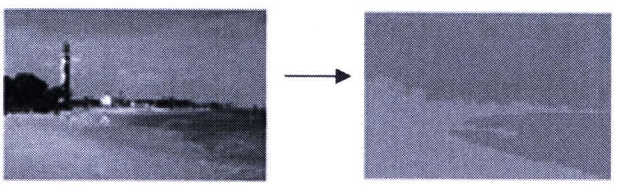
บทที่ 2
ทฤษฎีและงานวิจัย

2.1 ทฤษฎีเบื้องต้น

วิวัฒนาการของเครื่องมือและอุปกรณ์ถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็วจนทำให้ภาพถ่ายภาพดิจิทัล มีจำนวนเพิ่มมากขึ้น ปัญหาที่ตามมาก็คือการจัดเก็บข้อมูลภาพที่เพิ่มมากขึ้นอย่างไร้ขีดจำกัดนี้จะทำอย่างไรจึงจะสามารถจัดเก็บอย่างมีระบบและสามารถสืบค้นข้อมูลภาพ และจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้งานวิจัยในปัจจุบันที่เกี่ยวข้องกับการค้นคืนรวมทั้งการจัดกลุ่มภาพให้ตรงกับความต้องการเพิ่มมากขึ้น รวมไปถึงระยะเวลาในการสืบค้นที่น้อยลงกับปริมาณของภาพที่เพิ่มทวีคูณ ดังนั้นปัญหาดังกล่าวมาข้างต้นนั้นจึงได้รับความสนใจจากนักวิจัยหลายกลุ่ม ซึ่งเป็นงานด้านการประมวลผลภาพ (image processing) ด้านการค้นคืนสารสนเทศ (image retrieval) เป็นอีกหนทางหนึ่งในการแก้ปัญหาดังกล่าว เนื่องจากสามารถช่วยให้การค้นหาข้อมูลกระทำได้โดยสะดวกยิ่งขึ้นรวมถึง การค้นหาภาพและการจำแนกประเภทข้อมูลภาพ (image classification) เพื่อคัดเลือกภาพ เพื่อให้ตรงตามความต้องการของผู้ใช้ยิ่งมีความจำเป็นอย่างยิ่ง

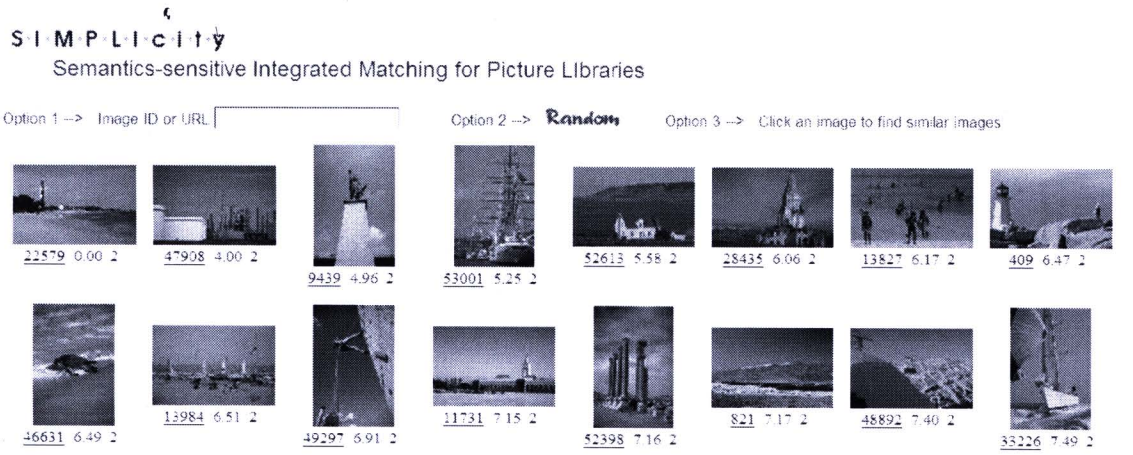
สำหรับงานวิจัยทางการประมวลผลภาพในการค้นคืนสารสนเทศกลุ่มแรก ๆจะมีการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่าง ๆ ยกตัวอย่างเช่น สี (color) ลวดลาย (texture) รูปทรง (shape) เป็นต้น กระบวนการนี้ถูกเรียกว่า การประมวลผลภาพระดับต่ำ (low-level image processing) [Jain A.K,1996][Cheng Y.C,2003] กระบวนการนี้สามารถค้นหาภาพได้ตามคุณลักษณะพื้นฐานที่นำไปสืบค้น โดยภาพผลลัพธ์ส่วนใหญ่มักจะเป็นภาพที่มีคุณลักษณะไม่ซับซ้อนมากนัก เช่น โทนสี หรือ รูปทรงที่แตกต่างกันอย่างเด่นชัด ดังภาพที่ 2.1 แสดงการค้นคืนของ SIMPLIcity [Jia Li,2003][James Z. Wang,2001] ด้วยคุณลักษณะฟีเจอร์ระดับต่ำด้วยสี ลวดลาย และตำแหน่งของพื้นที่ของภาพ จากผลลัพธ์จะสังเกตว่าผลลัพธ์ของภาพเป็นภาพที่มีโทนสีคล้ายกันเป็นหลัก แต่มีลักษณะวัตถุที่แตกต่างกันอย่างสิ้นเชิงที่อยู่ในหมวดหมู่เดียวกัน คือ ภาพชายหาด หรือ ชายทะเล แต่มีคุณลักษณะของโทนสี และลวดลาย ที่แตกต่างกันอย่างเห็นได้ชัดเจน

เมื่อนำมาจำแนกด้วยการประมวลผลภาพระดับต่ำ (low-level feature) แล้วนั้นค่อนข้างยากที่จะจัดให้หมวดหมู่เดียวกัน



ภาพต้นฉบับ ภาพต้นฉบับที่ถูกแปลงเป็นฟีเจอร์

ก. ภาพต้นฉบับสำหรับการค้นคืนที่ถูกแปลงเป็นฟีเจอร์ระดับต่ำ



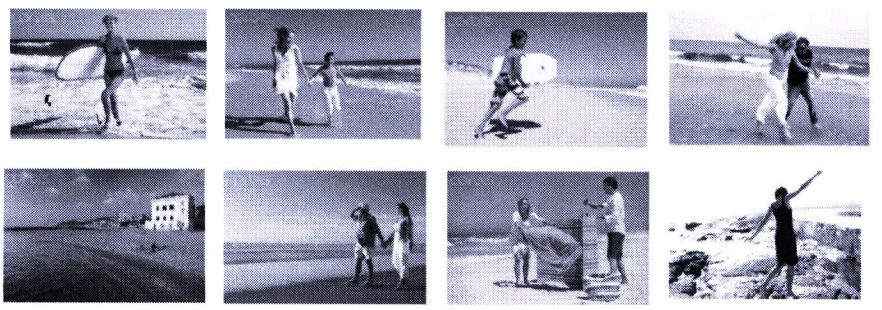
ข. ผลลัพธ์ของการค้นคืนด้วยภาพต้นฉบับ ก.

ภาพที่ 2.1 การประมวลผลแบบค้นคืนด้วยคุณลักษณะฟีเจอร์ระดับต่ำ[1]

แต่อย่างไรก็ตามได้มีกลุ่มนักวิจัยที่พยายามปรับปรุงแปลงอัลกอริทึมด้วยการประมวลผลภาพระดับต่ำ เพื่อทำการค้นคืนภาพที่มีลักษณะฟีเจอร์ที่ใกล้เคียงกับภาพที่ต้องการมากที่สุด [Andrew, 2005][M. Flickner, 1995][W. Ma, 1997] การปรับปรุงเทคนิควิธีการเพื่อให้กระบวนการค้นคืนภาพได้ผลลัพธ์ที่ถูกต้องเพิ่มมากขึ้น ด้วยการนำวิธีการมาผสมผสานกันระหว่างคุณลักษณะเพื่อให้สามารถวิเคราะห์ในรูปแบบที่ซับซ้อนได้มากขึ้น เช่นการรวมเทคนิคด้วยคุณลักษณะสีและรูปทรงของภาพเพื่อทำการค้นคืนภาพ [P.S. Hiremath, 2007] หรือมีการใช้อัลกอริทึมเพื่อทำการสกัดข้อมูลภาพเป็นฟีเจอร์ และนำฟีเจอร์นำมาใช้ในการ ค้นคืนภาพในรูปแบบที่แตกต่างกันออกได้ แต่ในความเป็นจริงแล้วนั้นลักษณะการมองภาพของคนโดยทั่วไปเป็นการมองจากความหมายของภาพ หรือ

[1] http://alipr.com/cgi-bin/zwang/regionsearch_show.cgi

มองจากชนิดของวัตถุของภาพ ดังแสดงในภาพที่ 2.2 แสดงภาพชายทะเล ภายในภาพประกอบด้วยวัตถุหลายชนิดได้แก่ ทะเล และ ท้องฟ้า เพียงอย่างเดียว แต่บางภาพอาจจะประกอบด้วยวัตถุอื่น ๆ มนุษย์ สุนัข หรืออาคาร เป็นภาพประเภทเดียวกันหรือภาพที่สื่อความหมายอย่างเดียวกัน โดยที่ไม่จำเป็นต้องมีคุณลักษณะสีหรือรูปทรงแบบเดียวกันก็สามารถเป็นภาพชนิดเดียวกันได้ ดังนั้นในการค้นคืนที่ใช้คุณลักษณะฟีเจอร์ระดับต่ำเพียงอย่างเดียว อาจจะได้ผลลัพธ์ที่ไม่ตรงกับความต้องการอย่างแท้จริง



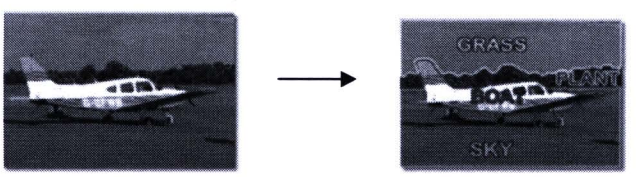
ภาพที่ 2.2 ตัวอย่างขององค์ประกอบของภาพ[2]

การวิจัยในส่วนแรกเป็นความพยายามในการค้นหาข้อมูลภาพเริ่มมาจากการสกัดข้อมูลภายในภาพ (feature extraction) ออกมาเป็นส่วน ๆแล้วใช้ข้อมูลระดับต่ำเพียงแบบเดียว เพื่อทำการสืบค้นข้อมูลภาพเท่านั้น วิธีการค้นคืนด้วยข้อมูลระดับต่ำเพียงแบบเดียวแล้วนั้นจะทำให้ได้ผลลัพธ์ของภาพที่มีความหมายไม่ครบถ้วน ทำให้ต้องมีการปรับปรุงพัฒนาอย่างต่อเนื่อง ด้วยเหตุนี้เองทำให้มีกลุ่มนักวิจัยที่สนใจในเรื่องของ การจัดกลุ่มภาพตามความหมายเรียกว่า semantic image เป็นการจัดกลุ่มภาพตามความหมายของภาพโดยแปลความหมายของภาพจากองค์ประกอบ หรือวัตถุ (object) ที่ปรากฏในภาพนั้น ซึ่งมีวิธีการค้นคืนที่แตกต่างกันเพื่อให้ได้ผลลัพธ์ที่ตรงตามความต้องการมากที่สุด

