาก็คือการจัดเก็บข้อมูลภาพที่เพิ่มมากขึ้นอย่างไร้ขีดจำกัดนี้ จะทำด้วยวิธีการใดจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถค้นคืนข้อมูลภาพ และจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้มีงานวิจัยมากมายที่พยายามจะคิดค้นกระบวนการ อัลกอริทึมต่าง ๆ เพื่อทำให้เกิดความพึงพอใจของผู้ใช้มากที่สุด แต่อย่างไรก็ตาม การสืบค้นหรือการจำแนกข้อมูลภาพส่วนใหญ่มักจะคำนึงเพียงแต่ความเหมือนกันของวัตถุที่ปรากฏบนภาพเท่านั้น ซึ่งยังไม่เพียงพอต่อการแปลความหมายที่แท้จริง

งานวิจัยนี้จึงนำเสนอวิธีการจำแนกความหมายของภาพ โดยใช้หลักทฤษฎีการเรียนรู้ภาพของมนุษย์ที่เรียกว่า โครงสร้างสเกตรี้ตรอน เพื่อทำการเชื่อมโยงประสานกับวัตถุบนภาพให้สอดคล้องตามความหมายของกระบวนการคิดแปลความหมายภาพของมนุษย์ โดยข้อมูลทั้งหมดถูกคัดเลือกโดยใช้กระบวนการคัดเลือกข้อมูลเพื่อให้ได้ฟีเจอร์ที่มีคุณสมบัติที่ดีที่สุดเข้ามาทำการทดลอง จากผลการทดลองสามารถจำแนกภาพด้วยวิธี โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าหลายชั้น ได้รับความถูกต้องถึง 94.8%



**Title** : Semantic Images Classification with Structure Skeleton

**Researcher** : Nutchanan Chinpanthana

**Institution** : Dhurakijpundit University.

**Year of Publication** : 2010

**Publisher** : Dhurakijpundit University.

**Sources** : Dhurakijpundit University Research Center.

**Number of Pages** : 59 Pages

**Copyright** : Dhurakijpundit University.

**Keyword** : Business Information Technology, Business Computer

## **Abstract**

Searching images their semantic is an active problem in multimedia image retrieval. Many researchers have attempted to improve semantic models by using high level concept based on keyword annotation. However, the annotation is tedious, inconsistent, and erroneous. The retrieval process of such approaches is done by keyword searching. This model is rather rudimentary and it does not specific enough for representing the actual meaning.

In this paper, we present a technique of the semantic image classification by using the human perception. The structure skeleton is used to combine the object components and image meaning. The feature selection methods are introduced to select the essential features from existing features. The experimental results indicate that our proposed approach offers significant performance improvements in the interpretation of semantic image classification, compare with other features, with the maximum of 94.8%

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาของปัญหา

ในสมัยก่อนการประมวลผลภาพนั้น จะอยู่ในรูปของการประมวลผลสัญญาณแอนะล็อก (analog) โดยใช้อุปกรณ์จับภาพแตงภาพที่ถ่ายออกมายังคงเป็นภาพขาวดำและไม่ชัดเจนเท่าที่ควร ทำให้ส่วนใหญ่ภาพถ่ายมักจะถูกถ่ายเฉพาะภาพเหตุการณ์ที่สำคัญหรือบุคคลสำคัญเท่านั้น เนื่องจากอุปกรณ์ที่นำมาใช้ในการถ่ายภาพ หรือบันทึกภาพนั้นซับซ้อนและมีความยุ่งยากในการใช้งาน รวมทั้งราคาสูง แต่อย่างไรก็ตามได้มีการพัฒนาปรับปรุงเทคโนโลยีเพื่อตอบรับกับความต้องการที่เพิ่มขึ้น ทำให้อุปกรณ์ถ่ายภาพในปัจจุบันมีราคาถูกลง และสะดวกต่อการใช้งานมาก และพัฒนาต่อเนื่องไปเป็นอุปกรณ์ถ่ายภาพแบบดิจิทัล (digital) ทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้นอย่างต่อเนื่อง จึงทำให้เกิดปัญหาในการจัดเก็บข้อมูลภาพรวมถึงการค้นหาภาพ (image retrieval) ที่เพิ่มมากขึ้นอย่างรวดเร็ว จะทำอย่างไรจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถที่จะค้นหาภาพที่ต้องการได้อย่างถูกต้อง

การประมวลผลภาพดิจิทัล (digital image processing) เป็นสาขาที่ได้รับความนิยมในการวิจัยค่อนข้างมาก เพราะสามารถที่จะนำไปประยุกต์ได้หลายด้าน เช่น ทางด้านการแพทย์ ทางด้านอุตสาหกรรม หรือนำมาประยุกต์ใช้ในส่วนของการจราจร ในการตรวจจับทะเบียนรถยนต์ เป็นต้น สิ่งทีกล่าวมาข้างต้นเป็นเพียงส่วนหนึ่งที่น่าการประมวลผลภาพดิจิทัลเข้ามาปรับปรุงและไปประยุกต์ใช้งาน ปัจจุบันมีงานวิจัยหลายกลุ่ม [Qian Huang,1995][Aditaya Vailaya,2001][W. Ma,1997][C. Carson,1999] พยายามนำเทคนิคต่างๆเข้ามาใช้ในการค้นหาภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ ไม่ว่าจะเป็นการสืบค้นข้อมูลโดยใช้ การประมวลผลภาพระดับต่ำ (low level image processing) เป็นการนำเอาคุณลักษณะของภาพ หรือ ฟีเจอร์ (feature) ต่าง ๆ ของภาพที่ถูกสกัด (extraction) อย่างเช่น สี (color) ลวดลาย (texture) หรือ รูปทรง (shape) ด้วยวิธีการอัลกอริทึมต่าง ๆ เพื่อใช้ในการสืบค้นข้อมูล แต่อย่างไรก็ตามได้มีงานวิจัยอีกกลุ่มที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้น ที่เรียกว่า การประมวลผลภาพระดับสูง (high level image processing) งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ

(object) ที่มีความหมาย [Benitez A.B,2002] และแทนวัตถุอื่น ๆ ด้วยคำศัพท์ (keyword) บนภาพ และใช้ความหมายหรือคำศัพท์นั้นเพื่อทำการสืบค้นข้อมูลแทน ซึ่งสิ่งที่ได้เมื่อเปรียบเทียบแล้วจะ ได้ผลที่ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าอัลกอริทึมที่นำมาใช้นั้นจะเป็นลักษณะใด ยกตัวอย่างเช่น การใช้เทคนิคความสัมพันธ์ของแท็กชื่อวัตถุบนภาพ [Benitez A.B,2001][ R. Zhao ,2002][ Philippe Mulhem,2002] ซึ่งเป็นการสอดคล้องกันด้วยความหมายตามพจนานุกรม การใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำ (synonym) เช่น “stone” มีความหมาย สอดคล้องกันกับ “rock” เป็นต้น การค้นหาภาพด้วยเทคนิคแบบนี้จะได้ผลลัพธ์ที่ค่อนข้างขึ้นกับ คำศัพท์ที่ถูกให้ความหมายไว้ในภาพส่วนใหญ่ เพราะฉะนั้นผลลัพธ์ของการค้นคืนภาพคือการค้นหา คำศัพท์ที่อยู่ในภาพนั่นเอง จึงทำให้ความหมายภาพที่แท้จริงจะไม่สามารถเกิดขึ้นได้เลย นอกจากเป็น การค้นหาตาม คำศัพท์ที่ถูกให้ความหมายขึ้นมาเท่านั้น ไม่ว่าจะมีการใช้เทคนิคการค้นหาภาพดี อย่างไรก็ตามแต่วิธีการนั้นยังไม่ได้คำนึงถึง ความหมายที่เกิดขึ้นของภาพนั้นอย่างแท้จริง (semantic image) เป็นเพียงการค้นหาคำศัพท์ที่ถูกกำหนดไว้เท่านั้น เพราะฉะนั้นการวิจัยครั้งนี้สนใจเฉพาะการ รวมกันของคำที่อยู่บนภาพเพื่อนำมาใช้ในการจัดกลุ่มความหมายของภาพโดยรวม เพื่อให้ได้ ความหมายของภาพอย่างแท้จริง โดยนำเอาทฤษฎีการรับรู้ภาพของมนุษย์ที่มีการพิจารณาจาก ตำแหน่งและขนาดของวัตถุที่เกิดขึ้นเพื่อนำมาใช้ในการแปลความหมายของภาพโดยรวม

## 1.2 วัตถุประสงค์

- เพื่อค้นหาเทคนิคใหม่ที่สามารถนำมาใช้ในการจำแนกความหมายของภาพ
- เพื่อลดจำนวนพีเจอร์ในการการจำแนกความหมายของภาพ
- เพื่อสร้างแบบจำลองวิเคราะห์การจำแนกความหมายของภาพ
- เพื่อสำรวจพฤติกรรมการมองวัตถุและการแปลความหมายของมนุษย์บนภาพ
- เพื่อพัฒนาเทคนิคในการแปลความหมายภาพ

## 1.3 สมมติฐาน

งานวิจัยนี้จึงเสนอวิธีการจำแนกความหมายของภาพ โดยใช้หลักทฤษฎีการรับรู้ภาพของ มนุษย์ที่มีการพิจารณาดำแหน่งของวัตถุ (object position) และขนาดของวัตถุ (object size) บน ภาพและนำมาผ่านโครงสร้างสเกตริตรอน (structure skeleton framework) เพื่อทำการเชื่อมโยง ประสานกับวัตถุบนภาพให้สอดคล้องตามความหมายของกระบวนการคิดแปลความหมายภาพของ

มนุษย์ และได้ทำการจำแนกกลุ่มภาพความหมายของภาพ โดยใช้วิธีการจำแนกภาพ (classification) เปรียบเทียบผลการทดลองทั้งหมด 3 วิธีการ naive-Bayes, multiple feedforward และ supporting vector machine

## 1.4 ขอบเขตของการวิจัย

ข้อมูลภาพที่นำมาวิเคราะห์เป็นข้อมูลภาพดิจิทัล เฉพาะภาพธรรมชาติ (natural images) เช่น ภาพทิวทัศน์ ชายทะเล ภูเขา หรือ สวนสาธารณะ ภาพภายในอาคาร ภายในบ้าน สำนักงาน หรือ โรงงาน เป็นต้น และเป็นภาพที่มีวัตถุเกิดปรากฏขึ้น สำหรับภาพที่ไม่มีความหมายหรือเป็นภาพที่มีความหมายกำกวม หรือภาพที่สามารถแปลได้หลายความหมาย หรือภาพที่โฟกัสระยะใกล้ จะไม่นำภาพนั้นเข้ามาวิเคราะห์สำหรับการหาความหมายภาพ

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

### 1.5.1 ผลต่อสังคม

1. เพื่อนำเอากระบวนการนี้มาประยุกต์ใช้กับการค้นหาภาพบนห้องสมุดดิจิทัลได้
2. เพื่อได้แนวทางการแปลและการตีความหมายภาพ
3. ช่วยให้การใช้คำศัพท์ในการค้นหาข้อมูลภาพได้ผลตามความหมายของภาพมากขึ้น

### 1.5.2 ผลต่อมหาวิทยาลัย

1. สร้างกลุ่มนักวิจัยที่เป็นลักษณะของสาขาทางด้าน Images Processing ในสาขาย่อย Semantic Image Processing
2. สร้างชื่อเสียงให้กับมหาวิทยาลัย เมื่อมีบทความวิจัยลงในวารสารต่างประเทศ บทความวิจัยที่นำเสนอในการประชุมระดับชาติ และนานาชาติ

### 1.5.3 ผลต่อกลุ่มผู้วิจัย

1. พบแนวทางในการคิดค้นสิ่งใหม่ๆ ในการจัดกลุ่มและกระบวนการแปลความหมายของภาพ
2. สามารถนำผลวิจัยมาเขียนบทความลงวารสารนานาชาติ และร่วมประชุมวิชาการระดับชาติและนานาชาติ
3. พัฒนาความรู้ใหม่ๆ หลักการแนวคิดใหม่ ให้เกิดขึ้นในกลุ่มของนักวิจัย

## 1.6 นิยามคำศัพท์

1. การประมวลผลภาพ (Image processing) หมายถึง การนำภาพมาผ่านกระบวนการเพื่อประมวลผลสัญญาณบนสัญญาณ 2 มิติ เช่น ภาพนิ่ง (ภาพถ่าย) หรือภาพวีดิทัศน์ (วีดิโอ) และนำมาใช้งาน
2. การจำแนกประเภทข้อมูลภาพ (Image Classification) หมายถึง การแยกแยะภาพที่มีคุณลักษณะเดียวกันออกเป็นกลุ่ม ๆ
3. การจำแนกความหมายของภาพ (Semantic Image Classification) หมายถึง การนำภาพมาแบ่งเป็นกลุ่มโดยพิจารณาจากความหมายโดยรวมของภาพ
4. การค้นคืนข้อมูล (Image Retrieval) หมายถึง การค้นหาข้อมูลของภาพจากระบบฐานข้อมูล
5. วัตถุ (Object) หมายถึง ส่วนของวัตถุนภาพ ยกตัวอย่างเช่น คน, ต้นไม้ เป็นต้น
6. คำศัพท์ (Keyword) หมายถึง คำที่มีความหมายได้ใจความใช้แทนวัตถุนภาพ
7. แท็ก (Tag) หมายถึง การกำหนดคำ หรือ คำศัพท์บนภาพ หรือเรียกว่าการ labeled ภาพด้วยคำศัพท์
8. คุณลักษณะข้อมูล (Feature) หมายถึง เป็นฟีเจอร์ หรือตัวแปร ที่ถูกสกัดออกมาจากภาพ เช่น สี (color) ลวดลาย (texture) หรือ รูปทรง (shape) รวมทั้ง วัตถุ ที่ปรากฏบนภาพเพื่อนำมาใช้ในการสืบค้นข้อมูลต่อไป
9. โครงข่ายประสาทเทียม (Artificial neural network) หมายถึง เทคนิคการพยากรณ์เหมาะสำหรับลักษณะข้อมูลซึ่งไม่ใช่เชิงเส้น

## บทที่ 2

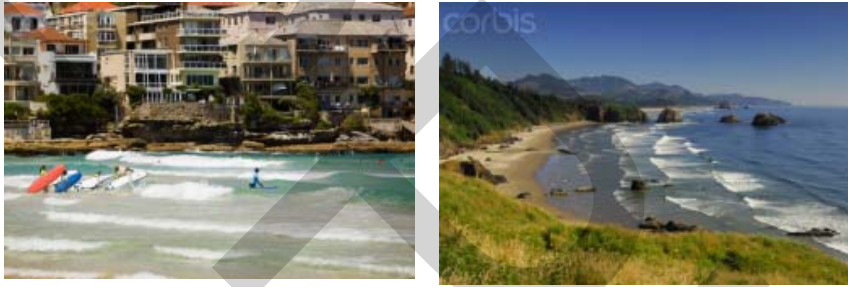
### ทฤษฎีและงานวิจัย

#### 2.1 ทฤษฎีเบื้องต้น

วิวัฒนาการของเครื่องมือและอุปกรณ์ถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็วจนทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้น ปัญหาที่ตามมาก็คือการจัดเก็บข้อมูลภาพที่เพิ่มมากขึ้นอย่างไร ซึ่ดจำกัดนี้จะทำอย่างไรจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถสืบค้นข้อมูลภาพ และจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้งานวิจัยในปัจจุบันที่เกี่ยวข้องกับการค้นคืนรวมทั้งการจัดกลุ่มภาพให้ตรงกับความต้องการเพิ่มมากขึ้น รวมไปถึงระยะเวลาในการสืบค้นที่น้อยลงกับปริมาณของภาพที่เพิ่มทวีคูณ ดังนั้นปัญหาดังกล่าวมาข้างต้นนั้นจึงได้รับความสนใจจากนักวิจัยหลายกลุ่ม ซึ่งเป็นงานด้านการประมวลผลภาพ (image processing) ด้านการค้นคืนสารสนเทศ (image retrieval) เป็นอีกหนทางหนึ่งในการแก้ปัญหาดังกล่าว เนื่องจากสามารถช่วยให้การค้นหาข้อมูลกระทำได้โดยสะดวกยิ่งขึ้นรวมถึง การค้นหาภาพและการจำแนกประเภทข้อมูลภาพ (image classification) เพื่อคัดเลือกภาพ เพื่อให้ตรงตามความต้องการของผู้ใช้ ยิ่งมีความจำเป็นอย่างยิ่ง

ในงานวิจัยทางการประมวลผลภาพในการค้นคืนสารสนเทศกลุ่มแรก ๆ จะมีการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่างๆ ยกตัวอย่างเช่น สี (color) ลวดลาย (texture) รูปร่าง (shape) เป็นต้น กระบวนการนี้ถูกเรียกว่า การประมวลผลภาพระดับต่ำ (low-level image processing) [Jain A.K,1996][Cheng Y.C,2003] กระบวนการนี้สามารถค้นหาภาพได้ตามคุณลักษณะพื้นฐานที่นำไปสืบค้น โดยภาพผลลัพธ์ส่วนใหญ่มักจะเป็นภาพที่มีคุณลักษณะไม่ซับซ้อนมากนัก เช่น โทนสี หรือรูปร่างที่แตกต่างกันอย่างเด่นชัด ดังภาพที่ 2.1 จะสังเกตุว่าภาพเป็นภาพที่อยู่ในหมวดหมู่เดียวกัน คือ ภาพชายหาด หรือ ชายทะเล แต่มีคุณลักษณะของโทนสี (color) และลวดลาย (texture) ที่แตกต่างกันอย่างเห็นได้ชัดเจน เมื่อนำมาจำแนกด้วยการประมวลผลภาพระดับต่ำ (low-level feature) แล้วนั้นค่อนข้างยากที่จะจัดให้หมวดหมู่เดียวกัน แต่อย่างไรก็ตามได้มีกลุ่มนักวิจัยที่พยายามดัดแปลงอัลกอริทึมเพื่อโดยใช้การประมวลผลภาพระดับต่ำเพื่อค้นคืนภาพให้มีลักษณะที่ใกล้เคียงกับภาพที่ต้องการมากที่สุด [Andrew, 2005][M. Flickner,

1995][W. Ma, 1997] การปรับปรุงเทคนิควิธีการเพื่อให้กระบวนการสืบค้นภาพได้ผลลัพธ์ที่ถูกต้องเพิ่มมากขึ้น ด้วยการนำวิธีการมาผสมผสานกันระหว่างคุณลักษณะเพื่อให้สามารถวิเคราะห์ในรูปแบบที่ซับซ้อนได้มากขึ้น เช่นการรวมเทคนิคการใช้สีและรูปร่างของภาพเพื่อทำการค้นหา [ P.S. Hiremath, 2007] หรือมีการใช้อัลกอริทึมเพื่อสกัดข้อมูลภาพเป็นฟีเจอร์ เพื่อนำมาใช้ในการสืบค้นภาพในรูปแบบที่แตกต่างกันออกได้ แต่ในความเป็นจริงแล้วนั้นลักษณะการมองภาพของคนโดยทั่วไปเป็นการมองจากความหมายของภาพ หรือมองจากชนิดของวัตถุของภาพเช่น ภาพที่ 2.1 เป็นในหมวดหมู่ ชายทะเล หรือชายหาด บางภาพอาจจะประกอบด้วย ทะเล และ ท้องฟ้า เพียงอย่างเดียว แต่บางภาพอาจจะประกอบด้วย มนุษย์ สุนัข หรืออาคาร เป็นบางส่วนมารวมด้วย ก็เป็นภาพประเภทเดียวกัน เป็นต้น



ภาพที่ 2.1 แสดงคุณลักษณะพื้นฐานที่มีความแตกต่างในภาพตัวอย่างหมวดหมู่ชายทะเล<sup>1</sup>

เพราะฉะนั้นการที่จะใช้คุณลักษณะและวิธีการค้นคืนด้วยข้อมูลระดับต่ำแล้วนั้นจะผลลัพธ์ของความหมายภาพจะขาดหายไป ได้มีกลุ่มนักวิจัยหันมาสนใจในเรื่องของ การจัดกลุ่มภาพตามความหมายเรียกว่า semantic image เป็นการจัดกลุ่มภาพตามความหมายของภาพโดยแปลความหมายของภาพจากองค์ประกอบ หรือวัตถุ (object) ที่ปรากฏในภาพนั้น ซึ่งมีวิธีการค้นคืนที่แตกต่างกันเพื่อให้ได้ผลลัพธ์ที่ตรงตามความต้องการมากที่สุด

ปัจจุบันมีวิจัยหลายกลุ่ม [Zhao, 2001] [Benitez, 2002] [Kobus, 2001][Philippe, 2002] การพยายามนำเทคนิคต่าง ๆ เข้ามาใช้ในการค้นคืนข้อมูลภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ เช่นการแท็ก (tag)หรือการให้ความหมายของวัตถุบนภาพเป็น ชื่อวัตถุ หรือคำศัพท์ ที่สอดคล้อง

