KKU Engineering Journal

http://www.en.kku.ac.th/enjournal/th/

โครงข่ายประสาทเทียมแบบไม่มีผู้สอนและการประยุกต์

Unsupervised Neural Networks and Its Applications

บัญชา อานนท์กิจพานิช*

Banchar Arnonkijpanich*

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น 40002

Department of Mathematics, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand, 40002.

Received April 2012

Accepted August 2012

บทคัดย่อ

เนื่องด้วยข้อมูลทางอุตสาหกรรมและวิทยาศาสตร์กำลังเพิ่มขึ้นอย่างรวดเร็วทั้งในแง่จำนวนข้อมูลและรายละเอียดของข้อมูล โครงข่ายประสาทเทียมจึงถูกนำมาใช้ในการกำหนดลักษณะที่สำคัญของข้อมูลเช่นโครงสร้างภายในและการแบ่งกลุ่มของข้อมูล บทความนี้แนะนำระเบียบวิธีของโครงข่ายประสาทเทียมแบบไม่มีผู้สอน ระเบียบวิธีเหล่านี้ถูกใช้ในการสกัดโครงสร้างที่ฝังตัวอยู่ในข้อมูล ซึ่งโครงสร้างที่ได้รับถือเป็นข้อมูลสำคัญในการวิเคราะห์ข้อมูลลำดับถัดไป โครงข่ายประสาทเทียมแบบไม่มีผู้สอนยังถูกแบ่งออกเป็นสองประเภทคือ วิธีแบ่งกลุ่มข้อมูลและวิธีลดจำนวนมิติของข้อมูล ซึ่งแต่ละวิธีจะถูกนำไปประยุกต์ใช้กับการแบ่งพื้นที่ในภาพและการสกัดโครงสร้างภายในจากข้อมูลพหุมิติ

คำสำคัญ : โครงข่ายประสาทเทียมแบบไม่มีผู้สอน การแบ่งกลุ่มข้อมูล การลดจำนวนมิติของข้อมูล การสกัดโครงสร้างของข้อมูลพหุมิติ

Abstract

Due to the rapid increase of size and resolution of industrial and scientific data sets, artificial neural networks have become essential tools for identifying the important aspects of the clustered data structure. In this article, we reviewed unsupervised neural network methods which can be applied to the task of extracting hidden structures as useful features for subsequent processing. The unsupervised learning algorithms reviewed here are grouped into two sections: data clustering methods and dimensionality reduction methods. Each of these major sections concludes with a discussion of successful applications of the methods to image segmentation and data visualization.

Keywords : Unsupervised neural networks, Data clustering, Dimensionality reduction, Data visualization

*Corresponding author. Tel.: +66 8 3140 9462

Email address: abanchar@kku.ac.th

1. บทนำ

ปัจจุบันนี้ การวิเคราะห์ข้อมูลด้วยโครงข่ายประสาทเทียม (Artificial neural networks) มีบทบาทสำคัญกับงานวิจัยหลายสาขาเช่น การจดจำรูปแบบ (Pattern recognition) การประมวลผลภาพและสัญญาณ (Signal/image processing) ชีวสารสนเทศศาสตร์ (Bioinformatics) และอื่นๆ จุดประสงค์หลักของการวิเคราะห์ข้อมูลคือการเข้าใจในความหมายแบบองค์รวมของข้อมูลและการสกัดเอาลักษณะสำคัญที่อยู่ในข้อมูลออกมาเช่น โครงสร้างของข้อมูลและความสัมพันธ์ระหว่างตัวแปรในชุดข้อมูล ซึ่งโครงข่ายประสาทเทียมสามารถใช้กับงานวิเคราะห์เหล่านี้ได้อย่างมีประสิทธิภาพ ทั้งนี้ โครงข่ายประสาทเทียมถูกแบ่งเป็นสามประเภทตามลักษณะการเรียนรู้คือ แบบมีผู้สอน (Supervised learning) แบบไม่มีผู้สอน (Unsupervised learning) และแบบเสริมกำลัง (Reinforcement learning) ในบทความนี้จะมุ่งไปที่ระเบียบวิธีของโครงข่ายประสาทเทียมแบบไม่มีผู้สอน ซึ่งมีเป้าหมายคือการสกัดโครงสร้างที่ฝังตัวในข้อมูลออกมา และแบ่งวิธีการสกัดโครงสร้างได้เป็นสองประเภทคือ วิธีแบ่งกลุ่มข้อมูล และวิธีลดจำนวนมิติของข้อมูล นอกจากนี้แล้วยังได้แนะนำตัววัดระยะทางที่ใช้ในระเบียบวิธีได้แก่ ระยะทางแบบยูคลิด (Euclidean distance) และระยะทางแบบวงรี (Ellipsoidal distance) พร้อมทั้งอธิบายการประยุกต์ใช้งานจริงในงานประมวลผลภาพด้วย