แต่อย่างไรก็ตามได้มีงานวิจัยอีกกลุ่ม ที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้น เรียกว่า, การประมวลผลภาพระดับสูง (high level image processing) งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ (object) ที่มีความหมาย [Benitez A.B,2002] [Galleguillos C.,2010] และแทนวัตถุนั้น ๆ ด้วยคำหลัก (keyword) บน

[2] http://www.corbisimages.com/

ภาพ เรียกว่า การแท็ก (tag) หรือการให้ความหมายของวัตถุบนภาพเป็น ชื่อวัตถุ หรือคำศัพท์ ที่สอดคล้องกันเช่น “grass”, “plant” , “boat”, “sky” เป็นต้น ดังแสดงในภาพที่ 2.3 [Galleguillos C., 2010] และใช้ความหมายหรือคำศัพท์นั้นเพื่อทำการสืบค้นข้อมูลแทน ซึ่งเป็นการใช้ความหมายของคำศัพท์ที่มีสอดคล้องกันด้วยความหมายตามพจนานุกรม หรือในลักษณะใช้ความสัมพันธ์ของความหมายที่เหมือนกันของคำหลัก (synonym) [Zhao T., 2001] [Benitez, A.B., 2002] [Kobus B., 2001] [Philippe M., 2002] เข้ามาใช้ในการค้นคืนข้อมูลภาพเพื่อให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ ยกตัวอย่างเช่น “stone” มีความหมายสอดคล้องกันกับ “rock” เป็นต้น จะได้ผลที่ค่อนข้างดีกว่า แต่ขึ้นอยู่กับว่าอัลกอริทึมที่ถูกนำมาใช้นั้นจะเป็นลักษณะใด

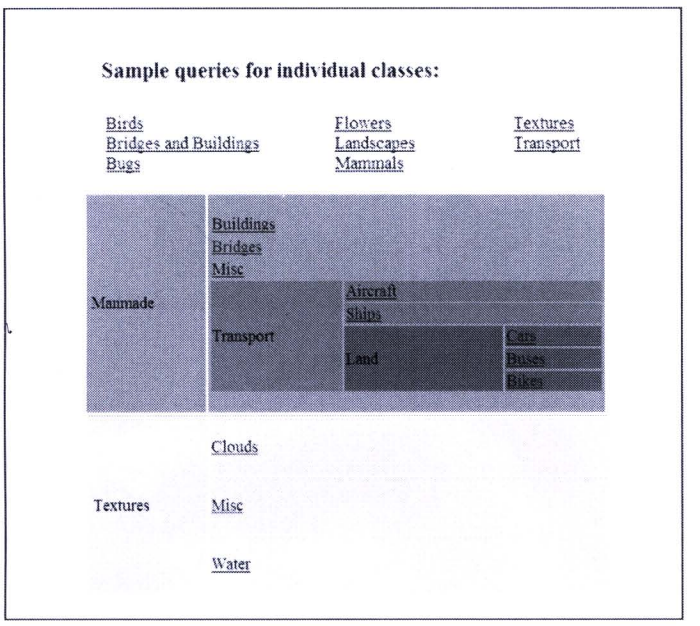


ภาพที่ 2.3 ตัวอย่างการแท็กองค์ประกอบของภาพ[3]

นักวิจัยบางกลุ่มใช้เทคนิคเพื่อสร้างความสัมพันธ์ (relationship) ของแท็กชื่อวัตถุบนภาพ [Benitez A.B, 2001] [R. Zhao, 2002] [Philippe M., 2002] ด้วยโมเดลเหตุการณ์ (event model) [Joo-Hwee L., 2003] เช่น การเชื่อมวัตถุด้วยคำต่างๆ เช่น “touch” “on” “top” เป็นต้น บางงานวิจัยได้พยายามใส่ข้อความ (context) เพื่อบรรยายความหมายของภาพ [Mathias Lux, 2003] [Mathias Lux, 2009] ในหัวข้อที่สอดคล้องกับภาพนั้นๆ เช่น “birthday party of uncle Adam” หรือ “a picture showing a barking dog” แต่ภาพที่นำมาใช้จัดเก็บนั้นมักจะเป็นจะภาพส่วนตัว (personal images) ทำให้การค้นคืนจำกัดเพราะคำบรรยายส่วนใหญ่จะเป็นคำเฉพาะเจาะจงจึงไม่นิยมเท่าที่ควร บางงานวิจัยพยายามที่จะใส่ข้อมูลเหตุการณ์ต่างๆจนครบถ้วนในรูปแบบของ ใคร ทำอะไร ที่ไหน เมื่อไหร่ (who, what, when, where) ดังนั้นในการใส่ข้อมูลบางครั้ง เป็นข้อมูลที่นอกเหนือ หรือเกินความจำเป็นโดยใช่เหตุ ข้อมูลเหล่านี้อาจจะไม่มีความจำเป็นเลยสำหรับการสืบค้นข้อมูล อาจจะเป็นข้อมูลที่มีความเป็นส่วนตัวจนเกินไป และใช้คำศัพท์ซ้ำซ้อน ฟุ่มเฟือย ทำให้การค้นคืนยิ่งเกิดความสับสนปนเป กันของภาพผลลัพธ์ที่ได้มา ดังแสดงในภาพที่ 2.4 ก. [Q. Iqbal, 2002] แสดงตัวอย่างการค้นคืนภาพที่มีการจัดเป็นหัวข้อ (topic) จากภาพที่ 2.4 ข. แสดงภาพต้นฉบับเพื่อ

[3] http://vision.ucsd.edu/project/context-based-object-categorization

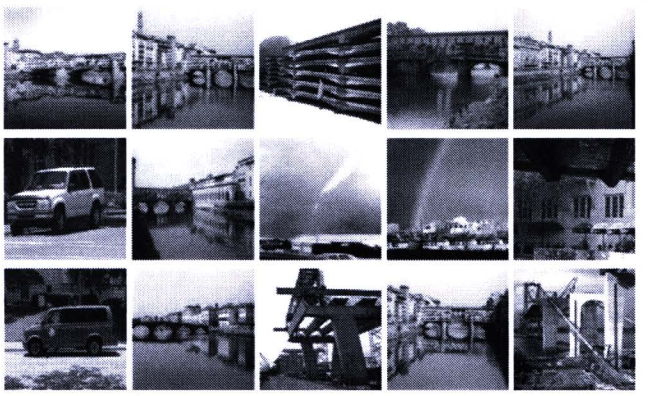
สร้างคุณลักษณะหรือฟีเจอร์ที่คล้ายกันเพื่อใช้ในการค้นคืนในที่เก็บข้อมูลภาพ จากภาพที่ 2.4 ค. แสดงภาพต้นฉบับมีคำหลักเป็น “bridges” และกลุ่มผลลัพธ์ที่ได้หลังจากการใช้ฟีเจอร์ในการค้นคืนภาพ แต่ผลลัพธ์ที่ได้ออกมานั้นจะมีการผสมกับภาพอื่น ๆที่ไม่ตรงกับคำหลักด้วย



ก. การค้นคืนด้วยการเลือกคำหลักที่เป็นหมวดหมู่

ข. การค้นคืนด้วยการเลือกภาพต้นฉบับ

ภาพต้นฉบับ



ผลลัพธ์ของการค้นคืน

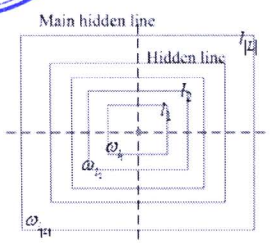
ค. ผลลัพธ์ของการค้นคืนด้วยภาพต้นฉบับ

ภาพที่ 2.4 การแท็กองค์ประกอบของภาพ[4]

การค้นหาภาพด้วยเทคนิคนี้จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกแท็กไว้บนภาพยิ่งมีการแท็กข้อมูลบนภาพมากยิ่งสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่ในความเป็นจริงแล้ว การแท็กข้อมูลบนภาพในปัจจุบันนั้นเป็นเพียงการหาคำศัพท์ที่ต้องการบนภาพ แต่ไม่ได้ให้ความหมายภาพโดยรวม ความหมายของภาพคือการนำวัตถุที่ปรากฏบนภาพมารวมกันเพื่อวิเคราะห์จากความคิดของมนุษย์เพื่อให้ได้ผลลัพธ์คือคำศัพท์ใหม่ที่แทนความหมายของภาพทั้งภาพ สำหรับการวิเคราะห์ความหมายของภาพจะได้จากการรับรู้ของมนุษย์ที่มองภาพนั้น จะมีกฎเกณฑ์ของการรับรู้จากการมองภาพได้กล่าวไว้เป็นทฤษฎี โครงสร้างสเกตตรีตรอน ดังแสดงในภาพที่ 2.5 ก. [Rudolph A., 1974] ได้มีการนำมาประยุกต์ใช้วิเคราะห์ในส่วนของเฉพาะการการจำแนกความหมายของภาพโดยรวมเพื่อให้ได้ความหมายของภาพอย่างแท้จริง โดยใช้ทฤษฎีการมองของมนุษย์ที่มีการพิจารณาจากตำแหน่งและขนาดของวัตถุที่เกิดขึ้นเพื่อนำมาใช้ในการแปลความหมายของภาพ ด้วยการวัดโครงสร้างกับภาพต้นฉบับที่มีการแท็ก ดังแสดงในภาพที่ 2.5 ข. และ ค. [N. Chinpanthana., 2010] [นัศพ์ชาณัณ 2552] แต่อย่างไรก็ตามการค่าตำแหน่งในส่วนของภาพนั้นยังไม่เพียงพอต่อการนำมาใช้สำหรับการแปลความหมายภาพ

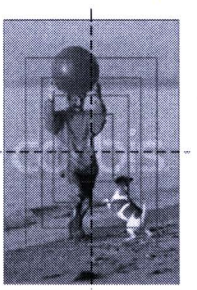
[4] http://amazon.ece.utexas.edu/~qasim/sample_queries.htm

สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย



ก. โครงสร้างสเกตตรีตรอน

ข. ภาพแท็กด้วยคำหลัก



ค.แม๊พด้วยโครงสร้างสเกตตรีตรอน

ภาพที่ 2.5 การหาตำแหน่งวัตถุของภาพด้วยโครงสร้างสเกตตรีตรอน

ยังมีวิธีการที่ผสมผสานระหว่างการใช้แท็กคำหลักบนภาพกับการสร้างความสัมพันธ์ระหว่างวัตถุด้วยมัลติเปิลเคอร์แนล (multiple kernel) ของพื้นที่วัตถุบนภาพ [Galleguillos C., 2011] [Galleguillos C., 2010] ผลลัพธ์ที่ได้จะเป็นกลุ่มของคำหลักที่เกิดขึ้นบนภาพเสียส่วนใหญ่ และมีนักวิจัยบางกลุ่มพยายามที่แก้ไขถึงความยุ่งยากลำบากในการใส่ข้อมูลบนภาพ (image annotation) และจำกัดขอบเขตคำศัพท์ของภาพให้รัดกุมยิ่งขึ้น จึงได้พัฒนาซอพต์แวร์เพื่อให้เกิดความสะดวกสบายในการใส่ข้อมูล และการใส่ข้อมูลที่มีแบบแผนแน่นอน ยกตัวอย่างเช่น Caliph & Emir [Mathias Lux, 2009], Annosearch [Xin-Jing, 2008], CAMEL [Apostol Paul N., 2001] เป็นต้น แต่อย่างไรซอพต์แวร์ส่วนใหญ่นั้น จะทำการสืบค้นภาพในรูปแบบของการเทียบคำหลักเป็นคำต่อคำตามที่มีการเก็บข้อมูลไว้เท่านั้น เพราะฉะนั้นคำตอบที่ได้จะไม่ได้ขึ้นกับความหมายของภาพแต่ขึ้นกับคำศัพท์ที่มีการเก็บข้อมูลลงไปเท่านั้น ความหมายของภาพยังไม่ได้อย่างแท้จริง เพราะฉะนั้นผลลัพธ์ของกลุ่มภาพที่ได้จะไม่ได้ขึ้นกับความหมายของภาพอย่างแท้จริง แต่ขึ้นกับคำหลักที่มีการเก็บข้อมูลลงไปบนภาพเท่านั้น จึงทำให้ความหมายของภาพโดยรวมนั้นยังไม่ได้สื่อออกมาให้เห็นอย่างเด่นชัด ซึ่งในความเป็นจริงแล้วความหมายของภาพจะเกิดจากการแปลความหมายรวมของทุกวัตถุที่ปรากฏบนภาพ ดังนั้นในงานวิจัยนี้ได้นำเสนอในรูปแบบของการแทนข้อมูลภาพ ด้วยความสัมพันธ์ของข้อมูลวัตถุภายในภาพ หรือเรียกว่า แนวคิดกราฟ (Conceptual Graph) ในลักษณะของกราฟที่นำเสนอนั้นจะมีแสดงความสัมพันธ์ของข้อมูลภายในภาพของวัตถุที่เกิดขึ้นทั้งหมดในรูปแบบของกราฟรวมทั้งความสัมพันธ์ (Relationship) ระหว่างวัตถุที่เกิดขึ้น และจะมีการวัดค่าความสัมพันธ์ของความเหมือนกันด้วยวิธีการที่เรียกว่า การจับคู่กราฟ (Graph Matching) ของข้อมูลภาพที่ถูกแทนค่าความหมายไว้ ดังนั้นภาพที่มีความหมาย เหมือนกันจะมีได้ค่าของความเหมือนกันมากที่สุด

สำนักงานคณะกรรมการวิจัยแห่งชาติ
ห้องสมุดงานวิจัย
วันที่ 24 ก.ค. 2555
เลขทะเบียน 247413
เลขเรียกหนังสือ

จากที่กล่าวมาข้างต้นในงานวิจัยนี้จึงได้นำเสนอ การปรับปรุงการสืบค้นข้อมูลภาพด้วยความหมายภาพเพื่อให้ได้ผลลัพธ์ของภาพที่มีอย่างแท้จริง โดยใช้วิธีการแทนข้อมูลภาพแบบแนวคิดกราฟ (Conceptual Graph) และทำเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการหาความเหมือนของภาพทั้งหมด 4 วิธีการ โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network) แผนผังการจัดระบบตัวเอง (Self-organizing maps) เครือข่ายแบบเบย์ (Bayesian Network) และการวัดความคล้ายของกราฟแบบจับคู่ (Similarity Measure with Conceptual Graph Matching)

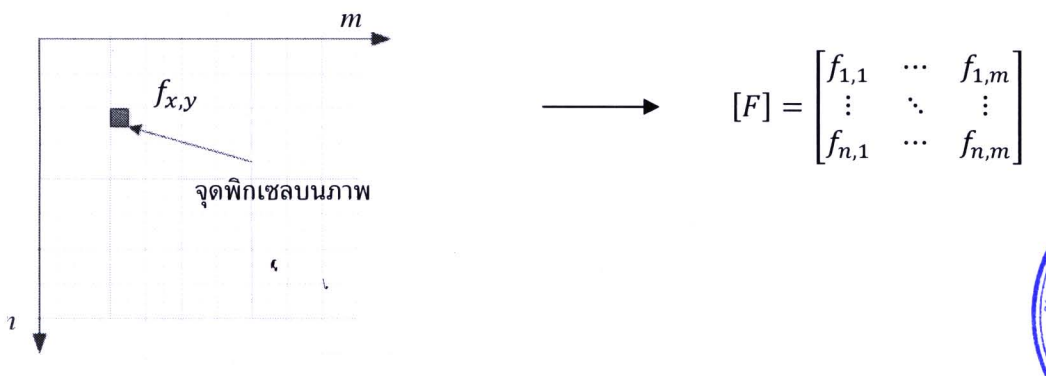
ในงานวิจัยนี้ได้นำทฤษฎีมาใช้ในการพิจารณาดังนี้

- การประมวลผลภาพดิจิทัล
- การแทนข้อมูลภาพแบบแนวคิดกราฟ
- การเปรียบเทียบความคล้ายของภาพ
- การใช้โมเดลในการสร้างการตัดสินใจ
- การวัดประสิทธิภาพ

2.2 การประมวลผลภาพดิจิทัล

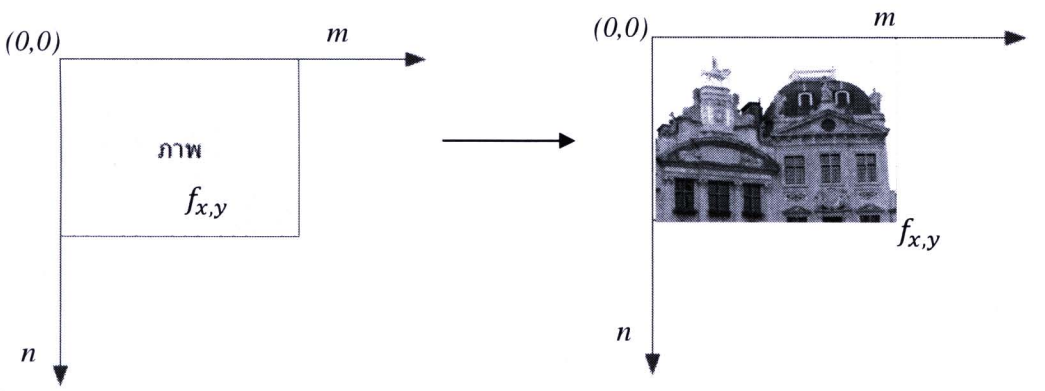
ภาพจะถูกจัดเก็บเป็นสัญญาณดิจิทัล (digital) [นัศพ์ชาณัณ 2553] [R.C. Gonzalez 2002] และทำการจัดเก็บข้อมูลภาพในหน่วยความจำ ในรูปของตัวแปรแบบอะเรย์ (array) โดยค่าในแต่ละช่องของอะเรย์แสดงถึงค่าความเข้มของแสง (intensity) ของแต่ละจุดของแต่ละจุดของภาพที่เรียกว่าพิกเซล (pixel) และตำแหน่งของช่องอะเรย์เป็นตัวกำหนดตำแหน่งของจุดภาพดังแสดงในภาพที่ 2.4 เมื่อนำภาพมาแสดงในรูปของ 2 มิติ ตามแนวแกน x และ y สามารถเขียนในรูปของฟังชันก์ของภาพสามารถแสดงในรูปชอง $i = f(x, y)$ ในรูปของ 2 มิติ โดย i หมายถึง ความสว่างหรือความเข้มของแสง และ (x, y) เป็นพิกัดในแบบจำลองของภาพ (image model) ค่าของฟังก์ชัน $f(x, y)$ ดังกล่าวเป็นค่าความเข้ม ซึ่งเปลี่ยนไปตามตำแหน่งหรือพิกัดแบบจำลองภาพ ค่าความเข้มของแสงนี้เป็นค่าดิจิทัลหรือเชิงตัวเลข ดังนั้น เมื่อแสดงเป็นตัวเลขของค่าความเข้มในแต่ละพิกเซลเรียงกันทั้งภาพจะเห็นว่าเป็นเมทริกซ์ สิ่งที่แบบจำลองต่างจากแบบจำลองทางคณิตศาสตร์ คือ จุดกำเนิด

(Coordinate Origin) ที่ตำแหน่ง (0,0) ซึ่งอยู่ที่จุดซ้ายบนสุดของสเกลต่างจากจุดกำเนิดของแบบจำลองทางคณิตศาสตร์ที่อยู่ที่จุดซ้ายล่างสุด ดังนั้นการนำเอาแบบจำลองทางคณิตศาสตร์มาประยุกต์ในการประมวลภาพ จำเป็นต้องเปลี่ยนจุดกำเนิดให้ถูกต้องด้วย ดังแสดงในภาพที่ 2.6



สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย

ภาพที่ 2.6 การแสดงตำแหน่งของพิกเซลบนภาพดิจิทัลด้วยเมทริกซ์



ภาพที่ 2.7 แบบจำลองของภาพด้วยระนาบ

ในหน่วยความจำ จะทำการจัดเนื้อที่ในการเก็บภาพ สามารถคำนวณได้จาก $m \times n \times b$ เมื่อ b เป็นจำนวนเต็มที่แทนจำนวนบิตของข้อมูลในแต่ละจุดภาพ ตัวอย่างถ้า b มีค่าเท่ากับ 8 บิต จะสามารถเก็บความแตกต่างของระดับสีที่เป็นไปสูงสุด 256 ระดับ ค่า m และ n จะเป็นตัวบอกถึงความละเอียดของภาพ สำหรับคอมพิวเตอร์ทั่วไปในระบบ VGA (Video Graphic Array) จะมีขนาด

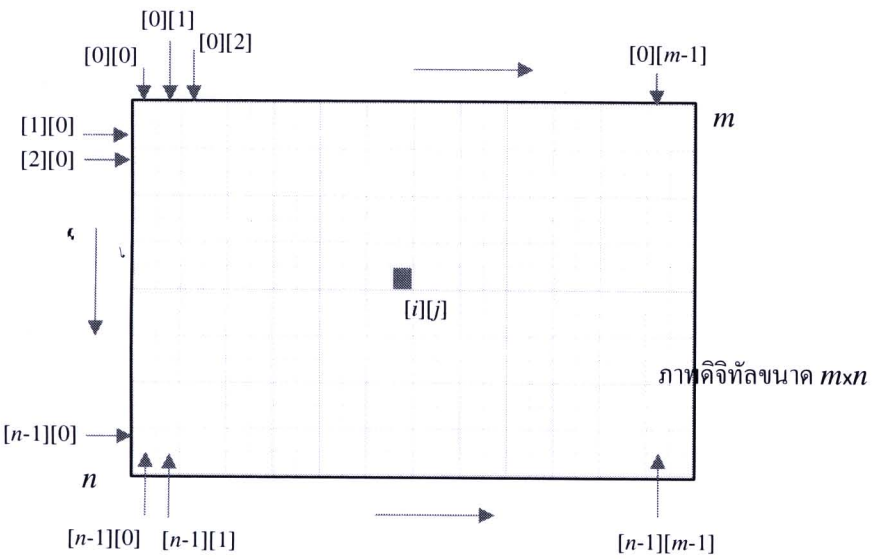
640 × 480, 800 × 600 และ 1024 × 768 จุด เป็นต้น การกำหนดความละเอียดจะขึ้นอยู่กับงานที่จะใช้ ในงานบางอย่างใช้ความละเอียดเพียง 30 × 50 จุด ก็พอซึ่งความละเอียดนั้นจะขึ้นกับงานที่จะใช้ ในบางงานจะใช้ความละเอียดถึง 1000 × 1000 จุด ก็ยังไม่พอ จากภาพที่ 2.7 สมมติให้ภาพแทนเป็นตัวแปรชื่อ x เป็นตัวแปรแบบอะเรย์ขนาด $m \times n$ (m แทน แถว และ n แทน คอลัมน์) ที่ใช้เก็บภาพขนาด $m \times n$ จุด และ ค่าของความเข้มของแสง (ค่าความสว่าง) ของจุดภาพในแถวที่ 5 คอลัมน์ที่ 4 จะตรงกับค่าของข้อมูล x เป็นรูปของ (5,4) จะเห็นว่าใช้ตำแหน่งของจุดภาพทั้งสองแกนเป็นตัวชี้ค่าข้อมูลในอะเรย์ ดังแสดงการเรียงตัวของข้อมูลบนอะเรย์ในภาพที่ 2.8 และการกำหนดความละเอียดของภาพ (image resolution) จากการกำหนดขนาดของพิกเซลตัวอย่างเช่น 1 ไมครอนต่อพิกเซล (μm/pix) 1 มิลลิเมตรต่อพิกเซล (mm/pix) เป็นต้น ในงานที่ต้องการทราบตำแหน่งหรือขนาดของวัตถุที่วัดเป็นค่าจริง เราสามารถที่จะคำนวณได้จาก

$$Resolution = \frac{\text{Field of vision in Y direction (mm)}}{\text{Number of pixels in Y direction}}$$