<sup>1</sup> The corbis stock: <http://pro.corbis.com>,  
The corel corporation: <http://www.corel.com>

กัน หรือเป็นการใช้เทคนิคหาความสัมพันธ์ของแท็กชื่อวัตถุบนภาพ ซึ่งเป็นการสอดคล้องกันด้วยความหมายตามพจนานุกรมใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำ (synonym) ยกตัวอย่างเช่น “stone” มีความหมายสอดคล้องกันกับ “rock” เป็นต้น [Vasileios, 2003] การค้นหาภาพด้วยเทคนิคแบบนี้จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกแท็กไว้บนภาพยังมีการแท็กข้อมูลบนภาพมากยิ่งขึ้นสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่ในความเป็นจริงแล้ว การแท็ก (tag) ข้อมูลบนภาพในปัจจุบันนั้นเป็นเพียงการหาคำศัพท์ที่ต้องการบนภาพ แต่ไม่ได้ให้ความหมายภาพโดยรวม ความหมายของภาพคือการนำวัตถุที่ปรากฏบนภาพมารวมกันเพื่อวิเคราะห์จากความคิดของมนุษย์เพื่อให้ได้ผลลัพธ์คือคำศัพท์ใหม่ที่แทนความหมายของภาพทั้งภาพ สำหรับการวิเคราะห์ความหมายของภาพจะได้จากการรับรู้ของมนุษย์ที่มองภาพนั้น จะมีกฎเกณฑ์ของการรับรู้จากการมองภาพได้กล่าวไว้เป็นทฤษฎีที่แน่นอน [Rudolph, 1974] เพราะฉะนั้นการวิจัยครั้งนี้ได้ทำการวิเคราะห์ในส่วนของการการจำแนกความหมายของภาพโดยรวม เพื่อให้ได้ความหมายของภาพอย่างแท้จริง โดยใช้ทฤษฎีการมองของมนุษย์ที่มีการพิจารณาจากตำแหน่งและขนาดของวัตถุที่เกิดขึ้นเพื่อนำมาใช้ในการแปลความหมายของภาพ

ในงานวิจัยนี้จึงได้เสนอวิธีการจัดกลุ่มความหมายของภาพ โดยใช้หลักทฤษฎีการจับจุดเด่นของการมองภาพของมนุษย์ [Rudolph, 1974] จากหลักทฤษฎีดังกล่าวได้มีการพิจารณาดำแหน่งของวัตถุและขนาดของวัตถุบนภาพประมวลผลเข้าด้วยกัน เรียกว่าโครงสร้างสเกตริตรอน (structure skeleton framework) ซึ่งจะถูกนำมาเชื่อมโยงประสานกับวัตถุบนภาพเพื่อให้สอดคล้องตามความหมาย กระบวนการคิดแปลความหมายภาพของมนุษย์ และได้นำโครงสร้างสเกตริตรอนไปใช้ในการจัดรูปแบบเพื่อให้สอดคล้องกับกระบวนการคิดในการแปลความหมายภาพ และในการทดลองได้มีการจำแนกภาพด้วยวิธีการดังต่อไปนี้ naïve-Bayes, artificial neural network, supporting vector machine โดยที่ข้อมูลฟีเจอร์ (feature) ที่ได้ถูกคัดเลือกจากอัลกอริทึม Chi-square และ Information gain ในงานวิจัยนี้ได้นำทฤษฎีมาใช้ในการพิจารณาดังนี้

- การประมวลผลภาพดิจิทัล (digital image processing)
- การคัดเลือกข้อมูล (data selection หรือ feature selection)
- การจำแนกประเภทข้อมูล (image classification)
- การใช้โมเดลในการสร้างการตัดสินใจ
- การวัดประสิทธิภาพ (evaluation)



## 2.2 การประมวลผลภาพดิจิทัล

การประมวลผลภาพดิจิทัล (digital image processing) เป็นการเรียกใช้ขั้นตอนหรือกระบวนการที่มากกระทำบนภาพ โดยมีวัตถุประสงค์เพื่อปรับปรุงคุณภาพของภาพ ให้ได้ภาพใหม่ที่มีคุณสมบัติตามวัตถุประสงค์ที่นำไปใช้งาน เช่น การปรับให้ภาพมีความคมชัดมากขึ้น (enhancement) หรือการบีบอัดข้อมูลภาพ (compression) เพื่อประหยัดพื้นที่ในการจัดเก็บข้อมูล หรือการฝังลายน้ำ (watermark) เพื่อป้องกันการลักลอบการใช้ภาพที่ไม่ได้รับอนุญาตเป็นต้น สำหรับการประมวลผลภาพระดับสูงด้วยคอมพิวเตอร์ สามารถทำได้โดย นำภาพที่ได้มาจากกล้องหรือ image source ต่าง ๆ ซึ่งเป็นสัญญาณอนาล็อก แล้วนำมาแปลงเป็นสัญญาณดิจิทัลที่มีลักษณะเป็นรหัสเชิงตัวเลขฐานสอง (binary) ประกอบด้วยตัวเลข 0 และ 1 ที่สามารถใช้รูปแบบทางคณิตศาสตร์เข้ามาช่วยในการคำนวณและการประมวลผลข้อมูลภาพ การดำเนินงานวิจัยสำหรับการจัดกลุ่มความหมายของภาพได้มีผู้วิจัยจำนวนมากที่ศึกษาและทำการทดลองโดยทั่วไป แบ่งเป็น 2 ระดับ [A. Gupta, 1997] ดังนี้

### 2.2.1 การประมวลผลภาพระดับต่ำ

การประมวลผลภาพระดับต่ำ (low-level image processing) เป็นการประมวลผลขั้นแรกหลังจากที่ได้ภาพเข้ามาในเครื่องคอมพิวเตอร์ ภาพที่ได้ก็จะประกอบด้วยองค์ประกอบที่ต้องการและรวมถึงสิ่งที่ไม่ต้องการ เรียกว่าสัญญาณรบกวน (noise) เป็นสัญญาณที่ทำให้ภาพที่ได้มีคุณภาพไม่ดี และทำให้ภาพนั้นยังไม่สามารถที่จะนำไปใช้ประมวลผลได้ ดังนั้น การประมวลผลภาพในระดับต่ำจึงประกอบไปด้วยการกำจัดสัญญาณรบกวน การทำภาพให้ชัด (highpass filter) การหาขอบภาพ (edge detection) การแปลง binary image สิ่งที่สำคัญสำหรับการประมวลผลระดับต่ำนั้นก็คือการใช้ อัลกอริทึมหรือขบวนการประมวลผลภาพเพื่อแบ่งแยกรูปร่างวัตถุ (image segmentation) หรือ ข้อมูลภายในภาพออกจากกันเป็นส่วนตามวัตถุประสงค์ของการนำไปใช้งาน เช่น การค้นคืนภาพที่โครงสร้างวัตถุเป็นทรงกลม หรือ การค้นคืนภาพที่มีสีเป็นสีแดง เป็นต้น ข้อมูลที่นำมาใช้จะถูกเรียกว่า ฟีเจอร์ (feature) ของภาพที่ถูกลักษณะจะมีได้หลายแบบยกตัวอย่างเช่น สี (color) ลวดลาย (texture) หรือ รูปร่าง (shape) ซึ่งปัจจุบันการประมวลผลภาพระดับต่ำถูกพัฒนาไปด้วยวิธีการหลากหลายมากขึ้นทั้งมีการรวมฟีเจอร์ เพื่อให้สามารถทำงานบรรลุวัตถุประสงค์ หรือการพัฒนาโปรแกรมอัลกอริทึมต่าง ๆ เพื่อให้สามารถสกัดฟีเจอร์จากภาพได้สมบูรณ์ยิ่งขึ้นก็ตาม การวิจัยในสิ่งเหล่านี้ยังมีข้อด้อยและยังไม่สามารถที่จะทำได้สมบูรณ์เท่าที่ควร

### 2.2.2 การประมวลผลภาพระดับสูง

การประมวลผลภาพระดับสูง (high-level image processing) เป็นการใช้ข้อมูลพีเจอร์หรือผลลัพธ์จากการประมวลผลข้างต้นเพื่อผ่านกระบวนการ หรืออัลกอริทึมทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพ (image understanding) รวมทั้งการหาความหมายของภาพ (semantic image) ได้ เพราะฉะนั้นการประมวลผลภาพระดับสูงจำเป็นต้องใช้ข้อมูลที่ได้มาจากการประมวลผลภาพระดับต่ำ ดังนั้นจะเห็นได้ว่าการประมวลผลภาพระดับต่ำมีความสำคัญมากสำหรับการทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพได้ โดยส่วนใหญ่กลุ่มนักวิจัยพยายามที่จะข้ามขั้นตอนการประมวลผลภาพระดับต่ำไป เนื่องจากยังไม่มีคุณสมบัติเท่าที่ควรและพยายามที่จะใช้การประมวลผลภาพระดับสูงอย่างเดียว แต่อย่างไรก็ตามการวิจัยทั้งสองกลุ่มนี้ยังคงมีการพัฒนาการวิจัยอย่างต่อเนื่องเพื่อนำผลลัพธ์ หรือพีเจอร์ต่างๆเข้ามาทำการค้นคืนภาพ (image retrieval) หรือ การจำแนกข้อมูลภาพ (image classification) รวมทั้งการแยกแยะความหมายของภาพ (semantic image classification) ซึ่งเป็นหัวข้อหลักในการทำวิจัยครั้งนี้

## 2.3 การคัดเลือกข้อมูล

การคัดเลือกข้อมูล (data selection หรือ feature selection) เป็นการคัดเลือกข้อมูล หรือพีเจอร์ที่มีประสิทธิภาพมาใช้ในการประมวลผลข้อมูล เนื่องจากข้อมูลที่ถูกลูกกัด (extraction) จากรูปภาพมีจำนวนของข้อมูลมากและข้อมูลบางตัวไม่จำเป็นที่จะต้องใช้ เพราะฉะนั้นต้องมีการคัดเลือกข้อมูลที่ดีเพื่อ เพิ่มประสิทธิภาพการทำนาย เพื่อการสังเคราะห์โมเดลได้อย่างรวดเร็ว และเพื่อลดความซับซ้อนของรูปแบบโมเดล ที่จะพิจารณาต่อไปจึงได้นำวิธี การคัดเลือกพีเจอร์มาใช้ทั้งหมด 2 วิธี

### 2.2.1 การคัดเลือกพีเจอร์แบบ ไคส์-สแควร์

การคัดเลือกพีเจอร์แบบ ไคส์-สแควร์ (chi-squared:  $\chi^2$ ) [Liu, 1995] เป็นการคัดเลือกข้อมูลหรือพีเจอร์โดยใช้การเปรียบเทียบค่าของ feature กับ  $X_{ij}$  ที่ได้จากการคำนวณของกลุ่มของข้อมูลทั้งหมด เพื่อทำการหาค่าคุณสมบัติของพีเจอร์ ในแต่ละตัวว่ามีความสำคัญมากหรือน้อยกว่ากัน ค่าสุดท้ายที่ได้จะสามารถบอกลำดับความสำคัญของ feature ได้ มีสูตรการคำนวณดังนี้

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(A_{ij} - X_{ij})^2}{X_{ij}}, \text{ เมื่อ } X_{ij} = \frac{R_i \times C_j}{N},$$

โดยที่  $m$  คือ จำนวนช่วงข้อมูล,  $k$  คือ จำนวนกลุ่มข้อมูล,  $A_{ij}$  คือ จำนวนกลุ่มข้อมูล ตัวอย่างที่  $i$  กลุ่มที่  $j$ ,  $R_i$  คือ จำนวนของข้อมูลในช่วงที่  $i$ ,  $C_j$  คือ จำนวนของข้อมูลในช่วงที่  $j$ ,  $N$  คือจำนวนรวมของข้อมูลทั้งหมด, ดังนั้น  $X_{ij}$  the expected frequency ของ  $A_{ij}$  ผลลัพธ์ที่ได้ ออกมาค่า  $\chi^2$  ที่สูงสุดแสดงว่าพีเจอร์ ตัวนั้นจะมีความสำคัญมากที่สุดในกลุ่มของข้อมูล

### 2.2.3 การคัดเลือกพีเจอร์แบบ Information Gain ratio

Information Gain ratio (IGR) [ R.C. Holte, 1993] เป็นการคัดเลือกข้อมูลหรือพีเจอร์โดยใช้การลดค่าของ entropy จากกลุ่มย่อยของข้อมูล (cluster) IGR ถูกนำมาใช้เป็นเครื่องมือเพื่อช่วยในการคัดเลือกพีเจอร์ที่ดีที่สุด ใน อัลกอริทึมแบบ C4.5 เป็นรูปแบบของ decision tree IGR แทนด้วย  $gain\_r(x, C)$  คือค่าของ IGR ที่มีข้อมูลแอทริบิวต์  $x$  ในกลุ่มย่อย  $C$  สามารถกำหนดสูตรของ IGR ได้ดังนี้

$$gain\_r(x, C) = \frac{gain(x, C)}{Split(C)},$$

$$gain(x, C) = entropy(x, C) - entropy_p(x, C),$$

$$entropy(x, C) = \frac{-p(x|C)\log_2 p(x|C) - (1-p(x|C))\log_2(1-p(x|C))}{1},$$

$$p(x|C) = \frac{freq(x, C)}{|C|},$$

$$entropy_p(x, C) = \sum_i \frac{|c_i|}{|C|} entropy(x, c_i),$$

$$Split(C) = -\sum_{i=1} \frac{|c_i|}{|C|} \log_2 \frac{|c_i|}{|C|},$$

เมื่อ  $freq(x, C)$  คือความถี่ที่เกิดขึ้นของข้อมูล  $x$  ใน  $C$ , กลุ่มย่อย  $c_i$  คือเซตย่อยที่  $i$  ใน  $C$ ,  $|c_i|$  คือจำนวนของข้อมูลทั้งหมดใน  $c_i$  เมื่อ  $C$  คือกลุ่มข้อมูล ผลลัพธ์ที่ได้ออกมาค่าของ IGR ที่สูงสุดแสดงว่าพีเจอร์นั้นจะมีความสำคัญมากที่สุดในกลุ่มของข้อมูล เพราะฉะนั้นการที่นำมาพีเจอร์ใดเข้ามาใช้ในการทดลองจึงควรที่จะหาความสำคัญของพีเจอร์ก่อนเสมอ ผลที่ได้จากการใช้งานจะช่วยลดความซับซ้อนได้และไม่เปลืองค่าที่ใช้งานจริง

## 2.4 การจำแนกข้อมูลภาพ

การจำแนกประเภทข้อมูล (image classification) เป็นการนำเอาข้อมูลทั้งหมดที่เก็บรวบรวมมาได้ มาผ่านกระบวนการขั้นตอนการแยกแยะข้อมูลลงในแต่ละกลุ่มที่จัดไว้ โดยในแต่ละกลุ่มของข้อมูลนั้นจะมีคุณลักษณะเด่นของแต่ละกลุ่มที่แตกต่างกัน ขึ้นกับข้อมูล หรือ ฟีเจอร์ที่เก็บรวบรวมมาได้รวมทั้งกระบวนการหรือวิธีการที่ใช้จำแนกข้อมูล ซึ่งในงานวิจัยนี้ได้ใช้วิธีการแยกประเภทข้อมูลของภาพเพื่อเปรียบเทียบความสามารถ ทั้งหมด 3 วิธี

- ทฤษฎีความเชื่อเบย์ (bayesian approach)
- โครงข่ายประสาทเทียม (artificial neural network)
- ซัพพอร์ตเวกเตอร์แมชชีน (support vector machines)

### 2.4.1 ทฤษฎีความเชื่อเบย์

ทฤษฎีความเชื่อเบย์ (bayesian approach) [Richard, 2001] เป็นการเลือกรูปแบบของการประมาณค่าของความน่าจะเป็นที่จะเกิดเหตุการณ์ โดยอาศัยความสัมพันธ์ที่เกิดขึ้นของข้อมูลเพื่อมาคำนวณเหตุการณ์ที่เกิดขึ้น ดังนั้นทฤษฎีความเชื่อเบย์จึงอยู่บนพื้นฐานของการประมาณค่าความน่าจะเป็นที่จะเกิดขึ้น ในเบื้องต้นจะกล่าวถึงทฤษฎีพื้นฐานของความน่าจะเป็น (basic formulas for probability) ดังนี้ ถ้ากำหนดให้ A และ B เป็นเหตุการณ์ที่เป็นอิสระต่อกัน คือเหตุการณ์ A และ B ไม่มีความเกี่ยวข้องกัน สามารถเขียนสมการได้

$$p(A \cap B) = p(A) \cdot p(B)$$

โดยที่  $p(A \cap B)$  หมายถึงความน่าจะเป็นที่เหตุการณ์ A และ B เกิดขึ้นพร้อมกัน แต่ถ้ากำหนดให้เหตุการณ์ A และ B เป็นเหตุการณ์ที่ขึ้นต่อกัน

$$p(A \cap B) = p(A) \cdot p(B|A)$$

$$p(A \cap B) = p(A) \cdot p(A|B)$$

$$p(A) \cdot p(B|A) = p(A) \cdot p(A|B)$$

เมื่อแทนค่าความน่าจะเป็นของ A โดยที่เหตุการณ์ B ได้เกิดขึ้นแล้ว จึงสามารถเขียนสูตรของเบย์ ได้ดังนี้

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)}$$

- วิธีการเรียนรู้แบบอย่างง่าย

วิธีการเรียนรู้แบบอย่างง่าย (naïve-bayes learning) [Sonka, 2007][Svoboda, 2007] เป็นวิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็น ซึ่งมีพื้นฐานมาจากทฤษฎีของเบย์ (bayes theorem) เข้ามาช่วยในการเรียนรู้ จุดมุ่งหมายก็เพื่อต้องการสร้างโมเดลที่อยู่ในรูปของความน่าจะเป็น ซึ่งเป็นค่าที่บันทึกได้จากการสังเกต จากนั้นนำโมเดลมาหาว่าสมมติฐานใดถูกต้องที่สุดโดยใช้ความน่าจะเป็นเข้ามาช่วยความรู้ก่อนหน้า หมายถึง ความรู้ที่เรามีเกี่ยวกับสมมติฐานแต่ละตัวก่อนที่เราจะเก็บข้อมูล เมื่อใช้งานเราจะนำความน่าจะเป็นของข้อมูลที่เก็บได้มาปรับสมมติฐานซ้ำอีกครั้ง เมื่อ  $C$  แทนกลุ่มข้อมูลที่จะถูกแบ่ง ประกอบด้วยกลุ่มที่ถูกสังเกต  $c$   $X \langle X_1, X_2, \dots, X_k \rangle$  แทนแอตทริบิวต์เวกเตอร์ข้อมูล ประกอบด้วยข้อมูลที่สังเกต  $x \langle x_1, x_2, \dots, x_k \rangle$  สามารถเขียน  $X = x$  โดยย่อได้ดังนี้

$$X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_k = x_k$$

สำหรับค่าผิดพลาดของการทำนายเป็น

$$\operatorname{argmax}_c (p(C = c | X = x))$$

สามารถเขียนสมการเบย์ได้ดังนี้

$$p(C = c | X = x) = \frac{p(C = c)p(X = x | C = c)}{p(X = x)},$$

$$p(C = c | X = x) \propto p(C = c)p(X = x | C = c),$$

โดยที่ค่าของ  $p(C = c)$  และ  $p(X = x | C = c)$  จะถูกประมาณค่าจากการเรียนรู้ซึ่งในความเป็นจริงแล้วไม่สามารถที่จะประมาณค่าของ  $p(X = x | C = c)$  ได้โดยตรง ดังนั้นเมื่อค่าของ  $X_1, X_2, \dots, X_k$  เป็นเหตุการณ์ที่สามารถเกิดขึ้นต่อกัน แล้ว

$$p(X = x | C = c) = p(\bigwedge_{i=1}^k X_i = x_i | C = c)$$

$$p(X = x | C = c) = \prod_{i=1}^k p(X_i = x_i | C = c)$$

โดยทั่วไปนิยมเขียนเป็น

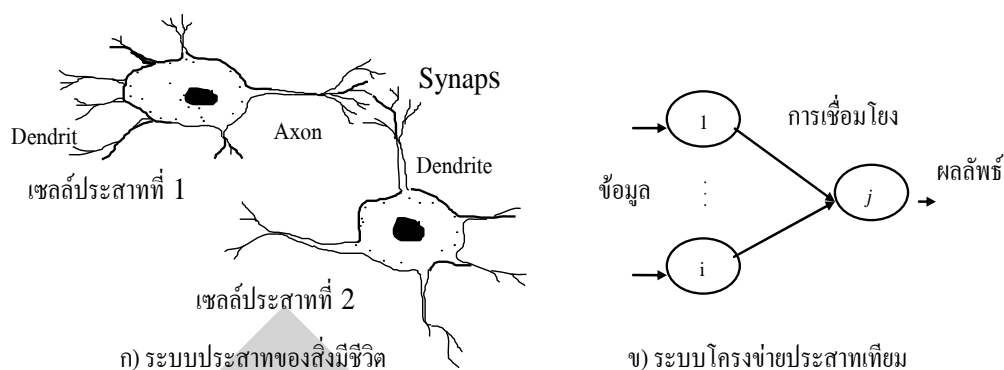
$$p(C = c | X = x) \propto p(C = c) \prod_{i=1}^k p(X_i = x_i | C = c)$$