2. การแบ่งกลุ่มข้อมูล

หลักการแบ่งกลุ่มข้อมูลคือการรวบรวมข้อมูลที่มีลักษณะคล้ายกันให้รวมอยู่ในกลุ่มเดียวกัน โดยแบ่งกลุ่มข้อมูลที่ให้มาออกเป็นกลุ่มข้อมูลย่อยหลายๆกลุ่ม และข้อมูลที่อยู่ในแต่ละกลุ่มย่อยจะมีตัวแทนคือจุดศูนย์กลางของกลุ่มย่อยนั้น หรือที่เรียกว่าโปรโทไทพ์ (Prototype) การแบ่งกลุ่มอยู่บนพื้นฐานของการเรียนรู้แบบแข่งขัน (Competitive learning) ระเบียบวิธีของการแบ่งกลุ่มที่นิยมใช้ได้แก่ K-means, Self-Organizing Maps (SOM), Neural Gas (NG) ซึ่งใช้ตัววัดระยะทางแบบยูคลิด และมีรายละเอียดดังต่อไปนี้

2.1 การแบ่งกลุ่มโดยใช้ตัววัดระยะทางแบบยูคลิด

กำหนดให้เวกเตอร์ $\mathbf{x}_j \in \mathbb{R}^m, j = 1,\ldots,p$ และตำแหน่งโปรโทไทพ์ $\mathbf{w}_i \in \mathbb{R}^m, i = 1,\ldots,n$ เป้าหมายหลักของการแบ่งกลุ่มข้อมูลคือ แทนแต่ละข้อมูล $\mathbf{x}_j$ ด้วย $\mathbf{w}_i$ ที่ใกล้ที่สุด ซึ่งระยะทางระหว่าง $\mathbf{x}_j$ และ $\mathbf{w}_i$ สามารถคำนวณได้จากตัววัดระยะทางแบบยูคลิด

$$d(\mathbf{x}_j, \mathbf{w}_i) = (\mathbf{x}_j - \mathbf{w}_i)^t (\mathbf{x}_j - \mathbf{w}_i) \quad (1)$$

และกำหนดให้ k_{ij} เป็น Kronecker delta มีค่าเท่ากับ 1 หาก $\mathbf{w}_i$ เป็นโปรโทไทพ์ที่อยู่ใกล้กับ $\mathbf{x}_j$ มากที่สุด แต่ถ้าไม่สอดคล้องกับเงื่อนไขดังกล่าว k_{ij} จะมีค่าเท่ากับ 0 ดังสมการ

$$k_{ij} = \begin{cases} 1, & \text{if} \quad d(\mathbf{x}_j, \mathbf{w}_i) = \min\, d(\mathbf{x}_j, \mathbf{w}_l), \forall l \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

แล้วระเบียบวิธี K-means จะทำการแบ่งกลุ่มข้อมูลโดยมีเป้าหมายว่าต้องการหาตำแหน่งโปรโทไทพ์ $\mathbf{w}_i$, $i = 1,\ldots,n$ ที่เหมาะสมที่สุด ที่ทำให้ค่าของสมการวัตถุประสงค์มีค่าน้อยที่สุด (Minimization) ซึ่ง K-means ได้ใช้สมการวัตถุประสงค์ดังนี้

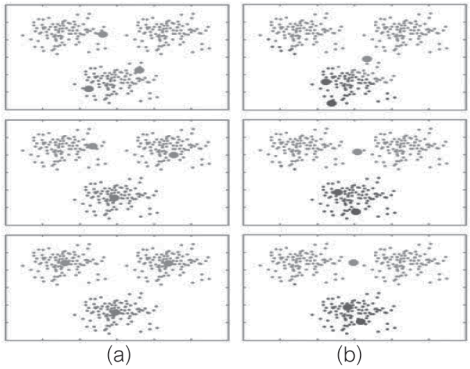
$$E(\mathbf{w}_i) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p} k_{ij} \cdot d(\mathbf{x}_j, \mathbf{w}_i) \quad (3)$$