โดยปกติแล้วในการเก็บข้อมูลภาพโดยเครื่องมือต่าง ๆ จะเก็บตามมาตรฐานของโทรทัศน์ซึ่งมีอัตราส่วน x ต่อ y เท่ากับ 4:3 สำหรับเครื่องมือเก็บข้อมูลภาพที่ไม่เป็นไปตามอัตราส่วน 4:3 เมื่อนำภาพนี้ไปแสดงในจอภาพมาตรฐาน จะทำให้ภาพที่แสดงนั้นมีขนาดของจุดภาพไม่เป็นสีเหลี่ยมจัตุรัส เช่น ในบางระบบอาจจะใช้ความละเอียดในการแสดง เท่ากับ 640 × 580 ซึ่งจะทำ ให้ขนาดของจุดภาพที่ได้มีขนาดของด้านกว้างมีความยาวมากกว่าด้านสูง เมื่อมีการกำหนดให้ขนาดของบิตต่อจุดมากขึ้น จะทำให้จำนวนของสีมากขึ้นด้วย ตัวอย่างเช่น 1 บิต = 21 จะได้ 4 สี 2 บิต = 22 จะได้ 4 สี 4 บิต = 24 จะได้ 16 สี 8 บิต = 28 จะได้ 256 สี 16 บิต = 216 จะได้ 65536 สี เป็นต้น

จากคุณลักษณะของภาพดิจิทัลที่กล่าวมาข้างต้นนั้น เป็นการเก็บข้อมูลภาพเป็นแบบ เมตริกซ์ที่มีถึง 3 ระนาบ (dimension) การประมวลผลภาพดิจิทัล (digital image processing) เ ป็ น การเรียกใช้ขั้นตอนหรือกระบวนการที่มากระทำบนภาพ โดยมีวัตถุประสงค์เพื่อปรับปรุงคุณภาพของภาพ ให้ได้ภาพใหม่ที่มีคุณสมบัติตามวัตถุประสงค์ที่นำไปใช้งาน เช่น การปรับให้ภาพมีความคมชัดมากขึ้น (enhancement) หรือการบีบอัดข้อมูลภาพ (compression) เพื่อประหยัดพื้นที่ในการจัดเก็บข้อมูล หรือการฝังลายน้ำ (watermark) เพื่อป้องกันการลักลอบการใช้ภาพที่ไม่ได้รับอนุญาตเป็นต้น สำหรับการประมวลผลภาพระดับสูงด้วยคอมพิวเตอร์ สามารถทำได้โดย นำภาพที่ได้มาจากกล้องหรือ image source ต่าง ๆ ซึ่งเป็นสัญญาณอนาล็อก แล้วนำมาแปลงเป็นสัญญาณดิจิตอลที่มีลักษณะเป็น

รหัสเชิงตัวเลขฐานสอง (binary) ประกอบด้วยตัวเลข 0 และ 1 ที่สามารถใช้รูปแบบทางคณิตศาสตร์เข้ามาช่วยในการคำนวณและการประมวลผลข้อมูลภาพ การดำเนินงานวิจัยสำหรับการจัดกลุ่มความหมายของภาพได้มีผู้วิจัยจำนวนมากที่ศึกษาและทำการทดลองโดยทั่วไป แบ่งเป็น 2 ระดับ [A. Gupta, 1997] ดังนี้



ภาพที่ 2.8 การจัดเก็บและเรียกใช้ข้อมูลภาพดิจิทัลลงในเมทริกซ์

2.2.1 การประมวลภาพเบื้องต้น

สำหรับขั้นตอนการประมวลภาพเบื้องต้น (image preprocessing) ในรูปแบบของการหาบริเวณที่ต้องการในรูปแบบอัตโนมัติ ยังคงเป็นงานวิจัยที่ยังหาข้อยุติไม่ได้ โดยเฉพาะที่เป็นลักษณะของ ระบบเรียลไทม์ (real time) ด้วยแล้วนั้นจะต้องคำนึงถึงเวลาในการคำนวณของอัลกอริทึม (computational cost of algorithm) ที่เป็นสิ่งที่จำเป็นค่อนข้างมาก สำหรับระบบเรียลไทม์ ที่มีการแปรเปลี่ยนรูปแบบของพื้นหลัง (modeling of background) จะทำให้เกิดกระบวนการในการสกัดวัตถุที่ต้องการขึ้นมา ซึ่งเทคนิคที่เรียกว่า Gaussians Mixture Model [C. Stauffer, 2000] เป็นเทคนิคที่ค่อนข้างประสบความสำเร็จมากแต่อย่างไรก็ตามยังคงต้องมีส่วนภาพพื้นหลังที่ไม่มีการเปลี่ยนแปลงมากนัก เพราะฉะนั้นปัญหาที่เกิดขึ้นยังคงเกิดขึ้นเมื่อพื้นหลังมีการแปรเปลี่ยนและยังคงเป็นปัญหาที่ยังคงต้องมีการแก้ไข ฉะนั้นในงานวิจัยนี้จึงข้ามในส่วนการวิเคราะห์การแปรเปลี่ยนของ

พื้นหลังที่ไม่คงที่ และได้มีการนำเทคนิคของฮีสโตแกรม (Histogram matching) [F.Porikli, 2005] [D.Comaniciu, 2003] [D. Comaniciu, 2000] ที่ไม่ได้มีการนำส่วนของพื้นหลังเข้ามาเกี่ยวข้อง เพื่อทำการเปรียบเทียบแต่อย่างไรก็ตาม ด้วยคุณลักษณะของเทคนิคฮีสโตแกรมไม่สามารถที่จะรองรับตำแหน่งของพิกเซล (pixel location) ทำให้ยังคงมีปัญหาเมื่อมีวัตถุที่ทับซ้อนกัน จนไม่สามารถที่จะประมวลผลได้อย่างถูกต้อง เทคนิคสหสัมพันธ์ (Correlation) [A.J. Lipton, 1998] [S. Wong, 2005] จะมีการทำงานที่ไม่สูญเสียในข้อมูลส่วนของสเปเชียลทำให้สามารถทดแทนวิธีการของฮีสโตแกรมได้ และได้มีการใช้เทคนิคสหสัมพันธ์เพื่อทำการหาขอบของวัตถุซึ่งวิธีการนี้ถูกเรียกว่า Edge-Enhanced Normalized Correlation (EENC) [Javed Ahmed, 2008] เพื่อทำการแก้ไขปัญหาต่าง ๆที่เกิดจากการหาวัตถุที่เกิดจากการทับซ้อนการ หรือมีปัญหาจากสัญญาณรบกวน หรือวัตถุที่มีการเปลี่ยนแปลงทิศทางจากการหมุน เป็นต้น และการใช้เทคนิค EENC สามารถทำงานได้ดีสำหรับการแมชชิ่งส่วนของพื้นที่ (matching region) พร้อมทั้งยังเป็นแทมเพลต (template) ที่ใช้ทำงานได้อย่างรวดเร็ว

2.2.2 การประมวลผลภาพระดับสูง

การประมวลผลภาพระดับสูง (high-level image processing) เป็นการใช้ข้อมูลฟีเจอร์หรือผลลัพธ์จากการประมวลผลข้างต้นเพื่อผ่านกระบวนการ หรืออัลกอริทึมทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพ (image understanding) รวมทั้งการหาความหมายของภาพ (semantic image) ได้ เพราะฉะนั้นการประมวลผลภาพระดับสูงจำเป็นต้องใช้ข้อมูลที่ได้มาจากการประมวลผลภาพระดับต่ำ ดังนั้นจะเห็นได้ว่าการประมวลผลภาพระดับต่ำมีความสำคัญมากสำหรับการทำให้คอมพิวเตอร์รู้จักและเข้าใจภาพได้ โดยส่วนใหญ่กลุ่มนักวิจัยพยายามที่จะข้ามขั้นตอนการประมวลผลภาพระดับต่ำไป เนื่องจากยังไม่มีความสมบูรณ์เท่าที่ควรและพยายามที่เข้าใช้การประมวลผลภาพระดับสูงอย่างเดียว แต่อย่างไรก็ตามการวิจัยทั้งสองกลุ่มนี้ยังคงมีการพัฒนาการวิจัยอย่างต่อเนื่องเพื่อนำผลลัพธ์ หรือฟีเจอร์ต่าง ๆเข้ามาทำการค้นคืนภาพ (image retrieval) หรือ การจำแนกข้อมูลภาพ (image classification) รวมทั้งการแยกแยะความหมายของภาพ (semantic image classification) ซึ่งเป็นหัวข้อหลักในการทำวิจัยครั้งนี้

2.3 การเปรียบเทียบความคล้ายของภาพ

การเปรียบเทียบความคล้ายของภาพ (image similarity measure) เป็นการนำเอาข้อมูลทั้งหมดที่เก็บรวบรวมมาได้ มาผ่านกระบวนการขั้นตอนการแยกแยะข้อมูลลงในแต่ละกลุ่มที่จัดไว้ โดยในแต่ละกลุ่มของข้อมูลนั้นจะมีคุณลักษณะเด่นของแต่ละกลุ่มที่แตกต่างกัน ขึ้นกับข้อมูล หรือฟีเจอร์ที่เก็บรวบรวมมาได้รวมทั้งกระบวนการหรือวิธีการที่ใช้จำแนกข้อมูล ซึ่งในงานวิจัยนี้ได้ใช้วิธีการแยกประเภทข้อมูลของภาพเพื่อเปรียบเทียบความสามารถ ทั้งหมด 3 วิธี

- โครงข่ายประสาทเทียมแบบหลายชั้น (Multiple Feedforward Neural Network),
- ท์ฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-organizing maps)
- ทฤษฎีเครือข่ายแบบเบย์ (Bayesian Network)