- ข้อดีของวิธีการเรียนรู้แบบเบย์

เบย์ เป็นวิธีการเรียนรู้ โดยใช้ ความรู้ก่อนหน้า (prior knowledge) เข้ามาช่วยในการเรียนรู้ ได้ซึ่งพบว่าวิธีนี้ให้ประสิทธิภาพในการเรียนรู้ได้ดีไม่ด้อยกว่าวิธีการเรียนรู้ประเภทอื่น สามารถลดข้อจำกัดอย่างง่ายในสมมุติฐานของความไม่ขึ้นต่อกันระหว่างคุณสมบัติ แต่ในความเป็นจริงคุณสมบัติบางตัวจะขึ้นต่อกัน และควรนำค่าความขึ้นต่อกันนี้เข้ามาใส่ไว้ในโมเดล จึงใช้ข่ายงานเบย์ในการอธิบายความไม่ขึ้นต่อกันอย่างมีเงื่อนไข (condition independent) ระหว่างตัวแปร เพื่อให้กระบวนการเรียนรู้มีประสิทธิภาพ โดยสามารถใส่ความรู้ก่อนในข่ายงานความเชื่อเบย์ให้อยู่ในรูปโครงสร้างข่ายงานและตารางความน่าจะเป็นแบบมีเงื่อนไข

#### 2.4.2 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (artificial neural network) [Flood, I.,1994][ Lippmann, R.P., 1988] หรือที่มักจะเรียกสั้น ๆ ว่า ข่ายงานประสาท (neural network หรือ neural net) เป็นโมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้รูปการจดจำแบบรูป (pattern recognition) และการอุปมาความรู้ (knowledge deduction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาข่ายงานไฟฟ้าชีวภาพ (bioelectric network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท หรือ “นิวรอน” (neurons) และ จุดประสานประสาท (synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า “เดนไดรต์” (dendrite) ซึ่งเป็นข้อมูลเข้า (input) และปลายในการส่งกระแสประสาทเรียกว่า แอคซอน (axon) ซึ่งเป็นเหมือนผลลัพธ์ (output) ของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรต์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินใจว่าต้องกระตุ้นเซลล์อื่นๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่น ๆ ต่อไปผ่านทางแอกซอน ดังแสดงในภาพที่ 2.2



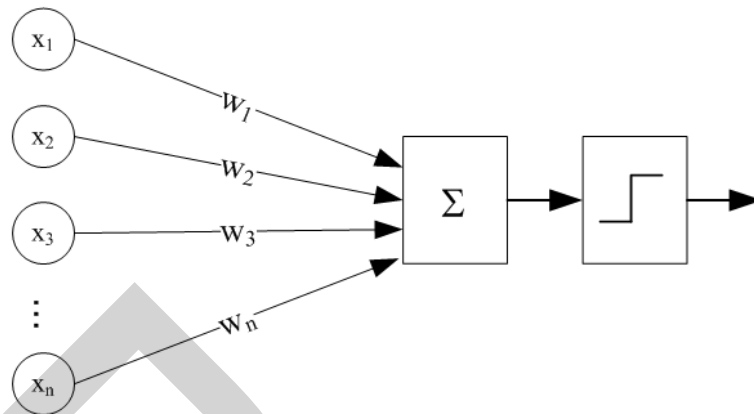
ภาพที่ 2.2 เซลล์ระบบประสาทของสิ่งมีชีวิตและโครงข่ายประสาทเทียม

#### ▪ หลักการทำงานของ neural networks

สำหรับในคอมพิวเตอร์ neurons [Chester M, 1993][ Smith M., 1993]ประกอบด้วย input และ output เหมือนกัน โดยจำลองให้ input แต่ละอันมีค่าน้ำหนัก (weight) เป็นตัวกำหนดน้ำหนักของ input โดย neuron แต่ละหน่วยจะมีค่า threshold เป็นตัวกำหนดว่าน้ำหนักรวมของ input ต้องมากขนาดไหนจึงจะสามารถส่ง output ไปยัง neurons ตัวอื่นได้ เมื่อนำ neuron แต่ละหน่วยมาต่อกันให้ทำงานร่วมกันการทำงานนี้ในทางตรรกแล้วก็จะเหมือนกับปฏิกิริยาเคมีที่เกิดในสมอง เพียงแต่ในคอมพิวเตอร์ทุกอย่างเป็นตัวเลข เพราะฉะนั้นเมื่อนำการทำงานเข้ามาใช้กับคอมพิวเตอร์สามารถเขียนได้ดังนี้

$$\text{if } (\text{sum}(\text{input} * \text{weight}) > \text{threshold}) \text{ then output}$$

หมายความว่า เมื่อมี input เข้ามายัง network ก็เอา input มาคูณกับ weight ของแต่ละขา ผลที่ได้จาก input ทุกขาของ neuron จะเอามารวมกันแล้วก็เอามาเทียบกับ threshold ที่กำหนดไว้ ถ้าผลรวมมีค่ามากกว่า threshold แล้ว neuron ก็จะส่ง output ออกไป output นี้ก็จะถูกส่งไปยัง input ของ neuron อื่น ๆ ที่เชื่อมกันใน network ถ้าค่าน้อยกว่า threshold ก็จะไม่เกิด output ดังแสดงในภาพที่ 2.3



ภาพที่ 2.3 แสดงรูปแบบของ Neuron ในคอมพิวเตอร์

สถาปัตยกรรมโครงข่ายประสาทเทียมที่พบทั่วไปจะมีลักษณะหลัก ๆ คือ มีการจัดเซลล์ประสาทเทียมเป็นชั้น ๆ (layer) ชั้นที่รับข้อมูลเข้าเรียกว่า ชั้นอินพุต (input layer) ชั้นที่ผลิตผลตอบของโครงข่ายเรียกว่า ชั้นเอาต์พุต (output layer) ส่วนชั้นอื่น ๆ ที่มีส่วนในการช่วยทำการประมวลผลอยู่ภายในเรียกว่า ชั้นซ่อน (hidden layer) ในโครงข่ายประสาทเทียมอาจมีชั้นซ่อนได้หลายชั้น โครงสร้างพื้นฐานจะมีลักษณะเป็นการประกอบกันของรูปแบบ ดังต่อไปนี้

### 1. แบบป้อนไปข้างหน้า

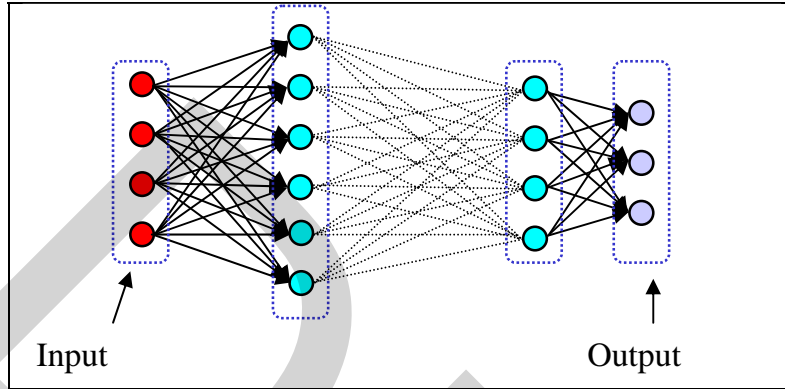
แบบป้อนไปข้างหน้า (feedforward network) ข้อมูลที่ประมวลผลในวงจรข่ายจะถูกส่งไปในทิศทางเดียวจาก input node ส่งต่อมาเรื่อย ๆ จนถึง output node โดยไม่มีการย้อนกลับของข้อมูล หรือแม้แต่ node ใน layer เดียวกันก็ไม่มี การเชื่อมต่อกัน สามารถจัดได้เป็น 2 แบบคือ แบบมีชั้นของเซลล์ประสาทชั้นเดียว และแบบมีชั้นของเซลล์ประสาทหลายชั้น โดยปกติแล้ว การเชื่อมโยงจะถูกกำหนดขึ้นระหว่างชั้นที่ติดกัน โดยจะมีการเชื่อมโยงระหว่างเซลล์ประสาทเทียมทุกตัว จากชั้นหนึ่ง ๆ ไปยังเซลล์ประสาทเทียมทุกตัวในชั้นต่อไป ในบางสถาปัตยกรรมอาจมีการเชื่อมโยงข้ามชั้นก็ได้ ดังแสดงในภาพที่ 2.4

### 2. แบบมีการป้อนไปเวียนกลับ

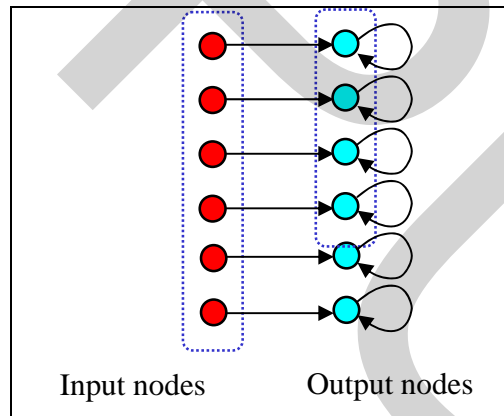
แบบมีการป้อนไปเวียนกลับ (feedback network) บางที่เรียกว่า recurrent network ข้อมูลที่ประมวลผลโครงข่ายประสาทเทียมจะมีการป้อนกลับเข้าไปยังวงจรข่ายหลาย ๆ ครั้ง จนกระทั่งได้



ค่าตอบออกมา จะมีการเชื่อมโยงที่ถูกกำหนดขึ้นระหว่างเซลล์ประสาทเทียมในชั้นหนึ่ง ๆ ย้อนกลับไปยังชั้นอื่น ๆ ก่อนหน้านั้น หรือแม้แต่ภายในชั้นเดียวกันเอง ดังแสดงในภาพที่ 2.5



ภาพที่ 2.4 แสดงสถาปัตยกรรมของ feedforward network



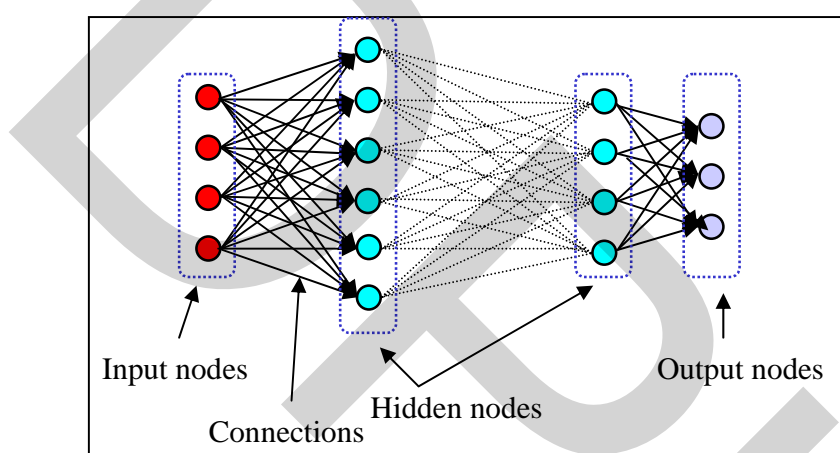
ภาพ 2.5 แสดงสถาปัตยกรรมของ feedback network

### 3. แบบป้อนไปข้างหน้าหลายชั้น

โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าหลายชั้น (multiple feedforward) นั้นใช้การเชื่อมโยงแบบป้อนไปข้างหน้าประกอบด้วย 3 layer ได้แก่ ชั้นของ input unit ที่ถูกเชื่อมต่อกับชั้นของ hidden unit ซึ่งเชื่อมต่อกับชั้นของ output unit คือเป็นการเพิ่มจำนวนชั้นใน hidden layer ดังแสดงในภาพ การทำงานของ input unit จะทำหน้าที่แทนส่วนของข้อมูลดิบ ที่จะถูกป้อนเข้าสู่

เครือข่าย และการทำงานของแต่ละ hidden unit จะถูกกำหนด โดยการทำงานของ input units และ ค่าน้ำหนักบนความสัมพันธ์ระหว่าง input unit และ hidden unit พฤติกรรมการทำงานของ output unit จะขึ้นอยู่กับการทำงานของ hidden unit และค่าน้ำหนักระหว่าง hidden unit และ output unit ดังแสดงในภาพที่ 2.6

ลักษณะการทำงานแบบนี้สามารถกำหนดการแทนค่าให้แก่ input unit ได้อย่างอิสระ ค่าน้ำหนักระหว่าง input unit และ hidden unit จะถูกกำหนดเมื่อ hidden unit กำลังทำงาน ฉะนั้น เวลาที่แก้ไขค่าน้ำหนัก hidden unit จะสามารถเลือกว่าอะไรคือค่าที่เราแทนเข้ามา



ภาพ 2.6 แสดงโครงสร้างแบบป้อนไปข้างหน้าหลายชั้น

ในโมเดลทางคณิตศาสตร์ของโครงข่ายประสาทเทียม แบบป้อนไปข้างหน้าหลายชั้น แสดงในภาพ มีเวกเตอร์อินพุต  $X = \{x_i\}$  กำหนดให้  $x_i \in \mathcal{R}^p$  และเวกเตอร์ผลลัพธ์  $Y = \{y_i\}$  กำหนดให้  $y_i \in \mathcal{R}^m$  ซึ่งประกอบด้วยสมการการคำนวณภายในเซลล์ประสาทแต่ละชั้นดังนี้

Hidden Layer

$$h_i = \sigma(u_j)$$

$$u_j = \sum_{i=1}^p w_{ji}^h x_i$$

กำหนดให้  $h_j$  คือผลลัพธ์ของเซลล์ใน Hidden layer ซึ่ง  $1 \leq j \leq L$

$\sigma(\cdot)$  คือ Activation function

$w_{ji}^h$  คือ ค่าของน้ำหนักบนเส้น  $ji$  เชื่อมเซลล์ในชั้นซ่อน  $j$  กับเซลล์ในชั้นอินพุต  $i$

$W^h$  คือ เวกเตอร์น้ำหนัก  $\{w_{ji}^h\}$

## Output Layer

$$y_k = \sigma(v_k),$$

$$y_k = \sum_{j=1}^m w_{kj}^y h_j$$

โดยที่  $y_k$  คือ ผลลัพธ์ของเซลล์ใน output layer ซึ่ง  $1 \leq k \leq m$

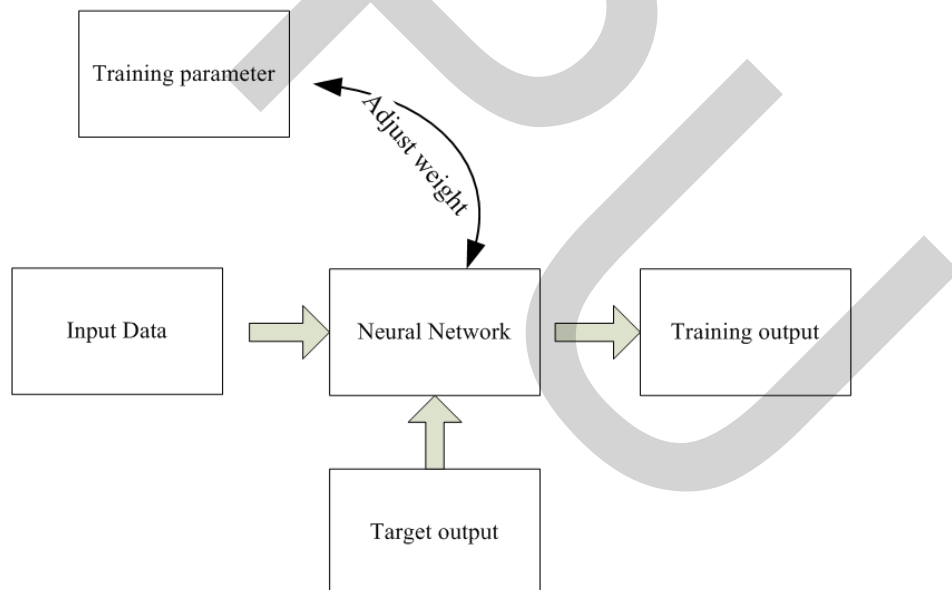
$\sigma(\cdot)$  คือ activation function

$w_{ki}^h$  คือ ค่าของน้ำหนักบนเส้น  $ki$  เชื่อมเซลล์ในชั้นซ่อน  $k$  กับเซลล์ในชั้นอินพุต  $j$

$W^y$  คือ เวกเตอร์น้ำหนัก  $\{w_{kj}^y\}$

- วิธีการเรียนรู้ของโครงข่ายประสาทเทียม

การเรียนรู้เป็นกระบวนการที่สามารถทำให้เปลี่ยนแปลงพฤติกรรม ความคิด สามารถเรียนรู้ได้จากการ ใฝ่เรียนการสัมผัส การอ่าน การใช้เทคโนโลยี สามารถแบ่งออกได้ 2 แบบ



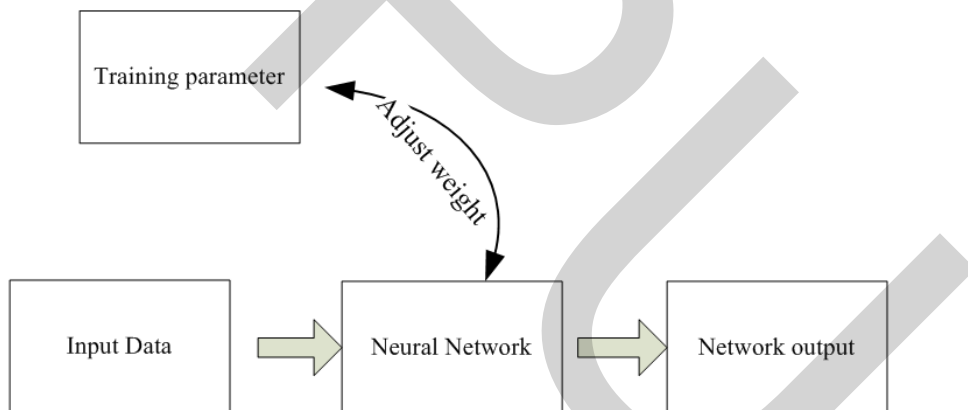
ภาพที่ 2.7 แสดงการเรียนรู้แบบมีการสอน

## 1. การเรียนรู้แบบมีการสอน

การเรียนรู้แบบมีการสอน (supervised learning) เป็นการเรียนรู้แบบที่มีการตรวจคำตอบ เพื่อให้วงจรรายปรับตัว ชุดข้อมูลที่ใช้สอนวงจรรายจะมีคำตอบไว้คอยตรวจดูว่าวงจรรายให้คำตอบที่ ถูกหรือไม่ ถ้าตอบไม่ถูก วงจรรายก็จะปรับตัวเองเพื่อให้ได้คำตอบที่ดีขึ้น ดังแสดงในภาพที่ 2.7

## 2. การเรียนรู้แบบไม่มีผู้สอน

การเรียนรู้แบบไม่มีการสอน (unsupervised learning) เป็นการเรียนแบบไม่มีผู้แนะนำ ไม่มีการตรวจคำตอบว่าถูกหรือผิด จะไม่มีค่าจริงในการฝึกสอน เครือข่ายจะจัดเรียงโครงสร้างของตัวเอง ตามลักษณะของข้อมูล ผลลัพธ์ที่ได้เครือข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ เช่นการแยก ลักษณะของคน พืช และสัตว์ เองได้โดยไม่มีผู้สอนจริงในการฝึกสอน เครือข่ายจะจัดเรียงโครงสร้าง ของตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้เครือข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ เช่นการ ที่เราสามารถแยกลักษณะของคน พืช และสัตว์ เองได้โดยไม่มีการสอนดังแสดงในภาพที่ 2.8



ภาพที่ 2.8 แสดงการเรียนรู้แบบไม่มีการสอน

### ▪ การเรียนรู้แบบแพร่ย้อนกลับ

การเรียนรู้แบบแพร่ย้อนกลับ (back-propagation learning) ถูกนำไปประยุกต์ใช้มากที่สุด ในการใช้งานในโครงข่ายประสาทเทียม โดยลักษณะของการเรียนรู้แล้ว การเรียนรู้แบบแพร่ย้อนกลับ ไม่จำเป็นต้องใช้เฉพาะโครงข่ายประสาทเทียมที่มีสถาปัตยกรรมแบบป้อนไปข้างหน้าหลายชั้นเพียง แบบเดียว แต่การจับคู่ดังกล่าวได้รับความนิยมนำไปใช้งานมากที่สุด การเรียนรู้แบบแพร่ย้อนกลับนี้