K-means ถือเป็นวิธีแบ่งกลุ่มข้อมูลที่ง่ายที่สุด โดยเริ่มต้นด้วยการกำหนดตำแหน่งเริ่มต้นของโปรโทไทพ์ แล้วแต่ละโปรโทไทพ์ $\mathbf{w}_i$ จะถูกคำนวณด้วยสมการที่ (4) เพื่อปรับตำแหน่ง แล้วคำนวณซ้ำเช่นนี้ไปจนกว่าตำแหน่งโปรโทไทพ์ไม่เกิดการเปลี่ยนแปลง จึงถือว่าลู่เข้าสู่คำตอบ

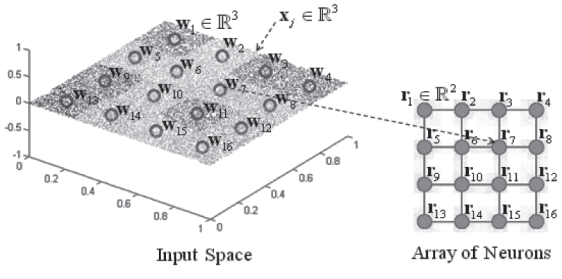
$$\mathbf{w}_i = \sum_j k_{ij}\mathbf{x}_j \Big/ \sum_j k_{ij} \quad (4)$$

แต่อย่างไรก็ตาม คำตอบสุดท้ายของวิธี K-means จะขึ้นอยู่กับการกำหนดตำแหน่งเริ่มต้นของโปรโทไทพ์ค่อนข้างมาก ถ้าหากกำหนดตำแหน่งเริ่มต้นไม่เหมาะสม กลุ่มของโปรโทไทพ์ $\mathbf{w}_i$ ที่ได้รับ เมื่อนำไปแทนค่าในสมการวัตถุประสงค์ดังสมการที่ (3) จะไม่ได้ค่าที่น้อยสุด ซึ่งเรียกปัญหาลักษณะนี้ว่า การลู่เข้าสู่ค่าที่น้อยที่สุดสัมพัทธ์ (Local minima) และลักษณะการปรับตำแหน่งของโปรโทไทพ์จะเป็นไปตามรูปที่ 1(b) ซึ่งโปรโทไทพ์จะไม่เคลื่อนตัวไปยังศูนย์กลางของแต่ละกลุ่มข้อมูล เพื่อแก้ปัญหาดังกล่าว SOM และ NG จึงได้เพิ่มขั้นตอนวิธีการรักษาคุณสมบัติโครงสร้าง (Topology preserving) เข้าไปในการแบ่งกลุ่มข้อมูลด้วย ซึ่งจะส่งผลให้ SOM และ NG ลดความผันผวนอันเนื่องมาจากการกำหนดตำแหน่งเริ่มต้นของโปรโทไทพ์ที่ไม่เหมาะสม และคำตอบ

สุดท้ายที่ได้รับซึ่งก็คือตำแหน่งโปรโทไทพ์ $\mathbf{w}_i$ เมื่อนำไปไปแทนในสมการวัตถุประสงค์ (5) และ (10) จะได้ค่าที่น้อยสุด ลักษณะเช่นนี้เรียกว่า การลู่เข้าสู่ค่าที่น้อยที่สุดสัมบูรณ์ (Global minima) นั่นเอง



รูปที่ 1 กำหนดให้ข้อมูลมีลักษณะการกระจายตัวออกเป็น 3 กลุ่ม จุดเล็กแสดงถึงเวกเตอร์ $\mathbf{x}_j$ แต่ละตัว และจุดวงกลมใหญ่แทนโปรโทไทพ์ $\mathbf{w}_i \in \mathbb{R}^2, i = 1,2,3$ (a) แสดงการปรับตำแหน่งโปรโทไทพ์จนลู่เข้าสู่จุดศูนย์กลางของแต่ละกลุ่มข้อมูล และเมื่อนำทุก $\mathbf{w}_i$ ไปแทนในสมการวัตถุประสงค์ตามสมการที่ (3) จะได้ค่าที่น้อยสุด ถือว่าลู่เข้าสู่ค่าที่น้อยที่สุดสัมบูรณ์ (Global minima) (b) แสดงปัญหาที่เกิดจากการกำหนดตำแหน่งโปรโทไทพ์เริ่มต้นที่ไม่เหมาะสม ซึ่งส่งผลให้โปรโทไทพ์ไม่ลู่เข้าสู่ตำแหน่งศูนย์กลางของข้อมูลแต่ละกลุ่ม และเมื่อนำ $\mathbf{w}_i$ ไปแทนในสมการที่ (3) จะไม่ได้ค่าที่น้อยสุด ซึ่งปัญหานี้เรียกว่า การลู่เข้าสู่ค่าที่น้อยที่สุดสัมพัทธ์ (Local minima)