2.3.1 โครงข่ายประสาทเทียม

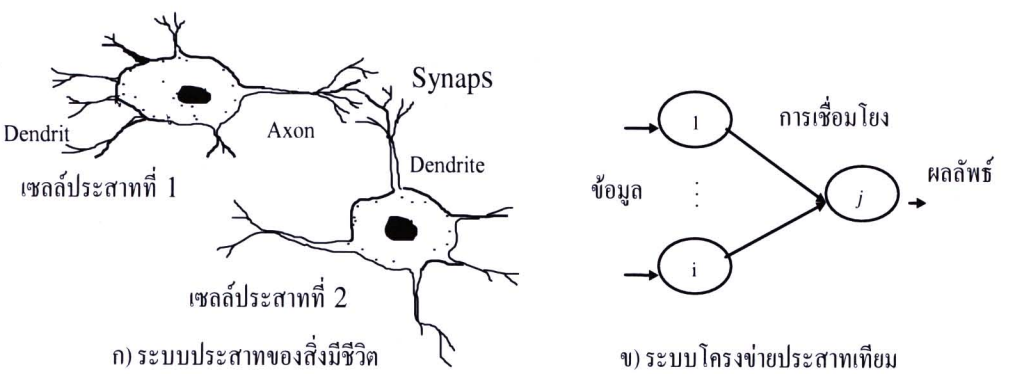
โครงข่ายประสาทเทียม (artificial neural network) [Foold, I.,1994] [Lippmann, R.P., 1987] หรือเรียกว่า ข่ายงานประสาท (neural network หรือ neural net) เป็นโมเดลทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรูปการจดจำแบบรูป (pattern recognition) และการอุปมานความรูป (knowledge deduction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ ข่ายงานประสาทเทียมถูกพัฒนามาจากวัตถุประสงค์ดังนี้

- การเลียนแบบการทำงานของเซลล์สมองของมนุษย์
- การหาความสัมพันธ์ของเหตุและผลซึ่งยังหาฟังก์ชั่นออกมาโดยตรงไม่ได้
- การทำให้คอมพิวเตอร์คิดและตัดสินใจได้ด้วยตัวเอง

แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาข่ายงานไฟฟ้าชีวภาพ (bioelectric network) ในสมอง ประกอบด้วย เซลล์ประสาท หรือ “นิวตรอน” (neurons) และ จุดประสานประสาท (synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า “เดนไดรท์” (dendrite) เป็นข้อมูลเข้า (input) และปลายในการส่งกระแสประสาทเรียกว่า “แอคซอน” (axon)

เป็น ผลลัพธ์ (output) ของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรท์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินว่าต้องกระตุ้นเซลล์อื่น ๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่น ๆ ต่อไปผ่านทางแอคซอน ดังแสดงในภาพที่ 2.9

จากรูป เซลล์ร่างกาย (cell body) คือ ตัวของเซลล์ประสาท โดยสัญญาณไฟฟ้าจะเข้ามาทางเดนไดรท์ซึ่งจะมีลักษณะเป็นแขนงย่อยและมีจำนวนมาก จากนั้น เซลล์ร่างกายจะทำหน้าที่ประมวลผลสัญญาณ และยิงสัญญาณออกไปทางเดนไดรท์ให้ เซลล์ประสาทอื่นต่อไป



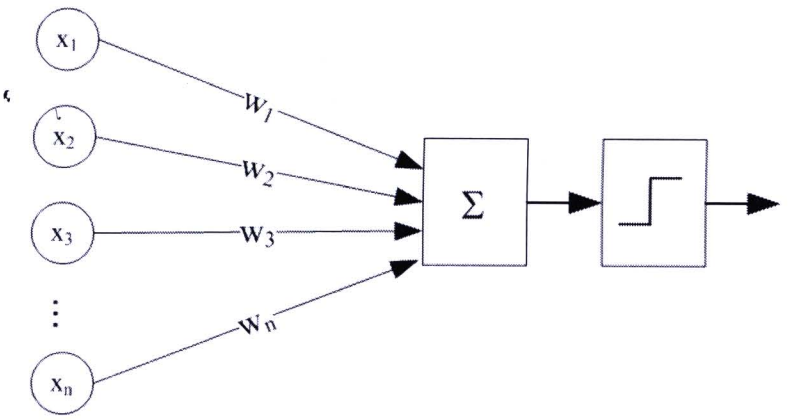
ภาพที่ 2.9 เซลล์ระบบประสาทของสิ่งมีชีวิตและโครงข่ายประสาทเทียม

▪ หลักการทำงานข่ายงานประสาท

สำหรับในคอมพิวเตอร์นิวรอน [Chester M, 1993][Smith M., 1993] ประกอบด้วย ข้อมูลเข้า และ ข้อมูลออกเหมือนกัน โดยจำลองให้ ข้อมูลเข้า แต่ละอันมีค่าน้ำหนัก (weight) เป็นตัวกำหนดน้ำหนักของ ข้อมูลเข้า โดย นิวรอนแต่ละหน่วยจะมีค่า threshold เป็นตัวกำหนดว่าน้ำหนักรวมของข้อมูลเข้าต้องมากขนาดไหนจึงจะสามารถส่งข้อมูลออกไปยัง นิวรอนตัวอื่นได้ เมื่อนำ นิวรอนแต่ละหน่วยมาต่อกันให้ทำงานร่วมกันการทำงานนี้ในทางตรรกจะเหมือนกับปฏิกิริยาเคมีที่เกิดในสมอง เพียงแต่ในคอมพิวเตอร์ทุกอย่างเป็นตัวเลข เพราะฉะนั้นเมื่อนำการทำงานเข้ามาใช้กับคอมพิวเตอร์สามารถเขียนได้ดังนี้

if (sum(input * weight) > threshold) then output

เมื่อมี input แทนข้อมูลเข้า เข้ามายังโครงข่าย จะใช้ข้อมูลเข้า คูณกับ weight แทนน้ำหนักของนิวรอนแต่ละขา ผลที่ได้จากข้อมูลเข้าทุกขาของนิวรอน จะรวมกันแล้วก็เอามาเทียบกับ threshold ที่กำหนดไว้ ถ้าผลรวมมีค่ามากกว่า threshold แล้วนิวรอนจะส่งข้อมูลออกออกไปข้อมูลออกนี้ก็จะถูกส่งไปยังข้อมูลเข้า ของนิวรอนอื่น ๆ ที่เชื่อมกันในโครงข่ายถ้าค่าน้อยกว่า threshold ก็จะไม่เกิดข้อมูลออกดังแสดงในภาพที่ 2.10



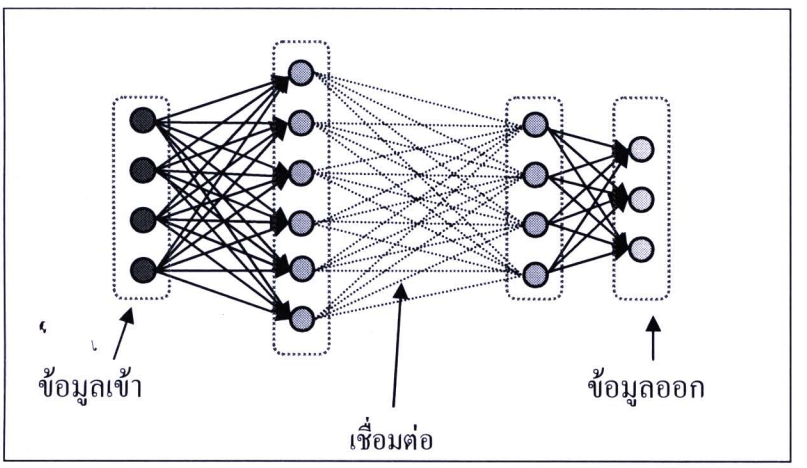
ภาพที่ 2.10 ตัวแบบของนิวรอนในคอมพิวเตอร์

สถาปัตยกรรมโครงข่ายประสาทเทียมที่พบทั่วไปจะมีลักษณะหลัก ๆ คือ มีการจัดเซลล์ประสาทเทียมเป็นชั้น (layer) ชั้นที่รับข้อมูลเข้าเรียกว่า ชั้นข้อมูลเข้า (input layer) ชั้นที่ผลิตผลตอบของโครงข่ายเรียกว่า ชั้นข้อมูลออก (output layer) ส่วนชั้นอื่น ๆ ที่มีส่วนในการช่วยทำการประมวลผลอยู่ภายในเรียกว่า ชั้นซ่อน (hidden layer) ในโครงข่ายประสาทเทียมอาจมีชั้นซ่อนได้หลายชั้นโครงสร้างพื้นฐานจะมีลักษณะเป็นการประกอบกันของรูปแบบ ดังต่อไปนี้

1. แบบป้อนไปข้างหน้า

แบบป้อนไปข้างหน้า (feedforward network) ข้อมูลที่ประมวลผลในวงจรข่ายจะถูกส่งไปในทิศทางเดียวจาก โหนดข้อมูลเข้า ส่งต่อมาเรื่อย ๆ จนถึง โหนดข้อมูลออก โดยไม่มีการย้อนกลับของข้อมูล หรือแม้แต่ โหนดในชั้นเดียวกันก็ไม่มีการเชื่อมต่อกัน สามารถจัดได้เป็น 2 แบบคือ แบบมีชั้นของเซลล์ประสาทชั้นเดียว และแบบมีชั้นของเซลล์ประสาทหลายชั้น โดยปกติการเชื่อมโยงจะถูกกำหนดขึ้นระหว่างชั้นที่ติดกัน โดยจะมีการเชื่อมโยงระหว่างเซลล์ประสาทเทียมทุกตัว จากชั้นหนึ่ง ๆ

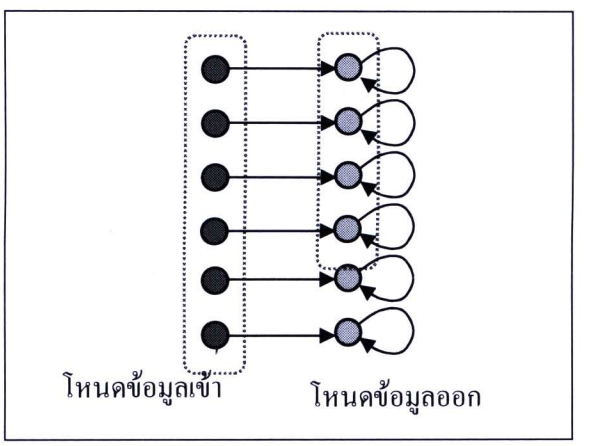
ไปยังเซลล์ประสาทเทียมทุกตัวในชั้นต่อไป ในบางสถาปัตยกรรมอาจมีการเชื่อมโยงข้ามชั้นก็ได้ ดังแสดงในภาพที่ 2.11



ภาพที่ 2.11 สถาปัตยกรรมแบบป้อนไปข้างหน้า

2. แบบมีการป้อนไปเวียนกลับ

แบบมีการป้อนไปเวียนกลับ (feedback network) บางทีเรียกว่า recurrent network ข้อมูลที่ประมวลผลโครงข่ายประสาทเทียมจะมีการป้อนกลับเข้าไปยังวงจรข่ายหลาย ๆ ครั้ง จนกระทั่งได้คำตอบออกมา จะมีการเชื่อมโยงที่ถูกกำหนดขึ้นระหว่างเซลล์ประสาทเทียมในชั้นหนึ่ง ๆ ย้อนกลับไปยังชั้นอื่น ๆ ก่อนหน้านั้น หรือแม้แต่ภายในชั้นเดียวกันเอง ดังแสดงในภาพที่ 2.12