เป็นการเรียนรู้แบบมีการสอน (supervise learning) ขั้นตอนในการเรียนรู้แบบแพร่ย้อนกลับจะเริ่มด้วยการขนส่งข้อมูลที่จะใช้สอนเข้าไปในเครือข่ายทางชั้นอินพุต ชั้นอินพุตซึ่งโดยปกติจะไม่มีผลกระทบ จะทำหน้าที่ส่งกระจายข้อมูลนั้นไปยังเซลล์ประสาทเทียมต่างๆในชั้นซ่อน จากนั้นเซลล์ประสาทเทียมทุกตัวในชั้นซ่อน จะทำการประมวลผลตามวิธีการของแบบจำลองเซลล์ประสาทเทียม การประมวลผลเกิดจากชั้นซ่อนชั้นแรกผ่านไปตามชั้นซ่อนต่างๆจนสร้างผลตอบที่ชั้นเอาต์พุต

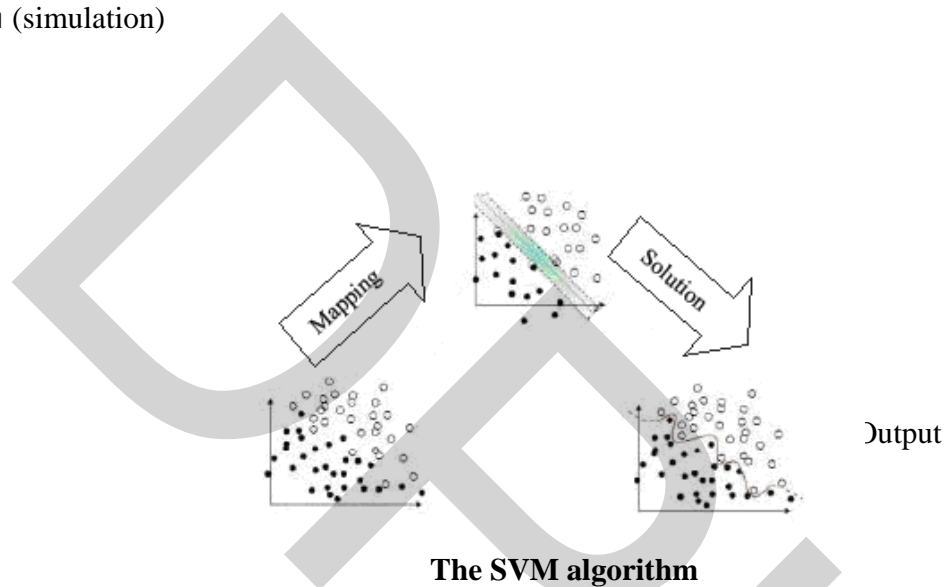
ผลที่ตอบออกมาจากชั้นเอาต์พุตจะถูกเปรียบเทียบกับค่าเป้าหมายที่กำหนด เพื่อคำนวณค่าความคลาดเคลื่อน (error) จากนั้นปริมาณการปรับแก้ค่าน้ำหนักของแต่ละข้อสัญญาณเข้าของแต่ละเซลล์ประสาทเทียมในชั้นเอาต์พุตก็就会被คำนวณขึ้น โดยดูจากความคลาดเคลื่อนและปริมาณการปรับแก้ค่าน้ำหนักที่ข้อสัญญาณของแต่ละเซลล์ประสาทเทียมในชั้นซ่อนต่างๆ แต่เนื่องจากไม่สามารถกำหนดค่าเป้าหมายของผลตอบจากเซลล์ประสาทเทียมในชั้นซ่อนจึงถูกมองว่า เป็นการไหลย้อนกลับของความคลาดเคลื่อนของเซลล์ประสาทเทียมต่างๆ ของชั้นเอาต์พุตหลังจากรวบรวมค่าความคลาดเคลื่อนที่ไหลย้อนกลับมาแล้ว หลักการเดียวกับชั้นเอาต์พุต คือ ดูจากผลรวมความคลาดเคลื่อนที่ไหลย้อนกลับมายังเซลล์ประสาทเทียมตัวนั้นๆ ประกอบกับค่าของสัญญาณเข้าที่วิ่งผ่านเข้ามาทางข้อสัญญาณเข้านั้นด้วย

เมื่อค่าน้ำหนักที่ข้อสัญญาณเข้าของเซลล์ประสาทเทียมทุกตัวถูกปรับแก้หมดแล้วโครงข่ายประสาทเทียมก็พร้อมจะเรียนรู้ข้อมูลตัวอย่างต่อไป ในทางปฏิบัติแล้ว จะต้องใช้ข้อมูลตัวอย่างจำนวนมาก และจะต้องสอนข้อมูลตัวอย่างเหล่านั้นซ้ำๆ กันหลายๆรอบ จึงจะสามารถสอนโครงข่ายประสาทเทียมชนิดนี้จนเรียนรู้ได้อย่างถูกต้อง

#### ▪ ข้อดีโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (Artificial Neural Networks: ANNs) เป็นแขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence: AI) ซึ่งเลียนแบบการทำงานและคุณสมบัติเซลล์สมองหรือระบบประสาทของมนุษย์ เมื่อโครงข่ายประสาทเทียมผนวกกับความสามารถของวิทยาการคอมพิวเตอร์ในปัจจุบัน เช่น หน่วยความจำ การประมวลผลที่รวดเร็ว แม่นยำ และค่าใช้จ่ายที่ไม่สูงนัก ทำให้ได้ระบบที่มีศักยภาพในการทำงาน มีคุณลักษณะและคุณสมบัติที่น่าสนใจ เช่น สามารถจำลองปัญหาได้โดยไม่จำเป็นต้องทราบรูปแบบการกระจายของข้อมูล (distribution free) มีข้อผิดพลาดได้บ้าง (fault tolerance) เรียนรู้ด้วยตนเองได้ (self-organization) ทำงานแบบขนาน (massively parallel process) รวดเร็ว (fast processing) ระบบทำงานโดยใช้เพียงฟังก์ชันทาง

คณิตศาสตร์อย่างง่ายแทนที่จะใช้กลไกทางชีวเคมี และมีได้ทำงานตามชุดคำสั่งแต่อย่างเดียวดังเช่นโปรแกรมคอมพิวเตอร์ทั่วไป คำตอบหรือผลลัพธ์น่าเชื่อถือ ด้วยเหตุผลดังกล่าวโครงข่ายประสาทเทียมจึงสามารถแก้ปัญหาใกล้เคียงกับเซลล์สมองหรือระบบประสาทของสิ่งมีชีวิตโดยเฉพาะมนุษย์ ระบบเรียนรู้หรือรู้จำจากตัวอย่างที่มีจำนวนและความหลากหลาย แหล่งที่มาของตัวอย่างอาจได้จากการสัมภาษณ์ตรง (direct interview) ข้อมูลในอดีต (historical record) หรือกระบวนการการจำลอง (simulation)



ภาพที่ 2.9 แสดงหลักการทำงานของซัพพอร์ตเวกเตอร์แมชชีน

#### 2.4.3 ซัพพอร์ตเวกเตอร์แมชชีน

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) [Richard, 2001] [Sonka, 2007][Svoboda, 2007] เป็นตัวแบบที่ใช้ในการแยกแยะข้อมูล โดย SVM จะทำการแบ่งชั้นของข้อมูลด้วยระนาบหลายมิติ จากข้อมูล 2 กลุ่มชุดข้อมูล โดยตัวแบบของ SVM มีเกี่ยวข้องกับโครงข่ายประสาทเทียม ตัวแบบของ SVM ใช้ sigmoid kernel function ซึ่งมีค่าเท่ากันทั้ง 2 ชั้น (layer) ตัวแบบของ SVM มีความคล้ายคลึงกับเพอร์เซพตรอน (perceptron) ซึ่งเป็นข่ายงานประสาทเทียมแบบง่ายมีหน่วยเดียวที่จำลองลักษณะของเซลล์ประสาทด้วยการใช้ kernel function ซึ่ง SVM จะใช้ข้อมูลของ คุณสมบัติและตัวแปรที่เปลี่ยนแปลงใช้ในการกำหนดระนาบหลายมิติ เรียกว่า โครงสร้าง ส่วนการเลือกที่มีความเหมาะสมที่สุดเรียกว่า โครงสร้างในการคัดเลือกฟีเจอร์ (feature selection) จำนวนเซตของโครงสร้างที่ใช้อธิบายในกรณีหนึ่ง (เช่น แถวของการค่า

คาตการณ์) เรียกว่า เวกเตอร์(vector) ดังนั้นจุดมุ่งหมายของตัวแบบ SVM คือการประโยชน์สูงสุดจากระนาบหลายมิติที่แบ่งแยกกลุ่มของเวกเตอร์ที่ถูกเรียกว่า การหาเส้นแบ่ง hyper-planes ซึ่งใช้แบ่งข้อมูลสองคลาสเพื่อให้ได้ผลลัพธ์ที่ดีโดยพิจารณาจากสมการเส้นตรง hyper planes และ SVMs จะทำการค้นหาเวกเตอร์ที่อยู่ใกล้เส้นแบ่ง hyper planes เรียกว่า ซัพพอร์ตเวกเตอร์ (support vectors) ดังแสดงในภาพที่ 2.9

▪ หลักการทำงานของ SVM

1. ข้อมูลที่อินพุตเข้าคำนวณหาค่า  $y$  ซึ่งค่าของ  $y \in \{-1,1\}$  ได้จากสมการ  $y = w^T \cdot x + b$  ถ้าค่าของ  $w^T \cdot x + b > 0$  จะกำหนดให้ค่า  $y=1$  ซึ่งจะจัดอยู่ใน class ที่ 1 ถ้าค่าของ  $w^T \cdot x + b < 0$  จะกำหนดให้ค่า  $y = -1$  ซึ่งจะจัดอยู่ใน class ที่ 2

2. คำนวณหาเส้นตรงที่แบ่งเอกสารซึ่งเรียกว่า เส้น optimal hyperplane จากสมการ

$$w^T \cdot x + b = 0$$

3. นำค่าที่ได้จากข้อที่ 1 และ 2 ไปเขียนบนเส้นตรงตามแนวแกนตั้งและแกนนอนจะได้ดังภาพที่ 9 โดยระยะทาง ( $d$ ) หรือ maximum margin จากเส้นขอบ ณ จุด  $x_i$  ไปยัง hyperplane สามารถแสดงได้ดังสมการ

$$d = \frac{|w^T \cdot x + b|}{\|w\|}$$

โดยกำหนดให้

$w$  คือ เวกเตอร์น้ำหนัก (weight vector)

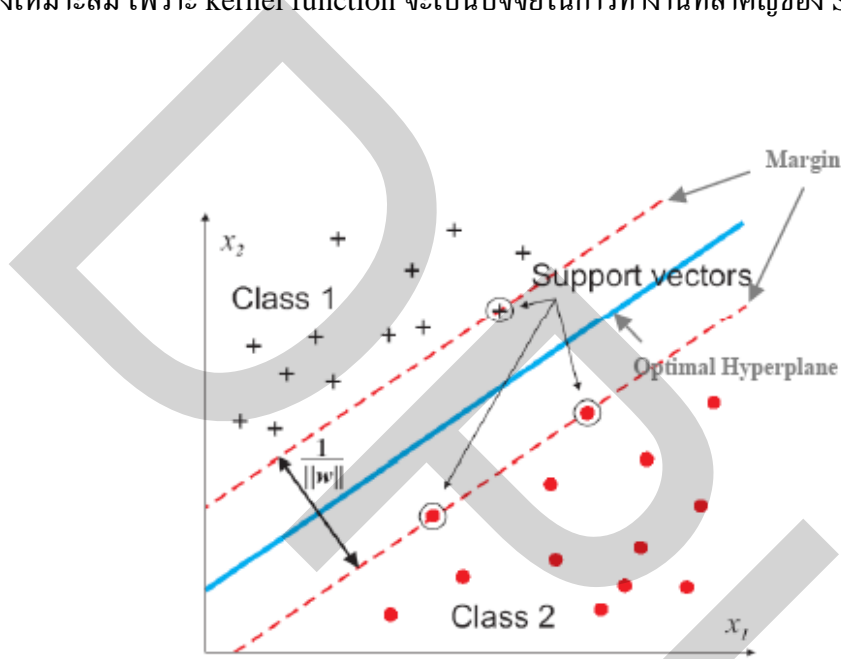
$x_i$  คือ input vector ของเอกสาร

$b$  คือ ค่าคงที่ที่กำหนดขึ้นเพื่อให้เหมาะสมกับการจัดกลุ่มเอกสาร

4. เลือกจุดที่อยู่ใกล้เส้นตรง optimal hyperplane ทั้งเหนือเส้นซึ่ง เรียกว่า “ขอบล่าง” ซึ่งเป็นขอบล่างสุดของ class เอกสารที่อยู่เหนือเส้นตรง optimal hyperplane และใต้เส้นเรียกว่า “ขอบบน” ซึ่งเป็นขอบบนสุดของ class เอกสารที่อยู่ใต้เส้นตรง optimal hyperplane เพื่อที่จะหาระยะทางระหว่างเส้นขอบทั้งสองโดยจะเลือกเอาค่าระยะทางที่ห่างจากเส้นตรง optimal hyperplane ที่น้อยที่สุดเป็นตัวเลือกในการจัดกลุ่มเอกสาร

อย่างไรก็ตามโดยพื้นฐานของ SVM นั้น จะสามารถแบ่งกลุ่มข้อมูลได้เพียง 2 กลุ่ม ดังนั้นการปรับเทคนิคของการเรียนรู้ด้วย SVM เพื่อให้ได้เป็นการจัดแบบหลายกลุ่มจึงเป็นสิ่งจำเป็น สำหรับ

งานวิจัยฉบับนี้จะปรับปรุงขั้นตอนการเรียนรู้ โดยเป็นการสร้างโมเดลการจัดกลุ่มด้วย One Class SVM นั่นคือ ให้แต่ละกลุ่มข้อมูลที่กำลังสนใจนั้นเป็น  $w^T \cdot x + b > 0$  โดยค่า  $y=1$  เมื่อผ่านข้อมูลแต่ละชุด (ที่ผ่านการจัดกลุ่มด้วยมือไว้ก่อนหน้า) เข้าสู่กระบวนการเรียนรู้ ก็จะสร้างโมเดลของการจัดกลุ่มเอกสารแต่ละกลุ่มด้วย SVM นอกจากนี้การจัดกลุ่มเอกสารด้วย SVM หากต้องการความถูกต้องอย่างมาก จะต้องคำนึงถึงลักษณะของข้อมูลที่มี เพื่อให้สามารถเลือก kernel function ของการทำงานได้อย่างเหมาะสม เพราะ kernel function จะเป็นปัจจัยในการทำงานที่สำคัญของ SVM



ภาพที่ 2.10 แสดงเส้นแบ่งข้อมูลด้วยซัพพอร์ตเวกเตอร์แมกซ์

ด้วยแนวคิดของการจัดกลุ่มด้วย SVM จะเป็นการสร้าง hyperplane เพื่อแยกกลุ่มเป็น 2 กลุ่ม แบบที่ต้องกำหนดค่า maximum margin ที่เหมาะสมที่สุดในการจัดกลุ่ม การใช้ maximum margin ตามทฤษฎีของ Vapnik Chervonenkis [Vapnik, 1989][ Vapnik, 2000] ด้วยการกำหนดค่าความผิดพลาดที่น้อยที่สุดเมื่อได้ค่า Margin ที่มากที่สุด แต่การทำลักษณะเช่นนี้แม้จะมีประโยชน์ แต่อาจจะก่อให้เกิดเส้นแบ่งเขตแดนที่ไม่เหมาะสมเพราะทำให้เกิดค่าความ ผิดพลาดสูง ดังนั้นจึงต้องกำหนดพีเจอร์ของค่า maximum margin hyperplane ดังนั้นจึงต้องมีการนำการทำงานอื่น ๆ เข้ามาช่วย ซึ่งก็คือการนำเอาหลักการเรื่อง kernel เข้ามาใช้เพื่อช่วยในการหาค่า maximum margin hyperplane



## 2.5 การใช้โมเดลในการสร้างการตัดสินใจ

การใช้โมเดลในการสร้างการตัดสินใจ เป็นวิธีการที่จะทำให้ความผิดพลาดเฉลี่ยในการจำแนกเกิดขึ้นน้อยที่สุด ขึ้นอยู่กับความสามารถในการแยกข้อมูลประเภทให้ถูกต้อง ซึ่งขึ้นอยู่กับความสามารถในการแยกความทับซ้อนกันของค่าการสะท้อนช่วงคลื่นระหว่างประเภท เพราะฉะนั้นการจำแนกจะรวมหลักการสำคัญหลายอย่างที่ต้องทำความเข้าใจในข้อกำหนด และปรัชญาของหลักการเหล่านั้น เพื่อให้การจำแนกได้ผลที่เชื่อถือได้และตรงกันทุกครั้ง (reliable and consistency)

การแบ่งข้อมูลให้เป็นแต่ละประเภท โดยทั่วไปจะใช้ขั้นตอนปฏิบัติ หรืออัลกอริทึมเพื่อทำการจำแนกข้อมูล เป็นเงื่อนไขที่แน่นอนที่กำหนดกระบวนการทางคอมพิวเตอร์ที่ใช้ข้อมูลเบื้องต้นที่หลากหลายไปสู่ผลลัพธ์ที่ต้องการ การใช้พีเอชหรือ พีเอชที่ได้มาจากการสกัดข้อมูลภาพเป็นอีกสิ่งหนึ่งที่ผลอย่างมากต่อผลที่นำมาใช้ในการตัดสินใจ ซึ่งในการทำให้ค่าที่ได้เปลี่ยนแปลงจะทำให้การตัดสินใจเปลี่ยนแปลงไปและจะมีผลต่อความแม่นยำในการจำแนกข้อมูล

ในการคาดคะเนค่าพารามิเตอร์โดยทั่วไปจะใช้ข้อมูลตัวอย่างที่เลือกสำหรับแต่ละวัตถุประสงค์ หรือเรียกว่า ข้อมูลฝึกหัด (training data) ข้อมูลตัวอย่างดังกล่าวจะต้องเป็นตัวแทนของประเภทหรือกลุ่มนั้น ๆ (class) ในขั้นแรกของการจำแนกข้อมูลใดๆ จะเป็นการฝึกให้โปรแกรมคอมพิวเตอร์จดจำลักษณะเฉพาะของประเภทที่สนใจ ในลักษณะนี้ของการจำแนกจะเป็นจุดวิกฤตของความสำเร็จในกระบวนการจำแนกโดยรวม และกินเวลาส่วนใหญ่ของผู้วิเคราะห์ วิธีการฝึกคอมพิวเตอร์ จำเป็นต้องให้กลุ่มตัวอย่างแก่เครื่องคอมพิวเตอร์ที่จะนำไปพัฒนา สัญลักษณ์ประเภท (class signature) สัญลักษณ์ประเภทนี้หมายถึง ลักษณะทางสถิติของกลุ่มข้อมูลตัวอย่าง โดยทั่วไปสามารถพัฒนากลุ่มสัญลักษณ์ของแต่ละประเภทได้ด้วยเทคนิคการจำแนก 2 แบบ

- เทคนิคการจำแนกแบบไม่มีการสอน

เทคนิคการจำแนกแบบไม่มีการสอน (unsupervised classification) การจำแนกแบบไม่มีการสอน เป็นทางหนึ่งในการสร้างกลุ่มที่มีความคล้ายคลึงกันที่อยู่ในพื้นที่เดียวกัน โดยผู้ใช้ไม่ต้องรู้ถึงรายการที่จะถูกจำแนกมาก่อน ค่าการสะท้อนช่วงคลื่นทั้งหมดของทุกจุดจะถูกตรวจสอบเพื่อหา

ความสัมพันธ์ทางสถิติ และรวมกลุ่มกันตามความคล้ายคลึงกันของกลุ่ม หลังจากนั้นจึงเป็นหน้าที่ของผู้แปล หรือวิเคราะห์ข้อมูลที่จะให้คำจำกัดความของแต่ละกลุ่มข้อมูลที่ถูกจำแนกไว้ ซึ่งการจำแนกแบบที่ไม่มีผู้สอนบางครั้งอาจจะถูกเรียกว่า การรวมกลุ่ม (clustering) ข้อเสียของการจำแนกข้อมูลในลักษณะนี้คือ ผลลัพธ์ที่ได้อาจจะไม่ได้ตรงกับที่ผู้ใช้ต้องการ หรือคาดหวังในหัวข้อที่กำหนดไว้ บางรายการที่ถูกจำแนกออกมานั้นอาจจะแปลผลได้ยาก เนื่องจากจำนวนกลุ่มที่ใช้ในการจำแนกเป็นสิ่งที่ไม่รู้มาก่อนล่วงหน้า