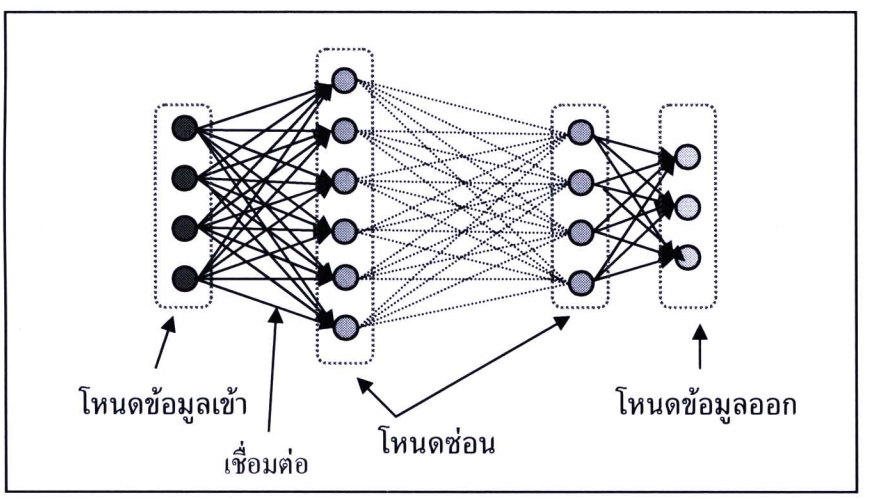


ภาพ 2.12 สถาปัตยกรรมแบบมีการป้อนไปเวียนกลับ

สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย

3. แบบป้อนไปข้างหน้าหลายชั้น

โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าหลายชั้น (multiple feedforward) ใช้การเชื่อมโยงแบบป้อนไปข้างหน้าประกอบด้วย 3 ชั้น ได้แก่ ชั้นโหนดข้อมูลเข้า ที่ถูกเชื่อมต่อกับชั้นของโหนดซ่อน และจะมีการเชื่อมต่อกับชั้นของโหนดข้อมูลเข้า เป็นการเพิ่มจำนวนชั้นใน ชั้นซ่อนดังแสดงในภาพที่ 2.11 การทำงานของโหนดข้อมูลเข้าจะทำหน้าที่แทนส่วนของข้อมูลดิบ ที่ถูกป้อนเข้าสู่เครือข่าย และการทำงานของแต่ละ โหนดซ่อน จะถูกกำหนด โดยการทำงานของโหนดข้อมูลเข้าและค่าน้ำหนักบนความสัมพันธ์ระหว่าง โหนดข้อมูลเข้า และ โหนดซ่อน พฤติกรรมการทำงานของ โหนดข้อมูลออก จะขึ้นอยู่กับการทำงานของ โหนดซ่อน และค่าน้ำหนักระหว่าง โหนดซ่อน และโหนดข้อมูลออก ดังแสดงในภาพที่ 2.13



ภาพ 2.13 โครงสร้างแบบป้อนไปข้างหน้าหลายชั้น

ลักษณะการทำงานแบบนี้สามารถกำหนดการแทนค่าให้แก่ โหนดข้อมูลเข้าได้อย่างอิสระ ค่าน้ำหนักระหว่าง โหนดข้อมูลเข้าและ โหนดซ่อน จะถูกกำหนดเมื่อ โหนดซ่อน กำลังทำงาน ฉะนั้นเวลาที่แก้ไขค่าน้ำหนัก โหนดซ่อน จะสามารถเลือกว่าอะไรคือค่าที่เราแทนเข้ามาในโมเดลทางคณิตศาสตร์ของโครงข่ายประสาทเทียม แบบป้อนไปข้างหน้าหลายชั้น แสดงในภาพ มีเวกเตอร์อินพุต $X = \{x_i\}$ กำหนดให้ $x_i \in \Re^p$ และเวกเตอร์ผลลัพธ์ $Y = \{y_i\}$ กำหนดให้ $y_i \in \Re^m$ ซึ่งประกอบด้วยสมการการคำนวณภายในเซลล์ประสาทแต่ละชั้นดังนี้

ชั้นซ่อน

$$h_i = \sigma(u_j)$$

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

กำหนดให้ h_j คือผลลัพธ์ของเซลล์ใน ชั้นซ่อน ซึ่ง $1 \le j \le L$

$\sigma(\cdot)$ คือ activation function

w_{ji}^{h} คือ ค่าของน้ำหนักบนเส้น ji เชื่อมเซลล์ในชั้นซ่อน j กับเซลล์ในชั้นข้อมูลเข้า i

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

ชั้นข้อมูลออก

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

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

โดยที่ y_k คือ ผลลัพธ์ของเซลล์ใน ชั้นข้อมูลออก ซึ่ง $1 \le k \le m$

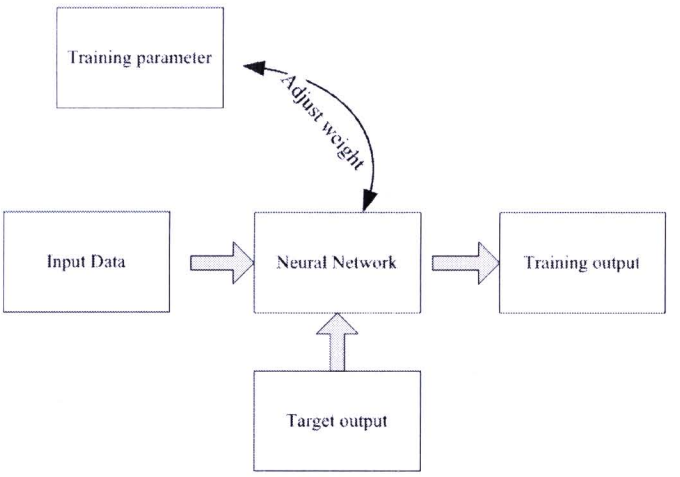
$\sigma(\cdot)$ คือ activation function

w_{ki}^{h} คือ ค่าของน้ำหนักบนเส้น ki เชื่อมเซลล์ในชั้นซ่อน k กับเซลล์ในชั้นข้อมูลเข้า j

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

- รูปแบบการเรียนรู้ของโครงข่ายประสาทเทียม

การเรียนรู้เป็นกระบวนการที่สามารถทำให้เปลี่ยนแปลงพฤติกรรม ความคิด สามารถเรียนได้จากการได้ยินการสัมผัส การอ่าน การใช้เทคโนโลยี สามารถแบ่งออกได้ 2 แบบ



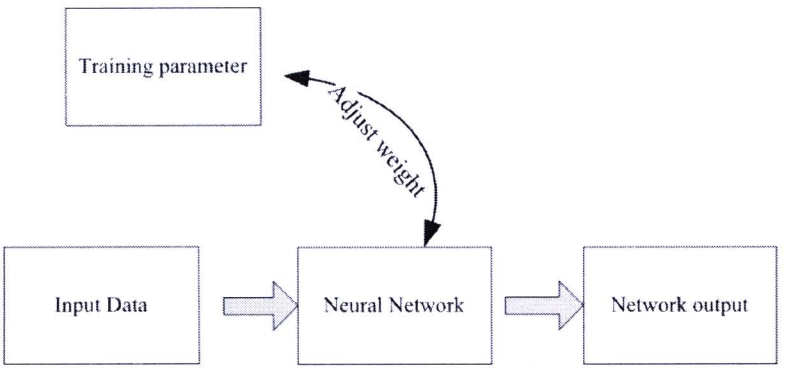
ภาพที่ 2.14 การเรียนรู้แบบมีการสอน

1. การเรียนรู้แบบมีการสอน

การเรียนรู้แบบมีการสอน (supervised learning) เป็นการเรียนรู้แบบที่มีการตรวจคำตอบเพื่อให้วงจรข่ายปรับตัว ชุดข้อมูลที่ใช้สอนวงจรข่ายจะมีคำตอบไว้คอยตรวจดูว่าวงจรข่ายให้คำตอบที่ถูกหรือไม่ ถ้าตอบไม่ถูก การเรียนรู้โดยให้ข่ายงานปรับตัวเข้ากับค่าจริง (target) หรือคำตอบที่มีอยู่เพื่อให้ได้คำตอบที่เหมือนหรือใกล้เคียงกับคำตอบจริงดังแสดงในภาพที่ 2.14

2. การเรียนรู้แบบไม่มีผู้สอน

การเรียนรู้แบบไม่มีการสอน (unsupervised learning) เป็นการเรียนแบบไม่มีผู้แนะนำ ไม่มีการตรวจคำตอบว่าถูกหรือผิด จะไม่มีค่าจริงในการฝึกสอน เครือข่ายจะจัดเรียงโครงสร้างของตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้เครือข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ เช่น การแยกลักษณะของคน พืช และสัตว์ เองได้โดยไม่มีผู้สอนจริงในการฝึกสอน เครือข่ายจะจัดเรียงโครงสร้างของตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้เครือข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ เช่นการที่เราสามารถแยกลักษณะของคน พืช และสัตว์ เองได้โดยไม่มีการสอนดังแสดงในภาพที่ 2.15



ภาพที่ 2.15 การเรียนรู้แบบไม่มีการสอน

- การเรียนรู้แบบแพร่ย้อนกลับ

การเรียนรู้แบบแพร่ย้อนกลับ (back-propagation learning) ถูกนำไปประยุกต์ใช้มากที่สุดในการใช้งานในโครงข่ายประสาทเทียม โดยลักษณะของการเรียนรู้แล้ว การเรียนรู้แบบแพร่ย้อนกลับไม่จำเป็นต้องใช้เฉพาะโครงข่ายประสาทเทียมที่มีสถาปัตยกรรมแบบป้อนไปข้างหน้าหลายชั้นเพียงแบบเดียว แต่การจับคู่ดังกล่าวได้รับความนิยมนำไปใช้งานมากที่สุด การเรียนรู้แบบแพร่ย้อนกลับนี้เป็นการเรียนรู้แบบมีการสอน (supervise learning) ขั้นตอนในการเรียนรู้แบบแพร่ย้อนกลับจะเริ่มด้วยการขนส่งข้อมูลที่จะใช้สอนเข้าไปในเครือข่ายทางชั้นข้อมูลเข้า ชั้นข้อมูลเข้าซึ่งโดยปกติจะไม่มีการประมวลผล จะทำหน้าที่ส่งกระจายข้อมูลนั้นไปยังเซลล์ประสาทเทียมต่าง ๆในชั้นซ่อน จากนั้นเซลล์ประสาทเทียมทุกตัวในชั้นซ่อน จะทำการประมวลผลตามวิธีการของแบบจำลองเซลล์ประสาทเทียม การประมวลผลเกิดจากชั้นซ่อนชั้นแรกผ่านไปตามชั้นซ่อนต่าง ๆจนสร้างผลตอบที่ชั้นข้อมูลออก