- เทคนิคการจำแนกแบบมีการสอน

เทคนิคการจำแนกแบบมีการสอน (supervised classification) ผู้วิเคราะห์จะต้องมีความรู้เกี่ยวกับข้อมูลที่จะทำการจำแนกอยู่ก่อนแล้ว และจะต้องมีการหาความสัมพันธ์ระหว่างกลุ่มที่จำแนกและพารามิเตอร์ไว้ล่วงหน้า และจะต้องมีการใช้ข้อมูลฝึกหัด (training data) เพื่อเป็นกลุ่มตัวอย่างที่จะมีการนำค่าที่ถูกคำนวณมาใช้ในการประเมินต่อไป เพราะฉะนั้นในการเลือกข้อมูลฝึกหัด อาจใช้วิธีการสุ่มแบบธรรมดา การแบ่งสุ่ม (stratified random) หรือ การสุ่มเป็นระบบ (systematic random) วัตถุประสงค์เพื่อให้ได้ตัวอย่างข้อมูลฝึกหัดกระจายทั่วทุกกรณี ในแต่ละประเภทควรจะใช้ตัวอย่างประมาณ 30-40 ตัวอย่าง เพื่อให้สามารถวิเคราะห์ข้อมูลทางสถิติได้

## 2.6 การวัดประสิทธิภาพ

การวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้าย เพื่อทำการตรวจสอบวิธีการที่ทำการทดลองมาข้างต้นว่ามีประสิทธิภาพมากหรือน้อยเพียงใดเมื่อนำมาใช้งานจริง จะเป็นขั้นตอนที่สำคัญ เพราะการนำวิธีการที่นำเสนอไปข้างต้นมาใช้งานได้นั้นจะต้องสอดคล้องกับความต้องการ จึงต้องมีการทดสอบศักยภาพการนำไปใช้ สถาปัตยกรรมที่ใช้ตัววัดความสำเร็จหลังการนำไปใช้หากนำไปใช้แล้วไม่ประสบผลสำเร็จต้องย้อนกลับไปเริ่มกระบวนการแรกใหม่ จึงต้องมีการประเมินผลก่อนการใช้งาน ในการประเมินนั้นกระทำได้โดยการวัดประสิทธิภาพของการจัดกลุ่มภาพมักจะถูกพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูลซึ่งจะประกอบด้วย การวัดค่าความแม่นยำ ค่าความระลึกลับ ค่าความถูกต้อง และ F-measure โดยยกตัวอย่างของค่าที่เกิดขึ้นจากตารางที่ 2.1

		ค่าทำนาย (predicted)	
		ปฏิเสธ (false/negative)	ยอมรับ (true/positive)
ค่าความจริง (actual)	ปฏิเสธ (negative)	a	b
	ยอมรับ (positive)	c	d
ค่าความถูกต้อง (Accuracy)		Acc	

ตารางที่ 2.1 ตารางแสดงการวัดประสิทธิภาพ

- ค่าความแม่นยำ (false positive rate / Precision: Pr) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพทั้งหมดที่ทำการค้นหาได้

$$Pr = \frac{a}{(a+b)}, a+b > 0$$

- ค่าความระลึก (true positive rate / Recall: Re) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพที่ถูกต้องทั้งหมด

$$Re = \frac{a}{(a+c)}, a+c > 0$$

- ค่าความถูกต้อง (Accuracy: Acc) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องทั้งหมดจากจำนวนภาพที่มีอยู่

$$Acc = \frac{(a+d)}{(a+b+c+d)}$$

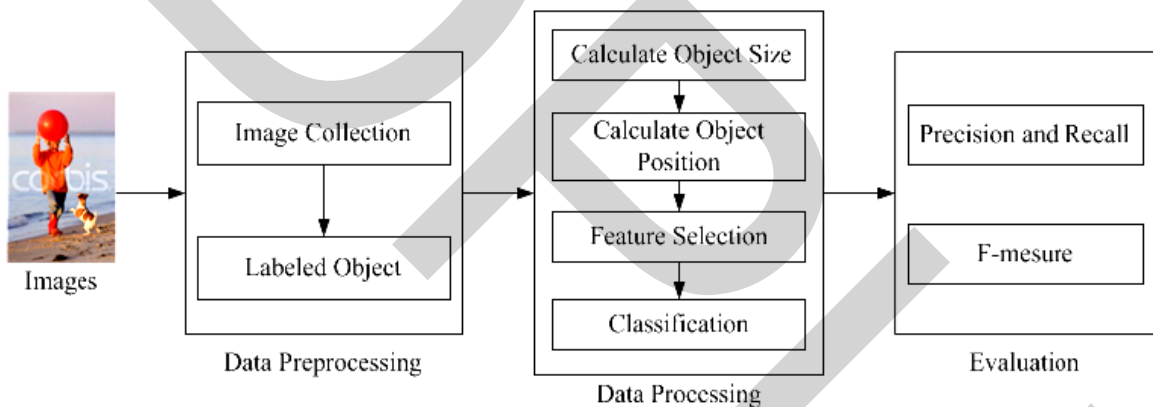
- ค่า F-measure เป็นการวัดค่าความสัมพันธ์ระหว่างค่าความระลึกและค่าความแม่นยำในเชิงฮาร์โมนิค (harmonic) เหมาะสำหรับฐานข้อมูลสารสนเทศที่มีขนาดใหญ่มาก และมักจะไม่สามารถหาข้อมูลภาพที่ถูกต้องทั้งหมดมีอยู่เท่าใด ทำให้ต้องทำการประมาณโดยใช้การสุ่มตัวอย่าง (sampling) ตามหลักทางสถิติหรือด้วยวิธีอื่นด้วย โดยทั่วไปจะเป็นการหาค่า F-measure ซึ่งแสดงสูตรได้ดังนี้

$$F = \frac{2(Pr \cdot Re)}{(Pr + Re)}$$

## บทที่ 3

### ขั้นตอนวิธีการดำเนินงานวิจัย

ในการศึกษางานวิจัยในครั้งนี้ได้มีการเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพที่เหมาะสมเพื่อเตรียมเป็นข้อมูลเบื้องต้น เป็นข้อมูลที่พร้อมจะทำการประมวลผลภาพ โดยในงานวิจัยจะมีการแบ่งขั้นตอนวิธีการดำเนินงานวิจัยออกเป็น 3 ส่วนหลักดังนี้ ขั้นตอนการเตรียมข้อมูล (data preprocessing) ขั้นตอนการประมวลผล (data processing) และ ขั้นตอนการวัดประสิทธิภาพ (evaluation) โดยขั้นตอนทั้งหมดถูกแสดงไว้ในภาพด้านล่าง



ภาพที่ 3.1 ขั้นตอนการจำแนกกลุ่มความหมายภาพ

### 3.1 ขั้นตอนการเตรียมข้อมูล

ขั้นตอนการเตรียมข้อมูล (data preprocessing) โดยการแยกข้อมูลซึ่งในงานวิจัยนี้ ข้อมูลที่กล่าวถึงเป็นข้อมูลภาพดิจิทัล ดังที่กล่าวมาข้างต้นแล้วนั้น สำหรับภาพบางภาพจะไม่นำเข้ามาทำการทดลองนั้นจะเป็นภาพที่ไม่สามารถนำมาประมวลผลได้ คือภาพที่มีความหมายกำกวม ภาพที่ไม่มี ความหมาย หรือภาพที่มีหลายความหมาย หรือภาพที่มีการโฟกัสระยะใกล้ ข้อมูลภาพที่มีความ ซ้ำซ้อน หรือไม่สอดคล้องกันจะถูกคัดเลือกภาพนั้นออกไป และทำการรวบรวมข้อมูลภาพที่ต้องการ

ที่มาจากหลายฐานข้อมูลจุดประสงค์ก็เพื่อให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นเหมาะสม ดังนั้นกระบวนการทั้งหมดนี้จะประกอบด้วย 2 กระบวนการดังนี้

### 3.1.1 การรวบรวมข้อมูลภาพ

การรวบรวมข้อมูลภาพ (image collection) จะมีการคัดเลือกภาพที่เหมาะสมจากแหล่งข้อมูลภาพ Fotosearch stock<sup>2</sup>, The Corbis Stock<sup>3</sup> และ The Corel Corporation<sup>4</sup> โดยภาพที่มีลักษณะผิดปกติ (outlier) หรือเป็นภาพที่มีลักษณะไม่สอดคล้องกับภาพส่วนใหญ่ ภาพเหล่านั้นจะถูกคัดเลือกออกไปจะไม่นำมาใช้ในการทดลอง เช่น ภาพถ่ายระยะใกล้ (close up) ภาพที่ไม่สามารถแปลความหมาย หรือ เป็นภาพที่ไม่สามารถวิเคราะห์ชื่อของวัตถุในภาพได้ หรือมีความหมายกำกวม เข้ามาพิจารณา ได้รวบรวมภาพทั้งหมด 900 ภาพ โดยแยกพิจารณาตามกลุ่ม กลุ่มละ 150 ภาพ ประกอบด้วยกลุ่มของ ภาพทะเล (beach), ภาพเมือง (city), ภาพโรงงาน (factory), ภาพทิวทัศน์ (landscape), ภาพภายใน (indoor)



( ก )



( ข )

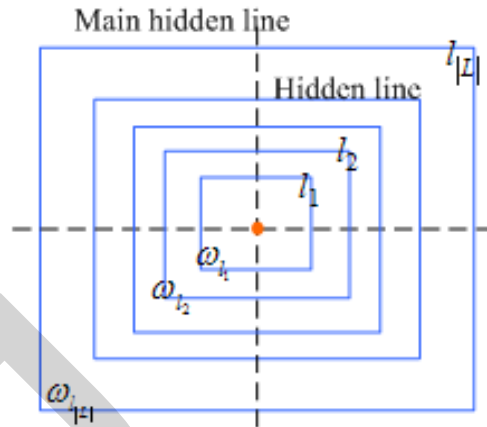
ภาพที่ 3.2 แสดงขั้นตอนการให้ความหมายกับภาพ

( ก ) ภาพที่ถูกคัดเลือกเข้าเป็นอินพุต ( ข ) ภาพถูกแท็กด้วยคำศัพท์<sup>2</sup>

<sup>2</sup> Fotosearch Stock: <http://www.fotosearch.com>

<sup>3</sup> The corbis stock: <http://pro.corbis.com>

<sup>4</sup> The Corel Corporation: <http://www.corel.com/>



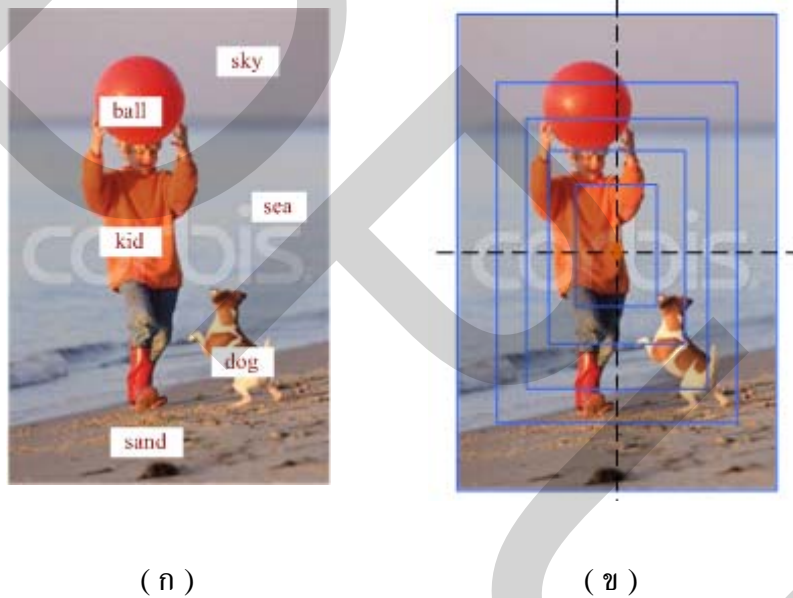
ภาพที่ 3.3 แสดงโครงสร้างสเกตตรีตรอน

### 3.1.2 การให้ความหมายกับวัตถุบนภาพ

ในกระบวนการทางการประมวลผลภาพ การแบ่งแยกวัตถุ หรือการทำ segmentation นั้นยังเป็นหัวข้อที่มีการวิจัยอย่างต่อเนื่อง จะทำการแบ่งแยกวัตถุที่มีรูปร่างลักษณะคล้ายคลึงกันหรือแตกต่างกันแต่มีความหมายเดียวกัน ซึ่งจะเป็นเรื่องที่ยากลำบาก รวมถึงกระบวนการที่การรู้จำรูปแบบวัตถุ เพื่อที่จะบอกความหมายของวัตถุหรือ ชื่อของวัตถุนั้น งานวิจัยบางกลุ่ม [Qian Huang, 1995][Vailaya, 2001] มักจะใช้การประมวลผลระดับต่ำ (low-level image processing) จะถูกเรียกว่า object recognition เป็นแขนงงานวิจัยที่มีนักวิจัยหลายกลุ่มให้ความสนใจและยังคงทำวิจัยอย่างต่อเนื่อง แต่อย่างไรก็ตามในการวิเคราะห์ความหมายของภาพในงานวิจัยนี้ได้เฉพาะเจาะจงในส่วนการประมวลผลระดับสูง (high-level image processing) [R. Zhao, 2002] จึงได้ข้ามในส่วนของการ object recognition ไปทำให้วัตถุนั้นถูกให้ความหมายด้วยวิธีการที่เรียกว่า การทำ labeled object หรือเรียกอีกอย่างว่าการ แท็ก (tag) [Ismail, 1998] [Tele, 2002] [Vasileios, 2003] [Zhao, 2001] ข้อมูลภาพให้อยู่ในรูปของชื่อวัตถุ หรือความหมายของวัตถุที่ปรากฏบนภาพ จากภาพที่ 3.2 (ก) เป็นภาพที่ถูกอินพุตเข้ามาในระบบและถูกให้ความหมายเป็นคำศัพท์ หรือแท็ก ข้อมูลวัตถุทั้งหมดและถูกจัดเก็บลงในรูปแบบของพีเจอร์เวกเตอร์ (feature vector) ดังแสดงในภาพที่ 3.2 (ข)

### 3.2 ขั้นตอนการประมวลผล

ขั้นตอนการประมวลผล (data processing) นำข้อมูลพีเจอร์เวกเตอร์ ทั้งหมดที่ได้ซึ่งประกอบด้วยข้อมูลวัตถุ ที่มีการแท็กมาเป็นพารามิเตอร์,  $O_i \in \{o_1, o_2, \dots, o_n\}$  และข้อมูลที่เป็น ขนาดของแต่ละวัตถุ (object size), ตำแหน่งของแต่ละวัตถุ (object position) เข้ามาประกอบในการพิจารณาร่วมกับโครงสร้างสเกตตรีตรอน เมื่อได้ข้อมูลครบถ้วนจะนำเข้าสู่กระบวนการคัดเลือกพีเจอร์ และการจำแนกข้อมูลภาพเป็นกระบวนการสุดท้าย วิธีการมีดังนี้



ภาพที่ 3.4 แสดงภาพที่ถูก Mapping ด้วยโครงสร้างสเกตตรีตรอน

( ก ) ภาพตัวอย่างที่ถูกแท็กด้วยคำศัพท์ ( ข ) ภาพที่ถูก Mapping ด้วยโครงสร้างสเกตตรีตรอน<sup>2</sup>

#### 3.2.1 การสกัดข้อมูลจากภาพ

การสกัดข้อมูลออกจากภาพ (feature extraction) จะเป็นวิธีที่ดึงข้อมูลที่ต้องการออกจากภาพซึ่งในขั้นตอนนี้จะเป็นการนำเอา โครงสร้างสเกตตรีตรอนเข้ามาประยุกต์เพื่อใช้ในการแปลความหมายภาพ ดังแสดงในภาพที่ 3.3 โดยพิจารณาจากข้อมูลวัตถุ  $O_i$  เริ่มต้น จากทฤษฎีของ Rudolph Arnheim [Rudolph, 1974] ที่กล่าวไว้ว่า การรับรู้รวมถึงการสนใจภาพครั้งแรกของมนุษย์นั้นจะรับรู้วัตถุที่เด่นก่อน โดยวัตถุที่เด่นนั้นจะเป็นวัตถุในแนวกึ่งกลางภาพ และจะต้องมีขนาดของ

วัตถุใหญ่เพียงพอ เพราะฉะนั้นวัตถุที่อยู่ด้านริมขอบภาพ หรือด้านข้างภาพจะรับรู้หรือสนใจเป็นส่วน ถัดไปโดยคิดเป็นสัดส่วนลดหลั่นกันตามขนาดและตำแหน่งบนภาพ ดังแสดงในภาพที่ 3.2 (ข) ประกอบด้วยวัตถุ *sea*, *sky*, *sand*, *kid*, *ball* และ *dog* โดยที่ขนาดของวัตถุที่เป็น *sea*, *sky* และ *sand* จะมีพื้นที่มากที่สุด ตามลำดับ แต่เมื่อพิจารณาโดยใช้หลักการของ Rudolph Arnheim ดังแสดงในภาพที่ 3.3 จะเห็นว่าเมื่อมนุษย์รับรู้ภาพครั้งแรก วัตถุที่มนุษย์สนใจ จะอยู่ตำแหน่งกึ่งกลางภาพจะมีขนาดที่เด่นพอ เพราะฉะนั้นสิ่งที่มนุษย์จะสนใจมากที่สุดคือ *kid* เป็นลำดับแรกมากกว่าที่จะสนใจ *sea* ทั้งที่มีขนาดใหญ่กว่ามาก เพราะฉะนั้น จากทฤษฎีข้างต้นจึงนำมาดัดแปลงเพื่อที่จะคิดคำนวณเป็นสมการเพื่อหาค่าของวัตถุที่ปรากฏบนภาพ และให้ค่าของวัตถุที่มีคุณลักษณะแตกต่างกัน มีค่าของวัตถุแตกต่างกันด้วย สามารถแสดงสมการของการประมวลผลดังนี้

- การหาขนาดของวัตถุ

การหาขนาดของวัตถุ (object size:  $s_i$ ) เป็นการนับจำนวนจุดพิกเซลทั้งหมดของแต่ละวัตถุที่ถูกแตกไว้แล้วโดยที่ขนาดของวัตถุใหญ่หรือพื้นที่มากที่สุดว่ามีจำนวนจุดพิกเซลมากกว่าวัตถุที่มีจำนวนพิกเซลน้อยหรือพื้นที่น้อย ยกตัวอย่างเช่น ภาพที่ 3.2 (ก) ขนาดของวัตถุ *sea* มีขนาดที่ใหญ่กว่า วัตถุ *kid* เป็นต้น

เมื่อกำหนดให้  $O_i \in \{o_1, o_2, \dots, o_n\}$  โดยที่แต่ละ  $o_i$  จะมีพื้นที่เป็น  $region(o_i) \in (x_i, y_i)$  สามารถเขียนสูตรนับจำนวนพิกเซลบนพื้นที่ของวัตถุทั้งหมดได้ดังนี้

$$s_i = \sum_{i=1}^{|o_n|} (pixel(x_i, y_i))$$

- การหาตำแหน่งวัตถุ

การหาตำแหน่งวัตถุ (object position:  $p_i$ ) จะใช้โครงสร้างสเกตตรีตรอนเพื่อทำการ mapping บนภาพเพื่อทำการคำนวณหาตำแหน่งของวัตถุและค่าน้ำหนักบนภาพ จากโครงสร้างที่กำหนดไว้ ดังแสดงในภาพที่ 3.2 (ข) เนื่องจากตำแหน่งของวัตถุที่อยู่จุดศูนย์กลางจะมีความสำคัญมากกว่า วัตถุที่อยู่ด้านริมหรือขอบภาพ ดังที่กล่าวมาแล้วข้างต้น

ดังนั้นจึงได้กำหนดค่าของน้ำหนักบนโครงสร้างสเกตตรีตรอนของแต่ละเส้นถูกกำหนดเป็นค่าของ  $\omega \in \{\omega_1, \omega_2, \dots, \omega_{|L|}\}$  ที่สัมพันธ์กับเส้นบนโครงสร้าง  $l \in \{l_1, l_2, \dots, l_{|L|}\}$  สามารถเขียนสูตรได้ดังนี้

$$p_i = \sum_{region(x_i, y_i) \in l_j, j=1}^{|o_n| |L|} \omega_{l_j}$$



### 3.2.2 การเลือกคุณลักษณะเฉพาะ

การเลือกคุณลักษณะเฉพาะ (feature selection) เป็นการคัดเลือกข้อมูลออกมาใช้งานเพียงบางส่วนซึ่งข้อมูลทั้งหมดที่ถูกจัดเก็บในรูปแบบของ ฟีเจอร์เวกเตอร์ นั้นมีจำนวนมากซึ่งคุณลักษณะบางตัวของ feature ไม่จำเป็นจึงจะต้องถูกคัดเลือกออกไป เพื่อให้ได้ผลลัพธ์โดยรวมแล้วมีค่าที่ดีที่สุด โดยได้ทำการเลือกอัลกอริทึม chi-square และ information gain ซึ่งเป็นอัลกอริทึมที่วัดความสามารถของ ฟีเจอร์ ทุกตัวเปรียบเทียบกัน สามารถอ่านเพิ่มเติมได้ในบทที่ 2