ผลที่ตอบออกมาจากชั้นเอาต์พุตจะถูกเปรียบเทียบกับค่าเป้าหมายที่กำหนด เพื่อคำนวณค่าความคลาดเคลื่อน (error) จากนั้นปริมาณการปรับแต่งค่าน้ำหนักของแต่ละขั้วสัญญาณเข้าของแต่ละเซลล์ประสาทเทียมในชั้นข้อมูลออกก็จะถูกคำนวณขึ้น โดยดูจากความคลาดเคลื่อนและปริมาณการปรับค่าน้ำหนักที่ขั้วสัญญาณของแต่ละเซลล์ประสาทเทียมในชั้นซ่อนต่าง ๆ แต่เนื่องจากไม่สามารถกำหนดค่าเป้าหมายของผลตอบจากเซลล์ประสาทเทียมในชั้นซ่อนจึงถูกมองว่า เป็นการไหลย้อนกลับของความคลาดเคลื่อนของเซลล์ประสาทเทียมต่าง ๆ ของชั้นข้อมูลออกหลังจากรวบรวมค่าความคลาดเคลื่อนที่ไหลย้อนกลับมาแล้ว หลักการเดียวกับชั้นข้อมูลออก คือ ดูจากผลรวมความคลาดเคลื่อนที่ไหลย้อนกลับมายังเซลล์ประสาทเทียมตัวนั้น ๆ ประกอบกับค่าของสัญญาณเข้าที่วิ่งผ่านเข้ามาทางขั้วสัญญาณเข้านั้นด้วย

เมื่อค่าน้ำหนักที่ขั้วสัญญาณเข้าของเซลล์ประสาทเทียมทุกตัวถูกปรับแต่งหมดแล้วโครงข่ายประสาทเทียมก็พร้อมจะเรียนรู้ข้อมูลตัวอย่างต่อไป ในทางปฏิบัติแล้ว จะต้องใช้ข้อมูลตัวอย่างจำนวนมาก และจะต้องสอนข้อมูลตัวอย่างเหล่านั้นซ้ำกันหลายรอบ จึงจะสามารถสอนโครงข่ายประสาทเทียมชนิดนี้จนเรียนรู้ได้อย่างถูกต้อง

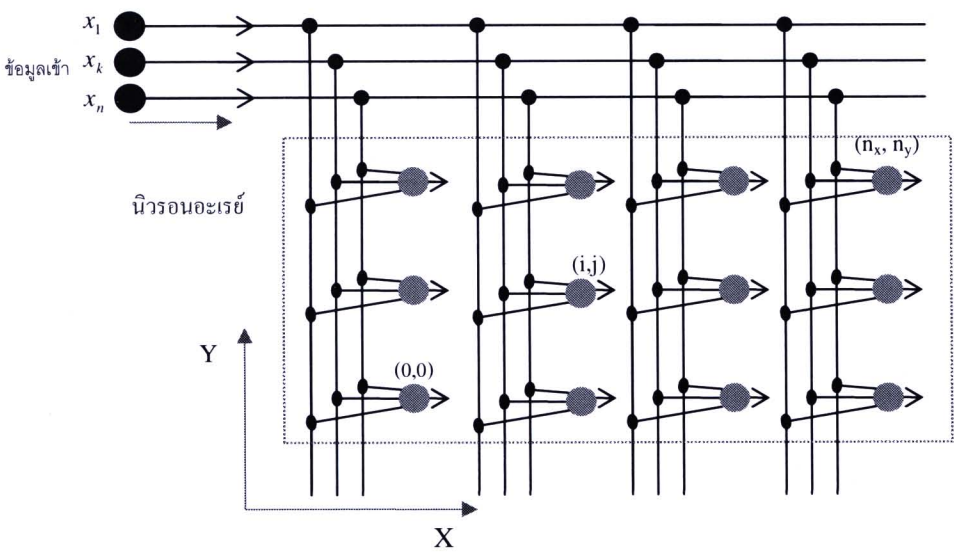
- ข้อดีโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม เป็นแขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence: AI) ซึ่งเลียนแบบการทำงานและคุณสมบัติเซลล์สมองหรือระบบประสาทของมนุษย์ เมื่อโครงข่ายประสาทเทียมผนวกกับความสามารถของวิทยาการคอมพิวเตอร์ในปัจจุบัน เช่น หน่วยความจำ การประมวลผลที่รวดเร็ว แม่นยำ และค่าใช้จ่ายที่ไม่สูงนัก ทำให้ได้ระบบที่มีศักยภาพในการทำงาน มีคุณลักษณะและคุณสมบัติที่น่าสนใจ เช่น สามารถจำลองปัญหาได้โดยไม่จำเป็นต้องทราบรูปแบบการกระจายของข้อมูล (distribution free) มีข้อผิดพลาดได้บ้าง (fault tolerance) เรียนรู้ด้วยตนเองได้ (self-organization) มีการทำงานแบบขนาน (parallel process) รวดเร็วระบบทำงานโดยใช้เพียงฟังก์ชันทางคณิตศาสตร์อย่างง่ายแทนที่จะใช้กลไกทางชีวเคมี และไม่ได้ทำงานตามชุดคำสั่งแต่อย่างเดียวดังเช่นโปรแกรมคอมพิวเตอร์ทั่วไป คำตอบหรือผลลัพธ์น่าเชื่อถือ ด้วยเหตุผลดังกล่าวโครงข่ายประสาทเทียมจึงสามารถแก้ปัญหาใกล้เคียงกับเซลล์สมองหรือระบบประสาทของสิ่งมีชีวิตโดยเฉพาะมนุษย์ ระบบเรียนรู้หรือรู้จำจากตัวอย่างที่มีจำนวนและความหลากหลาย แหล่งที่มาของตัวอย่างอาจได้จากการสัมภาษณ์ตรง (direct interview) ข้อมูลในอดีต (historical record) หรือกระบวนการการจำลอง (simulation)

2.3.2 ทฤษฎีแผนผังการจัดระเบียบตัวเอง

ทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-Organizing Maps : SOM) [T. Kohonen, 1995] [T. Mitchell,1997] เป็นอัลกอริทึมนิวรอนเน็ตเวิร์คที่นิยมใช้มากที่สุด โดยแนวคิดของ SOM คือ การทำซ้ำข้อมูลเพื่อหาค่าของน้ำหนักของข้อมูลที่มีอยู่ทั้งหมดตามจำนวนกลุ่มที่ต้องการ เป็นวิธีเกี่ยวกับการจัดกลุ่มด้วยตัวเองโดยใช้โครงสร้างตาข่ายระบบประสาท เป็นการเรียนรู้แบบไม่มีผู้สอน มีลักษณะการเรียนรู้ แบบเครือข่าย 2 ชั้น เป็นการจัดข้อมูลนำเข้าหลายมิติ ให้อยู่ในรูปแบบของข้อมูลสองมิติคือการจัดกลุ่มของข้อมูลที่มีลักษณะคล้ายกันจะอยู่ในโหนดใกล้ที่เคียงกัน โดยโครงสร้างการทำงานของ SOM ชั้นแรกทำหน้าที่นำเข้าข้อมูลและจัดส่งข้อมูลให้แก่นิวรอนชั้นที่สอง

ทุกโหนดในตาราง หรือ อะเรย์นิวรอน (array of neurons) ระหว่างชั้นจะเชื่อมต่อกันด้วยค่าน้ำหนัก (weight) จากนั้นข้อมูลจะถูกส่งไปยังนิวรอนในชั้นที่สอง เพื่อทำการเปรียบเทียบว่าใกล้เคียงกับค่ากลางกลุ่มใดมากที่สุด แต่ละโหนดในชั้นนี้จะมีความสัมพันธ์กันแบบ เพื่อนบ้าน (neighborhood relation) ทำให้เกิดเป็นรูปแบบ 2 มิติ จากนั้นจะทำการปรับค่า น้ำหนักของตัวที่เป็นผู้ชนะ (winner) ทำให้ข้อมูลที่ข้อมูลเข้า เข้ามาเกิดการปรับเปลี่ยน และเมื่อผ่านการเรียนรู้ไปหลายๆรอบจะทำให้ ได้กลุ่มข้อมูลออกมาเป็นผลลัพธ์ จากนั้นจึงนำผลลัพธ์ที่ได้นั้นมาผ่านกระบวนการ Visualization เพื่อแสดงผลลัพธ์ที่ได้นั้นออกมาเป็นกราฟชนิดต่างๆต่อไป



ภาพที่ 2.16 โหนดการเรียนรู้ของแผนผังการจัดระเบียบตัวเอง

ขั้นตอนการทำงานของทฤษฎีแผนผังการจัดระเบียบตัวเอง SOM

1. ทำการจัดกลุ่มของข้อมูลนำเข้าให้อยู่ในกลุ่มข้อมูลที่ใกล้เคียง ทำการ Normalize ค่าข้อมูลแต่ละตัว เพื่อให้ได้ค่ากลางเป็น $x_i = \frac{x_i - \min}{\max - \min}$, กำหนดให้ $x_1, x_2, ..., x_n$ เป็นข้อมูลที่ถูกนำเข้า n คือจำนวนตัวอย่างข้อมูลเข้า min คือ ค่าที่น้อยที่สุดของชุดข้อมูล max คือ ค่าที่มากที่สุดของชุดข้อมูล

2. กำหนดโครงสร้างโครงข่ายประสาทเทียมในลักษณะสองมิติ กำหนด แนวแกน x และแกน y กำหนดชุดของข้อมูล สุ่มค่าน้ำหนักเริ่มต้น ในโครงข่ายดังแสดงในภาพที่ 2.16
3. สุ่มค่าเริ่มต้นให้กับค่ากลางของกลุ่ม (cluster center) $W = \{w_1, w_2, ..., w_p\}$, โดยที่ p คือ จำนวนกลุ่ม เวกเตอร์ค่ากลางคือ $W_j = \{w_{j1}, w_{j2}, ..., w_{jm}\}$, เมื่อ $1 \le j \le p$
4. คำนวณหาระยะทางระหว่างโครงข่ายโดย ยูคลิเดียน (Euclidean distance) เพื่อหาผู้ชนะ (winner) ซึ่งจะหาได้จากโครงข่ายที่ใช้ระยะทางที่ได้จากการคำนวณ เพื่อหาค่าน้อยที่สุด

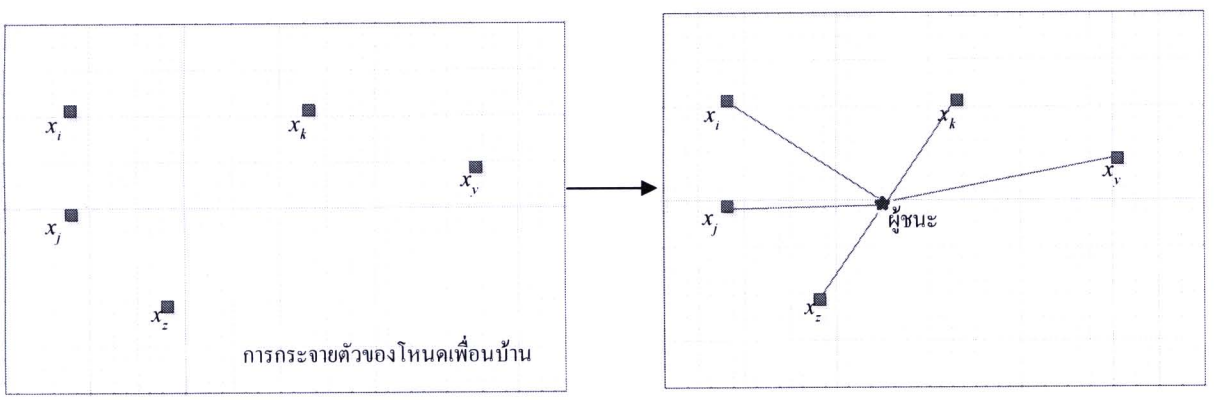
$$C(k_1, k_2) = \min_{i,j} C_{i,j}$$

เมื่อ k_1, k_2 คือดัชนีของโหนดผู้ชนะ

$$C_{i,j}^2 = \left\| x - w_j \right\|^2$$

$$C_{i,j}^2 = \sum_{i=1}^{w} (x_i - w_{ij})^2$$

เมื่อ $C_{i,j}$ ค่าความต่างระหว่างข้อมูลนำเข้า x_i กับเวกเตอร์น้ำหนัก $w_{i,j}$



ภาพที่ 2.17 การหาผู้ชนะสำหรับการกระจายตัวของข้อมูล x_i หรือโหนดเพื่อนบ้าน

5. คำนวณค่ากลางสำหรับกลุ่มที่เป็น ค่าผู้ชนะใหม่

$$w_j(t+1) = w(t) + \eta(x_i - w_j(t))$$

6. ทำซ้ำขั้นตอนที่ 4-5 จนครบชุดข้อมูลนำเข้า หรือจนกว่าค่าน้ำหนักเริ่มจะคงที่
7. โหนดที่เป็นผู้ชนะจะสอดคล้องตามสูตร $C(k_1, k_2) = \min_{i,j} C_{i,j}$ เมื่อ k_1, k_2 ดัชนีของโหนดที่เป็นผู้ชนะ

สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย

8. โหนดที่เป็นโหนดเพื่อนบ้านจะถูกกำหนดโดย $h(\rho,t)=\exp(\frac{\rho^2}{2\sigma^2(t)})$

9. หาระยะห่างระหว่างโหนดนั้น ๆ กับ โหนดที่เป็นผู้ชนะ $\rho=\sqrt{(k_1-i)^2+(k_2-j)^2}$

10. หาโหนดเพื่อนบ้านที่ใกล้เคียงที่สุด โดยขอบเขตจะลดลงตามเวลาตามรูปแบบดังนี้

$$h(\rho,t)=\exp\left(-\frac{\rho^2}{\sigma^2(t)}\right)\left(1-\frac{2}{\sigma^2(t)}\rho^2\right)$$

$$h(\rho)=\begin{cases}1, & |\rho|\le a,\\ -\frac{1}{3}, & a<|\rho|\le 3a,\\ 0, & |\rho|<3a,\end{cases}$$

11. ปรับค่าน้ำหนัก ของแต่ละโหนดด้วย

$$W_{i,j}(t+1)=W_{i,j}(t)+a(t)h(\rho,t)(X^l(t)-W(t))$$

- ข้อเสียของทฤษฎีแผนผังการจัดระเบียบตัวเอง SOM

การเรียนรู้นี้ไม่ทราบค่าที่ต้องการจึงไม่สามารถจะหยุดการเรียนรู้เมื่อไร ต้องมีการปรับค่าน้ำหนักไปเรื่อย ๆ ปัญหานี้สามารถแก้ไขได้ โดยอาศัยเงื่อนไขหรือหลักเกณฑ์บางอย่างในการใช้เป็นเครื่องมือในการหยุดการเรียนรู้

ค่ากลางบางตัวซึ่งเกิดจากการสุ่ม อาจจะตกในบริเวณที่ห่างจากกลุ่มค่ากลางตัวอื่นมาก ซึ่งไม่สามารถเป็น ตัวชนะได้เลย จึงต้องเพิ่มกระบวนการในการแก้ปัญหานี้

ไม่สามารถ กำหนดจำนวนชุดข้อมูลที่นำมาจัดกลุ่ม ควรมีกี่กลุ่มถึงจะเหมาะสม จึงควรมีการศึกษาวิจัยต่อไป

2.3.3 ทฤษฎีความเชื่อเบย์

ทฤษฎีความเชื่อเบย์ (ทฤษฎีความเชื่อเบย์ (bayesian approach) [R. O. Duda, 1973] [T. Mitchell, 1997] เป็นการใช้รูปแบบของการประมาณค่าของความน่าจะเป็นที่จะเกิดเหตุการณ์ โดยอาศัยความสัมพันธ์ที่เกิดขึ้นของข้อมูลเพื่อมาคำนวณเหตุการณ์ที่เกิดขึ้น ดังนั้นทฤษฎีความเชื่อเบย์จึงอยู่บนพื้นฐานของการประมาณค่าความน่าจะเป็นที่จะเกิดขึ้น ในเบื้องต้นจะกล่าวถึงทฤษฎีพื้นฐานของความน่าจะเป็น (basic formulas for probability) ดังนี้ ถ้ากำหนดให้ A และ B เป็นเหตุการณ์ที่เป็นอิสระต่อกัน คือเหตุการณ์ A และ B ไม่มีความเกี่ยวข้อกัน 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

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

- วิธีการเรียนรู้เบย์อย่างง่าย

วิธีการเรียนรู้เบย์อย่างง่าย (naïve-bayes learning) [T. Mitchell, 1997] วิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็น ซึ่งมีพื้นฐานมาจากทฤษฎีของเบย์ (bayes theorem) เข้ามาช่วยในการเรียนรู้ จุดมุ่งหมายก็เพื่อต้องการสร้างโมเดลที่อยู่ในรูปของความน่าจะเป็น ซึ่งเป็นค่าที่บันทึกได้จากการสังเกต จากนั้นนำโมเดลมาหาว่าสมมติฐานใดถูกต้องที่สุดโดยใช้ความน่าจะเป็นเข้ามาช่วยความรู้ก่อนหน้า หมายถึง ความรู้ที่เรามีเกี่ยวกับสมมติฐานแต่ละตัวก่อนที่เราจะเก็บข้อมูล เมื่อใช้งานเราจะนำความน่าจะเป็นของข้อมูลที่เก็บได้มาปรับสมมติฐานซ้ำอีกครั้ง เมื่อ C แทนกลุ่มข้อมูลที่จะถูกแบ่งประกอบด้วยกลุ่มที่ถูกสังเกต c $X\langle X_1, X_2, .., X_k \rangle$ แทนแอทริบิวต์เวกเตอร์ข้อมูล ประกอบด้วยข้อมูลที่ถูกสังเกต $x\langle x_1, x_2, .., x_k \rangle$ สามารถเขียน $X = x$ โดยย่อยได้ดังนี้

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

สำหรับค่าผิดพลาดของการทำนายเป็น

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

สามารถเขียนสมการเบย์ได้ดังนี้

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

สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย

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

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

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

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

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) เข้ามาช่วยในการเรียนรู้ได้ซึ่งพบว่าวิธีนี้ให้ประสิทธิภาพในการเรียนรู้ได้ดีไม่ด้อยกว่าวิธีการเรียนรู้ประเภทอื่น สามารถลดข้อจำกัดอย่างง่ายในสมมุติฐานของความไม่ขึ้นต่อกันระหว่างคุณสมบัติ แต่ในความเป็นจริงคุณสมบัติบางตัวจะขึ้นต่อกัน และควรนำค่าความขึ้นต่อกันนี้เข้ามาใส่ไว้ในโมเดล จึงใช้ข่ายงานเบย์ในการอธิบายความไม่ขึ้นต่อกันอย่างมีเงื่อนไข (condition independent) ระหว่างตัวแปร เพื่อทำให้กระบวนการเรียนรู้มีประสิทธิภาพ โดยสามารถใส่ความรู้ก่อนในข่ายงานความเชื่อเบย์ให้อยู่ในรูปโครงสร้างข่ายงานและตารางความน่าจะเป็นแบบมีเงื่อนไข

2.4 การใช้โมเดลในการสร้างการตัดสินใจ

การใช้โมเดลในการสร้างการตัดสินใจ เป็นวิธีการที่จะทำให้ความผิดพลาดเฉลี่ยในการจำแนกเกิดขึ้นน้อยที่สุด ขึ้นอยู่กับความสามารถในการแยกข้อมูลประเภทให้ถูกต้อง ซึ่งขึ้นอยู่กับความสามารถในการแยกความทับซ้อนกันของค่าการสะท้อนช่วงคลื่นระหว่างประเภท เพราะฉะนั้นการจำแนกจะรวมหลักการสำคัญหลายอย่างที่ต้องทำความเข้าใจในข้อกำหนด และปรัชญาของ

หลักการเหล่านั้น เพื่อทำให้การจำแนกได้ผลที่เชื่อถือได้และตรงกันทุกครั้ง (reliable and consistency)

การแบ่งข้อมูลให้เป็นแต่ละประเภท โดยทั่วไปจะใช้ขั้นตอนปฏิบัติ หรืออัลกอริทึมเพื่อทำการจำแนกข้อมูล เป็นเงื่อนไขที่แน่นอนที่กำหนดกระบวนการทางคอมพิวเตอร์ที่ใช้ข้อมูลเบื้องต้นที่หลากหลายไปสู่ผลลัพธ์ที่ต้องการ การใช้ฟีเจอร์หรือ ฟีเจอร์ที่ได้มาจากการสกัดข้อมูลภาพเป็นอีกสิ่งหนึ่งที่ผลอย่างมากต่อผลที่นำมาใช้ในการตัดสินใจ ซึ่งในการทำให้ค่าที่ได้เปลี่ยนแปลงจะทำให้การตัดสินใจเปลี่ยนแปลงไปและจะมีผลต่อความแม่นยำในการจำแนกข้อมูล

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

- เทคนิคการจำแนกแบบไม่มีการสอน

เทคนิคการจำแนกแบบไม่มีการสอน (unsupervised classification) การจำแนกแบบไม่มีการสอน เป็นทางหนึ่งในการสร้างกลุ่มที่มีความคล้ายคลึงกันที่อยู่ในพื้นที่เดียวกัน โดยผู้ใช้ไม่ต้องรู้ถึงรายการที่จะถูกจำแนกมาก่อน ค่าการสะท้อนช่วงคลื่นทั้งหมดของทุกจุดจะถูกตรวจสอบเพื่อหาความสัมพันธ์ทางสถิติ 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

2.5 การวัดประสิทธิภาพ

การวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้าย เพื่อทำการตรวจสอบวิธีการที่ทำการทดลองมาข้างต้นว่ามีประสิทธิภาพมากหรือน้อยเพียงใดเมื่อนำมาใช้งานจริง จะเป็นขั้นตอนที่สำคัญ เพราะการนำวิธีการที่นำเสนอไปข้างต้นมาใช้งานได้นั้นจะต้องสอดคล้องกับความต้องการ จึงต้องมีการทดสอบศักยภาพการนำไปใช้ สถาปัตยกรรมที่ใช้ตัววัดความสำเร็จหลังการนำไปใช้หากนำไปใช้แล้วไม่ประสบผลสำเร็จต้องย้อนกลับไปเริ่มกระบวนการแรกใหม่ จึงต้องมีการประเมินผลก่อนการใช้งาน ในการประเมินนั้นกระทำได้โดยการวัดประสิทธิภาพของการจัดกลุ่มภาพมักจะถูกพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูลซึ่งจะประกอบด้วย การวัดค่าความแม่นยำ ค่าความระลึก 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 การวัดประสิทธิภาพ

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

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

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

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

- ค่าความถูกต้อง (accuracy: Acc) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องทั้งหมดจากจำนวนภาพที่มีอยู่

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

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

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