### 3.2.3 การจำแนกข้อมูลภาพ

การจำแนกข้อมูลภาพ (image classification) เป็นการนำข้อมูลทั้งหมดที่ได้จากการรวบรวมข้อมูลมาจัดให้เป็นหมวดหมู่ ในลักษณะต่างๆ ตามอัลกอริทึมจะทำการสร้างตัวแบบที่เกิดจากการเรียนรู้ตามคุณลักษณะที่เกิดขึ้นประจำจากกลุ่มภาพตัวอย่าง (training data set) ซึ่งในการทดลองนี้ ได้ใช้ การเรียนรู้ทั้งหมด 3 รูปแบบเพื่อนำมาทำการจัดกลุ่มภาพ ประกอบด้วย 3 วิธีการ naïve-Bayes, artificial neural network, supporting vector machine อ่านเพิ่มเติมได้ในบทที่ 2 แต่ ละวิธีการจะประมวลผลด้วยการแบ่งแบบ cross validation เป็น 10 กลุ่ม (fold) [Richard, 2001] โดยแยกพิจารณาตามกลุ่มความหมายภาพ กลุ่มละ 150 ภาพ ประกอบด้วยกลุ่มของ ภาพทะเล (beach), ภาพเมือง (city) ภาพโรงงาน (factory) ภาพทิวทัศน์ (landscape) และ ภาพภายใน (indoor)

## 3.3 ขั้นตอนการวัดประสิทธิภาพ

จากภาพที่ 3.1 ขั้นตอนการวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้ายของกระบวนการจำแนกความหมายภาพ จะนำเอาผลการจำแนกกลุ่มของภาพที่จัดได้มาทำการวัดประสิทธิภาพของสิ่งที่นำเสนอข้างต้น โดยจะตรวจสอบกลุ่มของภาพที่จัดได้ว่ามีค่าเป็นอย่างไร เมื่อเทียบกับกลุ่มของภาพที่ถูกต้อง ซึ่งต้องมีการวัดค่าความระลึก (recall) และค่าความแม่นยำ (precision) จะเป็นค่าที่แสดงว่า การค้นคืนข้อมูลได้ตรงกับความต้องการเพียงใด ส่วนค่าความระลึกจะเป็นค่าที่แสดงถึงความครอบคลุมในการจัดกลุ่มภาพ หลังจากนั้นจะนำค่ามาคำนวณในรูปแบบของค่าความถูกต้อง (accuracy) และ F-measure ต่อไปและนำค่าทั้งหมดมาแปลผลและประเมินผลลัพธ์ที่ได้ว่ามีความเหมาะสม หรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ในรูปแบบที่สามารถเข้าใจได้ง่าย เช่น

ตาราง หรือกราฟซึ่งในงานวิจัยนี้ได้แสดงในรูปของตารางเชื่อมโยงความสัมพันธ์ความสับสน (confusion matrix) สามารถอ่านเพิ่มเติมได้ในบทที่ 2



## บทที่ 4

### ผลการทดลอง

ในการท่วิจัยครั้งนี้ได้นำเสนอวิธีการจำแนกข้อมูลภาพ (semantic image classification) โดยใช้วิธีการรับรู้ภาพจากการมองของมนุษย์ด้วยโครงสร้างสเกตตรีตรอน ซึ่งมีการสกัดข้อมูลพีเจอร์จากกระบวนการที่กล่าวมาแล้วในบทที่ 2 และถูกจัดเก็บลงในรูปแบบของพีเจอร์เวกเตอร์ ซึ่งข้อมูลเบื้องต้นที่ได้มาจะมีจำนวนและขนาดของพีเจอร์ที่ใหญ่ จึงได้มีการคัดเลือกข้อมูลหรือพีเจอร์ (data selection) ที่สำคัญเพื่อนำมาใช้ในการจำแนกข้อมูลภาพ (image classification) ด้วยวิธีการ naïve-Bayes, multiple feedforward neural network (MFNN), supporting vector machine (SVM) ดังแสดงผลการทดลองดังนี้

#### 4.1 ผลการคัดเลือกข้อมูล

การคัดเลือกข้อมูล (data selection) เพื่อคัดเลือกความสามารถของพีเจอร์ที่มีความสำคัญที่สุดมาใช้ในการทดลองต่อไป การคัดเลือกข้อมูลเป็นสิ่งที่จำเป็นเมื่อข้อมูลที่ได้รับมานั้นมีจำนวนมาก เพราะฉะนั้นการคัดกรองข้อมูลที่ดีที่สุดเพื่อนำเข้ามาใช้ในกระบวนการเป็นสิ่งที่ทำให้ประสิทธิภาพในการทดลองดีขึ้น เพราะฉะนั้นจึงจำเป็นต้องมีการคัดเลือกเข้ามาเป็นพีเจอร์ ในการทดลองครั้งนี้ได้ทำการทดลองเพื่อเปรียบเทียบโดยใช้ 2 วิธีการ chi-square และ information gain เป็นอัลกอริทึมที่ใช้วัดความสามารถของพีเจอร์แต่ละตัว ดังแสดงในตารางที่ 4.1

จากตารางที่ 4.1 แสดงให้เห็นว่าข้อมูล แต่ละพีเจอร์ต่างมีประสิทธิภาพในการจำแนกประเภทข้อมูลภาพที่ต่างกัน เมื่อนำค่า chi-square และ information gain มาเรียงลำดับความสำคัญจากมากไปน้อยจะเห็นว่า ลำดับของการเรียงตัวของการใช้พีเจอร์ มีลำดับที่เหมือนกันดังแสดงด้านล่าง

Chi-squared  $F_{chi} = \{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}, f_1, f_9, f_3, f_{12}, f_6\}$  และ

Info gain  $F_{Info} = \{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}, f_1, f_9, f_3, f_{12}, f_6\}$

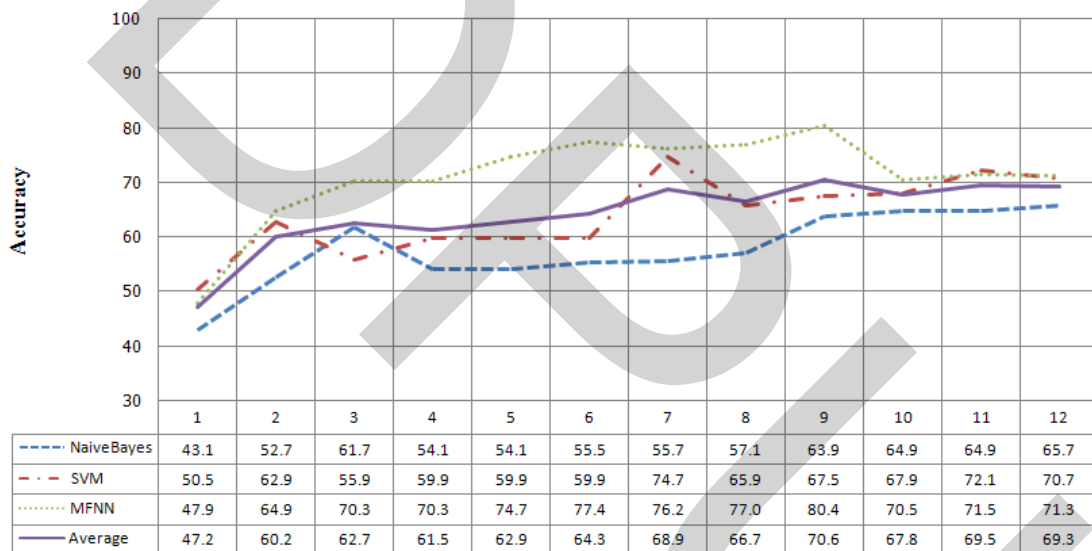
Features	ChiSquared	InfoGain ( $10^2$ )
1	69.38	9.67
2	119.82	14.92
3	39.30	5.29
4	156.31	23.03
5	162.19	23.54
6	13.37	1.97
7	455.54	65.7
8	213.99	28.01
9	59.56	8.95
10	91.51	12.6
11	90.36	11.99
12	32.25	4.29

ตารางที่ 4.1 แสดงผลการคัดเลือกข้อมูลวัตถุ ( $O_i$ )

จากผลการคัดเลือกข้อมูล ข้างต้นแสดงให้เห็นว่าข้อมูลในลำดับที่ 7 ( $f_7$ ) จะมีค่าของ chi-square และ information gain มีค่าสูงสุดเป็น 455.54 และ  $65.7 \times 10^2$  ตามลำดับ แสดงให้เห็นว่าข้อมูล  $f_7$  เป็นข้อมูลที่มีผลต่อการจำแนกข้อมูลมากที่สุดจากจำนวนฟีเจอร์ทั้งหมด 12 ตัว และในลำดับถัดมาข้อมูลลำดับที่ 8 ( $f_8$ ) จะมีค่าของ chi-square 213.99 และ information gain  $28.01 \times 10^2$  และถัดมาข้อมูลลำดับที่ 5 ( $f_5$ ) 162.19 และ  $23.54 \times 10^2$  เป็นค่า chi-square และ information gain ตามลำดับ แสดงให้เห็นว่าค่าของ  $f_8$  และ  $f_5$  จะมีความสำคัญต่อการจำแนกข้อมูลภาพถัดมาจาก  $f_7$  สำหรับข้อมูลลำดับที่ 6 ( $f_6$ ) ได้ค่าของ chi-square และ information gain น้อยที่สุดซึ่งมีค่าเป็น 13.37 และ  $1.97 \times 10^2$  ตามลำดับ แสดงให้เห็นว่าข้อมูลในลำดับนี้มีความสำคัญในการจำแนกข้อมูลที่น้อยที่สุด

เมื่อได้ลำดับความสามารถของข้อมูลทั้งหมดแล้ว นำฟีเจอร์ทั้งหมดมาทำการจำแนกข้อมูลภาพ ตามผลการทดลองข้างต้น โดยนำเอกลักษณ์ของการเรียงตัวของ การคัดเลือก มาทดลองเพื่อเปรียบเทียบการจำแนกโดยใช้การตัวจำแนก (classifier) ทั้งหมด 3 วิธี คือ naïve-Bayes, multiple feedforward neural network (MFNN), supporting vector machine (SVM) ได้ผล

การทดลองดังภาพที่ 4.1 เป็นการจำแนกข้อมูลภาพ ตามจำนวนของข้อมูล (number of features) โดยเริ่มต้นที่จำนวน 1 ฟีเจอร์ คือ ข้อมูลในลำดับที่ 7 ( $f_7$ ) ถูกจำแนกข้อมูลด้วยวิธี naïve-Bayes ได้ค่าความถูกต้อง (accuracy) เพียง 43% จำแนกข้อมูลด้วย SVM ได้ค่าความถูกต้อง 50% และจำแนกข้อมูลด้วย MFNN ได้ค่าความถูกต้อง 47.9% เมื่อพิจารณาค่าเฉลี่ยของความถูกต้องเป็น 47% แต่เมื่อมีการรวมกันของ ข้อมูลทั้งหมดจนกระทั่งครบทั้ง 12 ฟีเจอร์ ได้จำแนกข้อมูลด้วยวิธี naïve-Bayes ได้ค่าความถูกต้องเพียง 65.7% จำแนกข้อมูลด้วย SVM ได้ค่าความถูกต้อง 70.7% และจำแนกข้อมูลด้วย MFNN ได้ค่าความถูกต้อง 71.3% และค่าเฉลี่ยของความถูกต้องเพียงแค่ 69% ซึ่งนับว่ามีค่าที่น้อย เพิ่มขึ้นถึง 11 ฟีเจอร์ได้รับค่าความถูกต้องเพิ่มขึ้นเพียง 22% เท่านั้น



ภาพที่ 4.1 แสดงผลค่าความถูกต้องในการจำแนกข้อมูลภาพด้วยการเรียงลำดับฟีเจอร์ของวัตถุ ( $O_i$ )

ได้ทำการทดลองเพื่อพิจารณาค่าความถูกต้องโดยใช้การเรียงลำดับฟีเจอร์ของวัตถุ ( $O_i$ ) จากการทดลองคัดเลือกฟีเจอร์ ผลการทดลองที่ได้แสดงในภาพที่ 4.1 จะเห็นว่ากราฟ number of feature มีการเพิ่มขึ้นทีละหนึ่งฟีเจอร์ตามลำดับของการคัดเลือกข้อมูลข้างต้น โดยที่เริ่มต้นจากฟีเจอร์ตัวแรกที่ใช้ในการจำแนกคือ ฟีเจอร์ในลำดับที่ 7 ( $f_7$ ) ซึ่งเป็นฟีเจอร์ที่ถูกคัดเลือกมาให้เป็นฟีเจอร์ตัวแรกที่ใช้ในการทดลองเนื่องจากเป็นตัวที่สามารถแยกแยะกลุ่มของภาพได้มากที่สุด (จากการทดลองการคัดเลือกข้อมูลข้างต้น) โดยที่การที่ไม่มีการรวมกันของฟีเจอร์จะได้ค่าเฉลี่ยของความถูกต้องเป็น 47.2% ซึ่งเป็นเกือบครึ่งหนึ่ง ของการจำแนกข้อมูลทั้งหมด

จากภาพที่ 4.1 เมื่อพิจารณาค่าความถูกต้องที่ 7 9 และ 11 ฟีเจอร์มีค่าเฉลี่ยความถูกต้องเป็น 68.9% 70.6% และ 69.5% ตามลำดับ ในมีนี่จะมีการเปรียบเทียบฟีเจอร์ที่ 7 และ 9 เนื่องจากการใช้ข้อมูลถึง 11 ฟีเจอร์เป็นการใช้จำนวนของข้อมูลที่มากเกินไปจาก 12 ฟีเจอร์ และเมื่อเปรียบเทียบแล้วจะมีค่าความถูกต้องของการจำแนกด้วย SVM และ MFNN ที่น้อยกว่า การใช้ 7 ฟีเจอร์ เมื่อพิจารณาที่ 7 ฟีเจอร์ที่ประกอบด้วย  $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}\}$  เมื่อใช้เครื่องมือจำแนก naïve-Bayes จะได้ค่าความถูกต้องเพียง 55.7% และเมื่อมีการใช้เครื่องมือจำแนกข้อมูลด้วย SVM ได้ค่าความถูกต้อง 74.7% และ จำแนกข้อมูลด้วย MFNN ได้ค่าความถูกต้องเป็น 76.2% ส่วนค่าเฉลี่ยของความถูกต้องเป็น 68.9% และเมื่อทดลองแบบที่มีการรวมกันของข้อมูล ถึง 9 ฟีเจอร์ที่ประกอบด้วย  $\{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}, f_1, f_9\}$  จะได้ค่าความถูกต้องของ 63.9% 67.5% และ 80.4% โดยที่มีการใช้เครื่องมือจำแนกข้อมูลด้วย naïve-Bayes SVM และ MFNN ตามลำดับ ซึ่งมีการคิดเป็นค่าเฉลี่ยของความถูกต้องได้ 70.6% ซึ่งเป็นค่าที่มากกว่าการรวมกันโดยใช้ฟีเจอร์แบบ 7 ตัว

จากการทดลองทั้งสองส่วนข้างต้น เมื่อพิจารณาอย่างละเอียดการรวมกันของข้อมูล ทั้งหมด 7 และ 9 ฟีเจอร์ จะมีค่าเฉลี่ยของความถูกต้องอยู่ที่ 68.9% และ 70.6% ตามลำดับซึ่งมีค่าเฉลี่ยความถูกต้องห่างกันอยู่เพียง 1.8% แต่จะมีความต่างกันถึง 2 ฟีเจอร์ ได้แสดงการทดลองที่วิเคราะห์การจำแนกภาพ โดยใช้ข้อมูลแบบ 7 และ 9 ฟีเจอร์ ดังผลการทดลองต่อไปนี้

## 4.2 ผลการจำแนกความหมายของข้อมูลภาพ

ในขั้นตอนนี้จะทำการทดลองเพื่อจำแนกความหมายของข้อมูลภาพออกเป็นทั้งหมด 5 กลุ่ม ประกอบด้วยกลุ่มของ ภาพทะเล (beach), ภาพเมือง (city), ภาพโรงงาน (factory), ภาพทิวทัศน์ (landscape), ภาพภายใน (indoor) โดยใช้เครื่องมือการจำแนกทั้งหมด 3 วิธีด้วยกันคือ naïve-Bayes, MFNN, SVM และแสดงผลการทดลองในรูปแบบ ตารางเชื่อมโยงปฏิสัมพันธ์ความสับสน (confusion matrix) จุดประสงค์เพื่อทำการเปรียบเทียบการใช้ฟีเจอร์ในรูปแบบต่าง ๆ กัน ได้แสดงผลการทดลองดังนี้

#### 4.2.1 ผลการจำแนกข้อมูลด้วยวัตถุ

จากผลการทดลองข้างต้นได้ผลการคัดเลือกข้อมูลที่เป็นวัตถุ ( $O_i$ ) ข้างต้นจะสังเกตเห็นว่าลักษณะของรูปแบบชุดข้อมูลที่เป็นวัตถุบนภาพ จะมีชุดข้อมูลที่เรียงตัวกัน 2 แบบคือ การใช้วัตถุทั้งหมด 7 ฟีเจอร์ ( $O_i^7$ ) และ แบบใช้วัตถุทั้งหมด 9 ฟีเจอร์ ( $O_i^9$ ) ตามลำดับ ได้แสดงผลการทดลองเพื่อเปรียบเทียบจำแนกภาพด้วยตารางเชื่อมโยงปฏิสัมพันธ์ความสัมพันธ์ แสดงการทดลองผลดังนี้

- แบบใช้วัตถุทั้งหมด 7 ฟีเจอร์

จากผลการทดลองในตารางที่ 4.2 ถึง 4.4 จะแสดงถึงจำแนกข้อมูลในแต่ละกลุ่มและแต่ละวิธีการในการจำแนกภาพ จะเห็นว่าค่าเฉลี่ยของความถูกต้องได้เพียง 55.6% ด้วยการจำแนกแบบ naïve-Bayes ซึ่งจะมีค่าที่ค่อนข้างน้อยเมื่อเปรียบกับการจำแนกด้วยวิธี SVM และ MFNN จะได้มากถึง 74.6% และ 76% ซึ่งจะเห็นว่าเมื่อจำแนกด้วยวิธี MFNN จะได้ค่าความถูกต้องในกลุ่ม city ถึง 93% กลุ่ม factory 85% กลุ่ม indoor 75% ส่วนกลุ่มของ landscape ได้เพียง 53% ถ้าสังเกตในกลุ่มของ landscape ไม่ว่าจะมีการจำแนกด้วยวิธีใดจะได้รับค่าของ landscape ค่อนข้างจะต่ำกว่ากลุ่มอื่นๆ คือได้ค่าเพียง 39% 57% และ 53% ด้วยวิธีการจำแนก naïve-Bayes SVM และ MFNN ตามลำดับ

- แบบใช้วัตถุทั้งหมด 9 ฟีเจอร์

จากผลการทดลองในตารางที่ 4.5 ถึง 4.7 จะแสดงถึงจำแนกข้อมูลในแต่ละกลุ่มและแต่ละวิธีการในการจำแนกภาพ จะเห็นว่าค่าเฉลี่ยของความถูกต้องได้ 63.9% ด้วยการจำแนกแบบ naïve-Bayes แต่ เมื่อมีการจำแนกโดยใช้เครื่องมือแบบ SVM ได้ค่าความถูกต้องเป็น 67.5% และ MFNN ได้มากถึง 80.3% ซึ่งเมื่อมีจำแนกด้วยวิธี MFNN จะได้ค่าความถูกต้องในกลุ่ม city ถึง 95% กลุ่ม factory 89% กลุ่ม indoor 74% ส่วนกลุ่มของ landscape ได้เพียง 68% แต่เมื่อเปรียบเทียบกับวิธีการจำแนกด้วยวิธีจำแนก naïve-Bayes และ SVM ได้เพียง 36% และ 34% ตามลำดับ จะเห็นว่าในแต่ละกลุ่มมีค่าที่สูงกว่าแบบแรกเกือบทั้งหมด

แต่เมื่อวิเคราะห์อย่างละเอียดแล้วนั้น แบบใช้วัตถุทั้งหมด 7 ฟีเจอร์ กับแบบใช้วัตถุทั้งหมด 9 ฟีเจอร์ มีค่าความถูกต้องเฉลี่ยที่ห่างกันอยู่เพียง 1.8% และในค่าความถูกต้องของกลุ่ม landscape ในแบบใช้วัตถุทั้งหมด 9 ฟีเจอร์กลับมีค่าความถูกต้องที่ลดลงอย่างเห็นได้ชัดเจน ดังที่กล่าวมาแล้วข้างต้น เพราะฉะนั้นการวิเคราะห์ข้อมูลโดยใช้วัตถุเพียง 7 ฟีเจอร์ เท่านั้นก็มีความพอเพียงในการจำแนกความหมายภาพ

การจำแนกประกอบด้วย 7 ฟีเจอร์ ( $O_i^7$ )  $F = \{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}\}$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	53	28	0	0	19	54.1	53.0	53.5
City	15	60	8	1	16	40.8	60.0	48.6
Factory	4	17	58	18	3	71.6	58.0	64.1
Indoor	3	12	13	68	4	73.1	68.0	70.5
Landscape	23	30	2	6	39	48.1	39.0	43.1
Total accuracy rate						<b>55.60</b>		

ตารางที่ 4.2 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ในชุดข้อมูลแบบ 7 ฟีเจอร์

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	73	9	2	1	15	72.3	73.0	72.6
City	7	89	0	0	4	74.2	89.0	80.9
Factory	3	2	87	6	2	78.4	87.0	82.5
Indoor	3	6	17	67	7	80.7	67.0	73.2
Landscape	15	14	5	9	57	67.1	57.0	61.6
Total accuracy rate						<b>74.60</b>		

ตารางที่ 4.3 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ในชุดข้อมูลแบบ 7 ฟีเจอร์

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	74	13	2	0	11	71.2	74.0	72.5
City	5	93	0	0	2	75.6	93.0	83.4
Factory	3	0	85	11	1	81.7	85.0	83.3
Indoor	2	4	13	75	6	78.1	75.0	76.5
Landscape	20	13	4	10	53	72.6	53.0	61.3
Total accuracy rate						<b>76.00</b>		

ตารางที่ 4.4 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ในชุดข้อมูลแบบ 7 ฟีเจอร์



การจำแนกประกอบด้วย 9 feature ( $O_i^9$ )  $F = \{f_7, f_8, f_5, f_4, f_2, f_{10}, f_{11}, f_1, f_9\}$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	78	9	0	0	12	66.1	78.8	71.9
City	10	63	3	6	18	52.1	63.0	57.0
Factory	4	15	57	20	4	82.6	57.0	67.5
Indoor	4	8	7	75	6	71.4	75.0	73.2
Landscape	22	26	2	4	46	53.5	46.0	49.5
Total accuracy rate						<b>63.93</b>		

ตารางที่ 4.5 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ในชุดข้อมูลแบบ 9 ฟีเจอร์

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	82	9	0	0	8	52.9	82.8	64.6
City	28	71	0	0	1	65.1	71.0	67.9
Factory	4	2	79	13	2	80.6	79.0	79.8
Indoor	6	4	16	71	3	79.8	71.0	75.1
Landscape	35	23	3	5	34	70.8	34.0	45.9
Total accuracy rate						<b>67.54</b>		

ตารางที่ 4.6 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ในชุดข้อมูลแบบ 9 ฟีเจอร์

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	75	5	3	0	16	82.4	75.8	78.9
City	4	95	0	0	1	82.6	95.0	88.4
Factory	3	2	89	6	0	82.4	89.0	85.6
Indoor	3	3	14	74	6	78.7	74.0	76.3
Landscape	6	10	2	14	68	74.7	68.0	71.2
Total accuracy rate						<b>80.36</b>		

ตารางที่ 4.7 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ในชุดข้อมูลแบบ 9 ฟีเจอร์

feature	Accuracy rate (%)			
	Bayesian	SVM	NN	Average
$p_i$	37.68	32.87	43.49	38.01
$s_i$	35.67	38.08	48.90	40.88
$(p_i, s_i)$	47.09	47.70	65.53	53.44
$(O_i^7, p_i)$	64.33	73.55	76.75	71.54
$(O_i^7, s_i)$	83.37	87.78	89.78	86.97
$(O_i^7, p_i, s_i)$	93.80	94.80	85.20	91.27

ตารางที่ 4.8 ตารางแสดงผลค่าความถูกต้องของพีเจอร์ต่าง ๆ

#### 4.2.2 ผลการจำแนกข้อมูลภาพด้วยการรวมกันของพีเจอร์

จากการทดลองข้างต้นเป็นการหาความสมดุลของข้อมูลพีเจอร์วัตถุบนภาพ ซึ่งผลการทดลอง จะได้รับการจัดเรียงข้อมูลและขนาดของข้อมูลที่ได้ แต่อย่างไรก็ตามข้อมูลที่เกิดขึ้นยังมีอีกส่วนหนึ่งที่จะต้องนำมาพิจารณาด้วยคือ ข้อมูลที่ได้จะมีการสกัดออกมาจากภาพด้วยโครงสร้างสเกตตรีตรอน (ดังที่กล่าวในบทที่ 3) คือ ขนาดของวัตถุ (object size:  $s_i$ ) และตำแหน่งของวัตถุ (object position:  $p_i$ ) ที่สัมพันธ์สอดคล้องกันกับตัววัตถุบนภาพเข้ามา เพื่อนำมาช่วยให้การแยกแยะภาพได้ดีขึ้น ดังนั้นจึงได้ทำการทดลองคุณสมบัติของการจำแนกภาพตามแต่ละพีเจอร์ โดยใช้การผสมผสานพีเจอร์ในแบบที่ต่างกัน 6 วิธี โดยใช้วิธีการจำแนกแบบ naïve-Bayes SVM และ MFNN ดังแสดงในตารางที่ 4.8

จากตารางที่ 4.8 ค่าความถูกต้องในการจำแนกภาพที่มีการใช้เพียง ตำแหน่งของวัตถุ  $p_i$  และค่าของขนาดวัตถุ  $s_i$  เพียงอย่างเดียว จะมีค่าเฉลี่ยความถูกต้องเพียงแค่ 38% และ 40.8% ซึ่งจะมีค่าความถูกต้องที่น้อยกว่า การใช้พีเจอร์รวมกันของ  $p_i$  และ  $s_i$  แต่อย่างไรก็ตามเมื่อมีการรวมกันของ  $p_i$  และ  $s_i$  จะได้รับค่าความถูกต้องเฉลี่ยเพียงแค่ครึ่งหนึ่งเท่านั้น ยังคงเป็นค่าที่น้อยกว่าการใช้วัตถุ  $O_i^7$  เพียงอย่างเดียวในการจำแนกข้อมูล แต่เมื่อมีการนำวัตถุเข้ามาพิจารณาร่วมด้วยทำให้ค่าของความถูกต้องเพิ่มขึ้นอย่างเห็นได้ชัดทันที  $O_i^7$  และ  $p_i$  มีค่าความถูกต้องถึง 71.5% และ  $O_i^7$  และ  $s_i$  มีค่าความถูกต้องถึง 86.9% จะเห็นว่าเมื่อมีการรวมตัวกันของ  $O_i^7, p_i$  และ  $s_i$  จะทำให้ค่า

ความถูกต้องได้ถึง 91.2% ดังนั้นซึ่งจะแสดงการเปรียบเทียบการรวมกันของ feature และคุณสมบัติของ feature ที่นำมาใช้จำแนกดังนี้

- แบบใช้การรวมกันฟีเจอร์  $O_i^7$  และ  $p_i$

จากตารางที่ 4.9 ถึง 4.11 แสดงผลการทดลองในการจำแนกข้อมูลภาพ โดยใช้ feature  $O_i^7$  และ  $p_i$  จะสามารถจำแนกด้วยวิธี MFNN ได้ค่าความถูกต้องถึง 76.7% แต่อย่างไรก็ตามประสิทธิภาพของการจำแนกยังได้น้อยกว่า โดยใช้ข้อมูล  $O_i^9$  ซึ่งได้ค่าความถูกต้องถึง 80.3% แต่ในการจำแนกภาพด้วย MFNN ที่เป็นกลุ่มภาพของ beach จะได้ค่าความถูกต้องถึง 84% ในขณะที่ข้อมูล  $O_i^7$  ได้เพียง 75% แต่กลับจำแนกภาพในกลุ่มของ landscape ได้ค่าความถูกต้องเพียง 59% เท่านั้น แต่การจำแนกภาพในกลุ่มของ landscape โดยใช้ข้อมูล  $O_i^9$  เพียงอย่างเดียวมีค่าความถูกต้องได้ถึง 68% ข้อมูลจากตารางที่ 4.7

- แบบใช้การรวมกันฟีเจอร์  $O_i^7$  และ  $s_i$

จากตารางที่ 4.12 ถึง 4.14 แสดงผลการทดลองในการจำแนกข้อมูลภาพ โดยใช้ข้อมูล  $O_i^7$  และ  $s_i$  จะสามารถจำแนกด้วยวิธี MFNN ได้ค่าความถูกต้องถึง 89.6% ซึ่งเมื่อเปรียบกับแบบแรกที่มีการผสมระหว่างข้อมูล  $O_i^7$  และ  $p_i$  แล้วจะให้ค่าความถูกต้องที่สูงกว่าถึง 12.8% และสามารถจำแนกภาพในประเภทของ city ได้ค่าความถูกต้องถึง 99% และยังสามารถจำแนกกลุ่มข้อมูล landscape ได้ถึง 88% ในขณะที่ฟีเจอร์  $O_i^7$  และ  $p_i$  จำแนกด้วยวิธี MFNN ได้ค่าความถูกต้องเพียง 59% แสดงให้เห็นว่า  $s_i$  มีผลต่อการจำแนกประเภทข้อมูลมากกว่าการใช้  $p_i$

การจำแนกประกอบด้วย 2 ฟีเจอร์  $O_i^7$  และ  $p_i$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	79	13	0	0	7	67.5	79.8	73.1
City	5	72	1	8	14	58.5	72.0	64.6
Factory	8	9	60	17	6	78.9	60.0	68.2
Indoor	3	6	13	72	6	71.3	72.0	71.6
Landscape	22	23	2	4	49	59.8	49.0	53.8
Total accuracy rate						<b>66.53</b>		

ตารางที่ 4.9 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ด้วยชุดข้อมูล  $O_i^7$  และ  $p_i$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	89	10	0	0	0	71.8	89.9	79.8
City	8	88	0	0	4	62.4	88.0	73.0
Factory	4	5	79	10	2	78.2	79.0	78.6
Indoor	4	10	19	65	2	82.3	65.0	72.6
Landscape	19	28	3	4	46	85.2	46.0	59.7
Total accuracy rate						<b>73.55</b>		

ตารางที่ 4.10 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ด้วยชุดข้อมูล  $O_i^7$  และ  $p_i$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	84	7	0	4	4	80.8	84.8	82.8
City	5	93	0	0	2	72.7	93.0	81.6
Factory	1	4	78	12	5	80.4	78.0	79.2
Indoor	1	8	16	69	6	73.4	69.0	71.1
Landscape	13	16	3	9	59	77.6	59.0	67.0
Total accuracy rate						<b>76.75</b>		

ตารางที่ 4.11 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ด้วยชุดข้อมูล  $O_i^7$  และ  $p_i$

การจำแนกประกอบด้วย 2 ฟีเจอร์  $O_i^7$  และ  $s_i$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	96	4	0	0	0	94.1	96.0	95.0
City	0	83	13	4	0	82.2	83.0	82.6
Factory	0	2	78	20	0	74.3	78.0	76.1
Indoor	0	0	12	85	3	73.9	85.0	79.1
Landscape	6	12	2	6	74	96.1	74.0	83.6
Total accuracy rate						<b>83.20</b>		

ตารางที่ 4.12 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ด้วยชุด

ข้อมูล  $O_i^7$  และ  $s_i$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	96	1	0	0	3	96.0	96.0	96.0
City	0	95	5	0	0	86.4	95.0	90.5
Factory	0	8	77	12	1	81.1	78.6	79.8
Indoor	1	1	12	81	5	85.3	81.0	83.1
Landscape	3	5	1	2	89	90.8	89.0	89.9
Total accuracy rate						<b>87.95</b>		

ตารางที่ 4.13 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ด้วยชุดข้อมูล  $O_i^7$

และ  $s_i$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	96	1	0	0	3	96.0	96.0	96.0
City	0	99	0	1	0	86.8	99.0	92.5
Factory	0	10	81	7	2	85.3	81.0	83.1
Indoor	0	0	11	84	5	90.3	84.0	87.0
Landscape	4	4	3	1	88	89.8	88.0	88.9
Total accuracy rate						<b>89.60</b>		

ตารางที่ 4.14 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ด้วยชุดข้อมูล

$O_i^7$  และ  $s_i$

การจำแนกประกอบด้วย 3 ฟีเจอร์  $O_i^7$ ,  $p_i$  และ  $s_i$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	96	4	0	0	0	95.0	96.0	95.5
City	0	82	11	4	3	81.2	82.0	81.6
Factory	1	3	83	12	1	79.0	83.0	81.0
Indoor	0	2	9	86	3	80.4	86.0	83.1
Landscape	4	10	2	5	79	91.9	79.0	84.9
Total accuracy rate						<b>85.20</b>		

ตารางที่ 4.15 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ naïve-Bayes ด้วยชุด

ข้อมูล  $O_i^7$ ,  $p_i$  และ  $s_i$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	99	1	0	0	0	97.1	99.0	98.0
City	0	99	1	0	0	92.5	99.0	95.7
Factory	0	0	94	6	0	92.2	94.0	93.1
Indoor	0	0	4	90	6	93.8	90.0	91.8
Landscape	3	7	3	0	87	93.5	87.0	90.2
Total accuracy rate						<b>93.80</b>		

ตารางที่ 4.16 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ SVM ด้วยชุดข้อมูล

$O_i^7$ ,  $p_i$  และ  $s_i$

Class	Confusion Matrix (%)					Performance (%)		
	Beach	City	Factory	Indoor	Landscape	Prec	Rec	F1
Beach	96	1	0	0	3	99.0	96.0	97.5
City	0	99	1	0	0	93.4	99.0	96.1
Factory	0	0	97	2	1	91.5	97.0	94.2
Indoor	0	2	4	91	3	97.8	91.0	94.3
Landscape	1	4	4	0	91	92.9	91.0	91.9
Total accuracy rate						<b>94.80</b>		

ตารางที่ 4.17 ตารางแสดงผลลัพธ์การจำแนกข้อมูล ด้วยวิธีการจำแนกแบบ MFNN ด้วยชุดข้อมูล

$O_i^7$ ,  $p_i$  และ  $s_i$

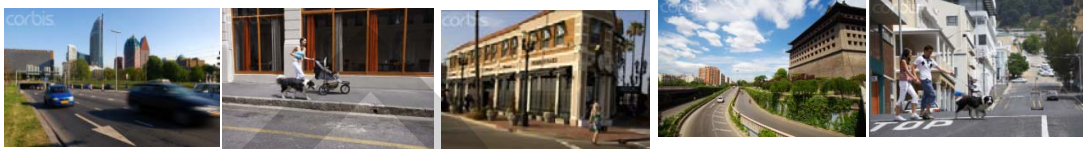
- แบบใช้การรวมกันพีเจอร์  $O_i^7$ ,  $p_i$  และ  $s_i$

จากตารางที่ 4.15 ถึง 4.17 แสดงผลการทดลองในการจำแนกข้อมูลภาพ โดยใช้ข้อมูล  $O_i^7$ ,  $p_i$  และ  $s_i$  ในการทดลองนี้เป็นการรวมพีเจอร์ทั้งหมดเพื่อใช้ในการจำแนกข้อมูลและสามารถจำแนกภาพด้วยวิธี MFNN ได้ค่าความถูกต้องถึง 94.8% และในการจำแนกกลุ่มของ landscape สามารถจำแนกด้วยวิธี naïve-Bayes ได้ค่าความถูกต้องถึง 79% ด้วยวิธี SVM ได้ค่าความถูกต้องถึง 87% และวิธี MFNN ได้ค่าความถูกต้องถึง 91% ซึ่งเมื่อเปรียบเทียบกับการจำแนกด้วยชุดข้อมูลแบบ  $O_i^7$  และ  $s_i$  แล้วนั้นแบบของการผสมกันทุกตัวข้อมูล สามารถที่จะจำแนกด้วยวิธี MFNN ในกลุ่มของ factory, indoor และ landscape ได้ดีขึ้นกว่าเดิมถึง 16% 7% และ 3% ตามลำดับ ได้แสดงผลลัพธ์ของการจำแนกความหมายภาพ ในกลุ่มต่างๆ ดังภาพที่ 4.2

จากการทดลองทั้งหมดเมื่อพิจารณาโดยรวมแล้วนั้น การใช้ชุดของข้อมูลที่มีการรวมกันของพีเจอร์ทั้งสามชุดข้อมูล ที่ประกอบด้วย  $O_i^7$ ,  $p_i$  และ  $s_i$  สามารถจะช่วยในการจำแนก กลุ่มข้อมูลภาพ ให้มีความหมายได้ดีขึ้น ด้วยวิธีการ MFNN (multiple feedforward neural network) จะได้ประสิทธิภาพถึง 94.8% และเมื่อพิจารณาค่าเฉลี่ยของความถูกต้องโดยรวมจะเป็น 91.3% ซึ่งเมื่อเปรียบเทียบกับวิธีการอื่นแล้วนั้นจะได้รับค่าความถูกต้องโดยรวมไม่ถึง 87% เพราะฉะนั้นการใช้วิธีการจำแนกแบบ MFNN จะสามารถจำแนกข้อมูลความหมายภาพได้ดีที่สุด ดังแสดงภาพผลลัพธ์ได้ในภาพที่ 4.2



(ก) ภาพตัวอย่างในกลุ่ม beach



(ข) ภาพตัวอย่างในกลุ่ม city



(ค) ภาพตัวอย่างในกลุ่ม factory



(ง) ภาพตัวอย่างในกลุ่ม landscape



(จ) ภาพตัวอย่างในกลุ่ม indoor

ภาพที่ 4.2 ผลลัพธ์ของการจำแนกความหมายภาพในแต่ละกลุ่ม



## บทที่ 5

### สรุปผลการทดลอง

ในบทนี้ได้ทำการสรุปผลการทดลอง และข้อเสนอแนะ รวมทั้งข้อผิดพลาดที่เกิดขึ้นจากการทดลอง รวมไปถึงสิ่งที่ควรปรับปรุงเพิ่มเติม เพื่อแก้ไขข้อผิดพลาดและแนวทางการทำวิจัยต่อในเรื่องของ semantic image ซึ่งเป็นงานวิจัยที่ปัจจุบันได้มีนักวิจัยให้ความสนใจอย่างแพร่หลาย เพราะสามารถนำไปประยุกต์ใช้ได้หลายด้านทั้งทางด้านการแพทย์ และการในอุตสาหกรรม เป็นต้น

#### 5.1 สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้เสนอหัวข้อวิจัยทางการประมวลผลภาพ ในส่วนของการแปลความหมายภาพ หรือเรียกอีกอย่างว่า semantic image โดยใช้วิธีการทฤษฎีการรับรู้จากการมองภาพของมนุษย์ด้วยโครงสร้างสเกตตรีตรอน ถูกนำมาดัดแปลงใช้ในการสกัดข้อมูลภาพ เพื่อนำมาใช้เป็นพารามิเตอร์ในการจำแนกข้อมูล (classification) โดยปรกติทั่วไปนั้นการใช้อัลกอริทึมที่มาสกัดข้อมูลภาพนั้นมักจะใช้สกัดเพียงข้อมูลที่เกิดขึ้นภายในภาพ แล้วนำมาประมวลผลเท่านั้น ไม่ได้มีการใช้ทฤษฎีของการรับรู้ภาพเข้ามาช่วยในการวิเคราะห์ จากการทดลองจะสามารถพิสูจน์ได้ว่าทฤษฎีนี้เหมาะสมที่จะนำมาใช้ร่วมกันเพื่อแปลความหมายของภาพ เพราะฉะนั้นในงานวิจัยนี้ยังคงเป็นอีกแนวทางหนึ่งที่สามารถนำวิธีการที่นำเสนอเข้ามาประยุกต์เพื่อให้สามารถนำวิธีการมาช่วยในแปลความหมายของภาพได้ ดังผลสรุปจากการทดลองดังนี้

จากขั้นตอนการทดลองเบื้องต้นในบทที่ 4 สามารถสรุปได้ว่า พารามิเตอร์ชื่อวัตถุบนภาพเป็นฟีเจอร์หลักที่นำเข้ามาพิจารณาเป็นลำดับแรก และเนื่องจากมีจำนวนชุดข้อมูลต่อภาพจำนวนหลายจำนวน เพราะฉะนั้นจึงได้นำชุดข้อมูลผ่านกระบวนการคัดเลือกฟีเจอร์ ซึ่งทำให้มั่นใจได้ว่าข้อมูลทุกตัวที่เข้ามานั้นเป็นข้อมูลที่มีคุณภาพ และข้อมูลส่วนที่เพิ่มเติมคือ ขนาดและตำแหน่งของวัตถุบนภาพ ซึ่งได้จากกระบวนการประมวลผลระดับต่ำ ที่เรียกว่า low level feature extraction และทำการแม็พวัตถุบนภาพในรูปแบบของโครงสร้างสเกตตรีตรอน เพื่อสกัดข้อมูลต่างๆออกมาเป็นในรูปแบบของฟีเจอร์เวกเตอร์ (feature vector) ดังที่กล่าวมาแล้วในบทที่ 2 และ 3

จากการทดลองได้ทำการแบ่งกลุ่มของภาพไว้ทั้งหมด 5 กลุ่ม ภาพทะเล (beach) ภาพเมือง (city) ภาพโรงงาน (factory) ภาพทิวทัศน์ (landscape) และ ภาพภายใน (indoor) จะเห็นว่าการจำแนกภาพ โดยใช้เพียงบางพีเจอร์ไม่สามารถที่จะจำแนกภาพให้มีความหมายตรงตามกลุ่มได้อย่างชัดเจน แต่อย่างไรก็ตามเมื่อมีการนำเอาพีเจอร์เข้ามาารวมกัน ทำให้เกิดการผสมผสานของพีเจอร์ โดยเฉพาะการรวมกันของพีเจอร์ทุกตัวสามารถที่จะจำแนกภาพได้ดีโดยเฉพาะวิธีการ MFNN สามารถจำแนกภาพได้มีประสิทธิภาพมากกว่าวิธีการอื่น และได้มากถึง 94.8% ในขณะที่ naïve-Bayes และ SVM จะได้เพียง 85.2% และ 93.8% ตามลำดับ มีค่าความถูกต้องที่เพิ่มขึ้นจากการใช้พีเจอร์ที่เป็นวัตถุอย่างเดียวยังถึง 22.7% ดังแสดงภาพตัวอย่างในบทที่ 4

เพราะฉะนั้นจากการทดลองสามารถสรุปได้ว่า การที่ใช้นาถฤษฎีของการรับรู้จากการมองของมนุษย์เข้ามาใช้เพื่อสกัดข้อมูลภาพ เช่น ตำแหน่งของวัตถุบนโครงสร้างสเกตริตรอน รวมทั้งค่าของขนาดวัตถุที่อยู่บนโครงสร้าง และใช้ผสมผสานกับข้อมูลทั้งหมดบนภาพ ทำให้สามารถใช้ในการจำแนกภาพในรูปแบบ ของความหมายของภาพโดยรวมได้ (semantic image classification) เป็นอย่างดี

## 5.2 ข้อเสนอแนะ

ในกระบวนการจำแนกข้อมูลภาพออกเป็นกลุ่มย่อยที่มีความหมายในรูปแบบของความหมายภาพโดยรวมนั้นจะเป็นการพิจารณาภาพที่ซับซ้อนมากขึ้น นอกจากการค้นคืนข้อมูลภาพหรือการจำแนกภาพโดยทั่วไปที่ต้องการค้นคืนเพียง ความเหมือนกันของวัตถุที่เกิดขึ้นบนภาพเท่านั้น ในงานวิจัยส่วนใหญ่จึงเน้นไปที่การสกัดข้อมูลภาพในรูปแบบของการสกัดด้วยอัลกอริทึมที่แตกต่างกัน และค้นหาภาพเพียงความเหมือนกันของรูปทรงหรือลักษณะเฉพาะ หรือเพียงแต่วัตถุนภาพ เท่านั้น ทั้งที่มีความหมายและไม่มี ความหมาย แต่อย่างไรก็ตาม เบื้องต้นของผลลัพธ์ที่ได้จากการค้นคืนหรือจำแนกภาพ คือความเหมือนกันทางกายภาพ เช่น รูปทรง สี หรือชนิดของวัตถุ แต่ลักษณะการวิเคราะห์และพิจารณาของการเหมือนกันทางความหมายภาพ หรือ semantic นั้นจะมีลักษณะการวิเคราะห์ที่แตกต่างกันออกไป ในอีกรูปแบบหนึ่งซึ่งเป็นรูปแบบที่เกิดจากความคิดของมนุษย์ที่มีการแปลงความจากภาพ แต่อย่างไรก็ตามสิ่งที่ได้จากการทดลองนั้นยังคงมีข้อที่ต้องปรับปรุงเพิ่มอยู่อีกหลายส่วนด้วยกัน

สิ่งควรจะมีการปรับปรุงเพิ่มเติมเพื่อให้การแปลความหมายของภาพได้ดียิ่งขึ้นอีก ก็คือการพิจารณารายละเอียดของวัตถุของภาพ ซึ่งคำศัพท์ที่นำมาใช้จะถูกจำกัดเพียงบางกลุ่มคำเท่านั้น ยังไม่หลากหลายมากเท่าที่ควร เพราะฉะนั้นควรจะมีการจัดกลุ่มของวัตถุที่เข้ามาทำการทดลองให้มีความหมายที่รัดกุมมากขึ้น และเมื่อภาพที่มีส่วนของสภาพแวดล้อมที่คล้ายคลึงกันมาก จะถูกจัดอยู่ในกลุ่มเดียวกันทั้งที่ วัตถุภายในภาพ โดยเฉพาะ วัตถุที่เป็น *มนุษย์* ที่เป็นวัตถุเด่นในภาพ ถ้ามนุษย์มีรูปร่างลักษณะท่าทางที่แปลงเปลี่ยน เช่น ยืน นั่ง นอน หรือ กระโดด ทำให้การตีความหมายของภาพโดยรวมมีความหมายที่แตกต่างกันออกไป จึงทำให้ท่าทางของมนุษย์ (human action) สามารถบอกได้ถึงความหมายของภาพโดยรวมได้เช่นกัน อาจจะต้องมีการใช้การพิจารณา อารมณ์ จากใบหน้าและท่าทางของการแสดงออกสัมพันธ์ควบคู่กันไป ซึ่งงานวิจัยในส่วนของอารมณ์มนุษย์ที่แสดงออกทางใบหน้า และท่าทางก็ยังคงเป็นงานวิจัยยังคงวิจัยกันอยู่มาก เพราะฉะนั้นในงานวิจัยที่นำเสนอนี้เป็นอีกแนวทางหนึ่งที่พยายามจะคิดค้นวิธีการที่จะหาความหมายที่เกิดขึ้นจากภาพ ในอีกมุมมองหนึ่งซึ่งยังคงต้องมีการพิจารณาและวิเคราะห์ปรับปรุงการทดลองต่อไป.

## บรรณานุกรม

- ชรินทร์ สุนทรกุล, การรู้จำตัวอักษรภาษาไทยและภาษาอังกฤษโดยใช้โครงข่ายประสาทเทียม, วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์, 2543.
- บุญเสริม กิจศิริกุล, ปัญญาประดิษฐ์, ภาควิชาวิศวกรรมคอมพิวเตอร์, คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2546.
- พิชญ์ฉัตรณ์ เมืองงาม, ปัทมา ผาโคตร, กวีพจน์ บรรลือวงศ์, การทำนายผลสำเร็จการศึกษาของ นักศึกษาระดับอุดมศึกษาด้วยการเรียนรู้แบบเบย์และการทำเหมืองข้อมูล, 3rd National cs&ict conference, กุมภภาพันธุ์, 2552.
- A. Gupta and R. Jain, "Visual Information Retrieval", *Communications of ACM (CACM)*, 40, 5, May 1997, 70-79.
- Aditaya Vailaya, M. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image Classification for Content-Based Indexing", *IEEE Transaction on Image Processing*, 10, 1, Jan 2001, 117-130.
- A. E. Savakis, S. C. Etz, and A. Loui, "Evaluation of Image Appeal in Photographic Images," SPIE Human Vision and Electronic Imaging V, January 2000.
- Aditaya Vailaya, M. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image Classification for Content-Based Indexing", *IEEE Transaction on Image Processing*, 10, 1, Jan 2001, 117-130.
- Allan Kuchinsky, Celine Pering, Michael L. Creech, Dennis Freeze, Bill Serra, and Jacek Gwizdka, "FotoFile: A Consumer Multimedia Organization and Retrieval System", *In Proceedings ACM CHI'99*, Pittsburgh, May, 1999.
- Andrew Payne and Sameer Singh, "Indoor vs. outdoor scene classification in digital photographs," *Pattern Recognition*, vol. 38, issue 10, October 2005, pp. 1533-1545.
- Benitez, A.B., and S.-F. Chang, "Semantic Knowledge Construction From Annotated Image Collections", *International Conference On Multimedia & Expo (ICME-2002)*, Lausanne, Switzerland, Aug 26-29, 2002.

- Benitez, A.B., S.-F. Chang, and J. R. Smith, "IMKA: A Multimedia Organization System Combining Perceptual and Semantic Knowledge", *In Proceeding of the 9th ACM International Conference on Multimedia (ACM MM-2001)*, Canada, Ottawa, 2001.
- C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, "Blobworld: A system for region-based image indexing and retrieval," *In Proceedings of the International Conference Visual Information System*, Amsterdam, The Netherlands, 1999, (509-516).
- Castleman, K.R., *Digital Image Processing*. 1996, N.J.: Prentice-Hall.
- Cheng, Y.C., Chen, S.Y., "Image Classification using Color, Texture and Regions," *IVC(21)*, No. 9, September , 2003, 759-776.
- Chester M., *Neural Network: A Tutorial*, Prentice-Hall Inc., 1993, New Jersey.
- Chinpanchana S. and B. Thipakorn, "Semantic Classification of Personal Images Based on Human Action and Associate Bayesian Rule", *International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Sendai, Japan, 2004.
- Flood, I., and Kartam, N., "Neural Network in Civil Engineering-I: Principles and Understanding", *Journal of Computing in Civil Engineering ASCE*, 1994, 131-148.
- Fotosearch Stock Photography, World Wide Web: <http://www.fotosearch.com>
- H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes", *In Proceedings of the 7th IEEE International Conference on Tools with Artificial Intelligence*, pp. 336–391. 1995.
- Ismail Haritaoglu, "A Real Time System for Detection and Tracking of People and Recognizing Their Activities", Phd. Proposal, University of Maryland at College Park, 1998.
- Ismail Haritaoglu, D.Harwood, and L.Davis, "Ghost: A Human Body Part Labeling System Using Silhouettes", *In Proceedings of the 4th International Conference on Pattern Recognition*, Brisbane, August 1998.

- J. Luo and A. Savakis, "Indoor vs. Outdoor Classification of Consumer Photographs," *International Conference on Image Proc. ICIP'01*, Thessaloniki, Greece, Oct. 2001.
- J.R. Smith and S.-F. Chang, "VisualSEEk: A Fully Automated Content-Based Image Query System", *In Proceedings of the ACM Multimedia*, Nov, 1996, 87-98.
- Jain, A.K., Vailaya, A., "Image Retrieval Using Color and Shape," *Pattern Recognition*, vol.29, no. 8, August 1996,1233-1244.
- Joo-Hwee Lim and Qi Tian, "Home photo content modeling for personalized event-based retrieval", *IEEE Multimedia*, vol. 10,no. 4, Oct-Dec, 2003, 28 - 37.
- Kobus Barnard and David Forsyth, "Exploiting Image Semantics for Picture Libraries", *The First ACM/IEEE-CS Joint Conference on Digital Libraries*, 2001.
- Lippmann, R.P., 1988. An Introduction to Computing with Neural Nets, In Vemuri, V., Editors, *Artificial Neural Networks; Theoretical Concept*, The Computer Society: 36-54.
- Liu Wenyin, Yanfeng Sun, Hongjiang Zhang, " MiAlbum - a system for home photo managemet using the semi-automatic image annotation approach", *In Proceedings of the ACM Multimedia*, Oct, 2000.
- M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, et al., "Query by Image and Video Content: The QBIC System", *IEEE Computer*, vol 28, no. 9, Sept, 1995, 23-32.
- N. Serrano, A. Savakis, and J. Luo, "A Computationally Efficient Approach to Indoor/Outdoor Scene Classification," *International Conference on Pattern Recognition*, ICPR'02, Quebec City, Canada, Aug. 2002.
- P.S. Hiremath, Jagadeesh Pujari, "Content Based Image Retrieval Using Color, Texture and Shape Features", *International Conference on Advanced Computing and Communications (ADCOM 2007)*, Dec, 2007, 780-784.
- Philippe Mulhem, Joo Hwee Lim, "Symbolic photograph content-based retrieval", *In Proceedings of the eleventh international conference on Information and knowledge management*, McLean, Virginia, USA, 2002, 94 – 101.

- Qian Huang, B. Dom, D.Steele, J. Ashley and W. Niblack, “Foreground/Background Segmentation of Color Images by Integration of Multiple Cues”, *Proceedings of the IEEE International Conference on Image Processing (ICIP95)*, 1995.
- R. Zhao and W. I. Grosky, “Narrowing the Semantic Gap–Improved Text-Based Web Document Retrieval Using Visual Features”, *IEEE Transactions on Multimedia*, Vol. 4, No. 2, 2002, 189-200.
- R.C. Holte, “Very simple classification rules perform well on most commonly used datasets”, *Machine Learning*, Vol. 11, No. 1, 1993, 63–90.
- Richard O. Duda, Peter E. Hart and David G. Stork, *Pattern classification*, New York, Wiley, 2nd, 2001.
- Rudolph Arnheim, *Art and Visual Perception A Psychology of the Creative Eye*, University of California Press, Ltd., 1974, 11-15.
- Smith, M., 1993. *Neural Networks for Statistical Modeling*, Van Nostrand Reinhold, New York.
- Sonka M., Hlavac V. , Boyle R., 2007, *Image Processing, Analysis, and Machine Vision*, 3rd edition, Thomson Learning, Toronto, April, 821 p, ISBN 049508252X (2nd edition Brooks/Cole, Pacific Grove, CA, 1999, 1st edition Chapman & Hall, London 1993.
- Svoboda T., Kybic J., Hlavac V., 2007, *Image Processing, Analysis, and Machine Vision, A MATLAB Companion*, Thomson Learning, Toronto.
- Tele Tan, Jiayi Chen, Philippe Mulhem and Mohan Kankanhalli, “SmartAlbum – A Multi-Modal Photo Annotation System”, *In Proceedings of the ACM Multimedia*, Juan-les-Pins, France, Dec, 2002.
- The Corbis Stock Photography and Pictures, World Wide Web: <http://pro.corbis.com/>
- The Corel Corporation, World Wide Web: <http://www.corel.com/>
- Vailaya, A. Figueiredo, M.A.T. Jain, A.K. Hong-Jiang Zhang, “Image classification for content-based indexing,” *IEEE Transactions on Image Processing*, vol. 10, issue 1, Jan 2001.

- Vapnik, Vladimir N., 1989, *Statistical Learning Theory*, Wiley-Interscience.
- ., 2000, *The Nature of Statistical Learning Theory*, Information Science and Statistics, Springer-Verlag.
- Vasileios Mezaris , Ioannis Kompatsiaris , Michael G. Strintzis, “An Ontology Approach to Object-Based Image,” *In Proc. IEEE International Conference on on Image Processing*, 2003.
- W. Ma, and B. Manjunath, “NETRA: A Toolbox for Navigating Large Image Database”, *In Proceedings of the IEEE International Conference on Image Processing (ICIP'97)*, 1, 568-571, Santa Barbara, 1997.
- Yanfeng Sun, Hongjiang Zhang, Lei Zhang, Mingjing Li, “MyPhotos – A System for Home Photo Management and Processing”, *In Proceedings of the ACM Multimedia*, France, Dec, 2002.
- Zhao T, Tang L H, Ip H, Qi F., “Visual Keyword Image Retrieval Based on Synergetic Neural Network for Web-Based Image Search”, *Journal of Real-Time Systems*, Vol.21, Issue 1/2 , 2001, 127 – 142.



## ประวัติผู้วิจัย

### อาจารย์นศัพธ์ชาณัณ ชินปัญชธนะ

วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์) มหาวิทยาลัยหอการค้าไทย  
 สถิติประยุกต์ (คอมพิวเตอร์) สถาบันบัณฑิตพัฒนบริหารศาสตร์

### ปัจจุบัน

อาจารย์ประจำภาควิชาเทคโนโลยีสารสนเทศธุรกิจ คณะเทคโนโลยีสารสนเทศ  
 มหาวิทยาลัยธุรกิจบัณฑิต  
 คณะกรรมการกลุ่มผลิตชุดวิชาการสื่อสารข้อมูลและระบบเครือข่ายคอมพิวเตอร์  
 สาขาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

### รางวัลและเกียรติประวัติ

ได้รับทุนการศึกษาประเภทเรียนดี มหาวิทยาลัยหอการค้าไทย พ.ศ. 2534 – 2537  
 ได้รับปริญญาตรีเกียรตินิยมอันดับ 1 จาก มหาวิทยาลัยหอการค้าไทย

### ประสบการณ์งานสอน

อาจารย์ประจำ  
 มหาวิทยาลัยธุรกิจบัณฑิต  
 ภาควิชาเทคโนโลยีสารสนเทศธุรกิจ คณะเทคโนโลยีสารสนเทศ

- วิชาวิชาการพัฒนาและเขียนโปรแกรมบนเว็บ (2550-2552) (Tool: html, java script)
- วิชาสำนักงานอิเล็กทรอนิกส์ (2550-2552) (Tool: Ms. Office)
- วิชาธุรกรรมอิเล็กทรอนิกส์เบื้องต้น (2/2552) (Tool: Ms. Office , Dreamwaver)
- วิชาการเขียนโปรแกรมเบื้องต้น (2/2550) (Tool: Java program)

## มหาวิทยาลัยธุรกิจบัณฑิต

## ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ

- วิชาการประมวลผลภาพ (Image processing) (2/2552) (Tool: Matlab program)
- วิชาโครงการ (IT project) (2551-2552)
  - โครงการ ระบบสนับสนุนการตัดสินใจในการเลือกอาชีพ (Online Career and Education Planning System) 2552
  - โครงการ ระบบสนับสนุนการตัดสินใจในการเลือกทัวร์ที่เหมาะสมกับลูกค้า (Intelligent Expert Tour System) 2552
  - โครงการ ระบบวิเคราะห์ดูแลสุขภาพ ( Health Care System Analysis) 2551
  - โครงการ ฐานข้อมูลเพื่อเรียกดูเส้นทางในเขตหลักสี่ (Route browsing database for Laksi ) 2551

## มหาวิทยาลัยธุรกิจบัณฑิต

## ภาควิชาคอมพิวเตอร์ คณะบริหารธุรกิจ

- วิชาการจัดการฐานข้อมูล (Database management system) (2542-2543)
- วิชาการเขียนโปรแกรมภาษา ซี (1/2542) (Tool: C program)
- วิชาการใช้โปรแกรมสำเร็จรูป (2/2541)
- วิชาคอมพิวเตอร์เบื้องต้น (Introduction to computer) (1/2541)

## มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ

## สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี

- วิชาโครงสร้างข้อมูล (Data Structure) (1/2541) (Tool: Borland Turbo C Programming language)

## อาจารย์พิเศษ

- มหาวิทยาลัยศรีปทุม บรรยายและปฏิบัติการ วิชาโครงสร้างข้อมูล (Data Structure) (Tool: C++ program)
- มหาวิทยาลัยเกริก บรรยายวิชา คลังข้อมูล (Data warehouse)
- วิทยาลัยภาคกลาง จ.นครสวรรค์ บรรยาย วิชาความปลอดภัยข้อมูล (Computer Security )

## งานวิจัยที่สนใจ

สนใจงานวิจัยทางการค้นคืนความหมายภาพ (Image retrieval) ทั้งภาพธรรมชาติทั่วไป (natural images) และ ภาพที่เป็นส่วนบุคคล (personal images) ในหัวข้อเกี่ยวกับการค้นคืน human activity หรือในหัวข้อการแปลความหมายภาพ (semantic human image) โดยเจาะจงทางด้านการใช้ท่าทางของมนุษย์เพื่อแสดงถึงความหมายของภาพ รวมทั้งการจำแนกภาพต่างๆ เป็นกลุ่ม (Image classification) และการประยุกต์หลักการประมวลผลภาพ เพื่อนำมาใช้ในทางอุตสาหกรรม ได้จริง

## ผลงานทางวิชาการ

N. Chinpanthana, Semantic Salient Images Based on Similarity Matching with Conceptual Graph, International Technical Conference on Circuits/Systems, Computers and Communications , Jeju, Korea 2009.

S. Chinpanchana, S. Maneewongvatana, and B. Thipakorn, "Semantic Human Image Classification Based on Energy Action Model with Essential Reference points," Int. Symposium on Communications and Information Technologies, 16-19 Oct, Sydney, AUS, 2007.

S. Chinpanchana, S. Maneewongvatana, and B. Thipakorn, "High-Level Semantic Image Classification by Using Energy Expenditure," 2007 International Workshop on Smart Info-Media Systems in Bangkok, November 1-2, 2007, Bangkok, Thailand.

S. Chinpanchana, "Semantic Human Action Classification Based on Energy-Action Model ", Tencon 2006 IEEE Region 10 , Hongkong, China, 2006.

S. Chinpanchana, S. Maneewongvatana, and B. Thipakorn, "Semantic Personal Image Classification by Energy Expenditure," Int. Symposium on Communications and Information Technologies, Beijing, China, 2005.

S. Chinpanchana, S. Maneewongvatana and B. Thipakorn, "Semantic Personal Image Pattern Classification Based on Human Body," Asia Information Retrieval Symposium, Beijing, China, Oct. 2004.

S. Chinpanchana and B. Thipakorn, "Semantic Classification of Personal Images Based on Human Action and Associate Bayesian Rule," International Technical Conference on Circuits/Systems, Computers and Communications, Sendai, Japan, 2004.

---