รูปที่ 2 สถาปัตยกรรมของ SOM บนปริภูมิของข้อมูลนำเข้า และบนอาร์เรย์สองมิติของกลุ่มประสาทเทียม

SOM จัดเป็นวิธีการแบ่งกลุ่มข้อมูลอีกวิธีหนึ่ง ซึ่งถูกเสนอโดย Kohonen [1] เป็นวิธีการหาตัวแทนของกลุ่มข้อมูลที่รักษาคุณสมบัติโครงสร้าง SOM ประกอบไปด้วยสถาปัตยกรรมสองส่วน ส่วนแรกคือปริภูมินำเข้า (input space) ซึ่งมีข้อมูลนำเข้า $\mathbf{x}_j \in \mathbb{R}^m$ และโปรโทไทพ์ $\mathbf{w}_i \in \mathbb{R}^m$ โดยปกติแล้ว ปริภูมินำเข้าจะมีจำนวนมิติที่สูง $(m > 2)$ ส่วนที่สองคืออาร์เรย์ของประสาทเทียม (Array of neurons) ที่เป็นโครงร่างตาข่ายบนมิติที่ต่ำกว่า (Lateral lattice space) ส่วนใหญ่จะเป็นโครงร่างตาข่ายแบบสองมิติที่มีการเรียงตัวของประสาทเทียมแต่ละตัวอย่างเป็นระเบียบ ดังแสดงในรูปที่ 2 ส่วนของปริภูมินำเข้าและส่วนของอาร์เรย์ของประสาทเทียมมีความสัมพันธ์กันคือ โปรโทไทพ์ $\mathbf{w}_i$ บนปริภูมินำเข้าจะสมนัยกับตำแหน่งของประสาทเทียมบนโครงร่างตาข่ายสองมิติ $\mathbf{r}_i \in \mathbb{R}^2$ ดังนั้นจำนวนโปรโทไทพ์ในปริภูมินำเข้าจะเท่ากับจำนวนของประสาทเทียมในโครงร่างตาข่าย เป้าหมายหลักของ SOM คือการใช้กลุ่มของโปรโทไทพ์เพื่อเป็นตัวแทนของกลุ่มข้อมูลและเพื่อประมาณการกระจายตัวของข้อมูล สมการวัตถุประสงค์ของ SOM แสดงในสมการที่ (5)

$$E_{SOM}(\mathbf{w}_i) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p} h(i^*,i,t)\cdot d(\mathbf{x}_j,\mathbf{w}_i) \tag{5}$$

เมื่อ $h(i^*,i,t)$ คือ Neighborhood function ตามสมการที่ 6

$$h(i^*,i,t) = \exp\left(-\frac{\left\| \mathbf{r}_i - \mathbf{r}_{i^*} \right\|^2}{2\sigma^2(t)}\right) \tag{6}$$

โดยที่ $\mathbf{w}_{i^*}$ คือโปรโทไทพ์ที่อยู่ใกล้กับเวกเตอร์ $\mathbf{x}_j$ มากที่สุด และ $\mathbf{r}_{i^*}$ คือตำแหน่งของประสาทเทียมบนโครงร่างตาข่ายสองมิติที่สมนัยกับ $\mathbf{w}_{i^*}$ และ $\sigma(t)$ เป็นค่าของฟังก์ชันลด ณ ขั้นเวลา t ใดๆ ทั้งนี้ $\mathbf{w}_{i^*}$ และ $\mathbf{r}_{i^*}$ จะถูกเรียกว่าโปรโทไทพ์ผู้ชนะ (Winner prototype) และประสาทเทียมผู้ชนะ (Winner neuron) ตามลำดับ และค่า i^* สอดคล้องกับสมการ

$$i^* = \arg\min_i \sum_{l=1}^{n} h(i,l,t)\cdot d(\mathbf{x}_j,\mathbf{w}_l) \tag{7}$$

และเมื่อนำสมการวัตถุประสงค์ของ SOM มาทำการหาค่าเหมาะสมที่สุดแบบ Batch optimization [2] โดยเทียบกับตัวแปร $\mathbf{w}_i$ จะได้รับสมการในการปรับตำแหน่งโปรโทไทพ์ดังสมการที่ (8)

$$\mathbf{w}_i = \sum_j h(i^*,i,t)\,\mathbf{x}_j \Big/ \sum_j h(i^*,i,t) \tag{8}$$

จะเห็นว่าการปรับตำแหน่งโปรโทไทพ์แต่ละตัว ต้องใช้ข้อมูลของตำแหน่งของประสาทเทียม $\mathbf{r}_i$ และ $\mathbf{r}_{i^*}$ บนโครงร่างตาข่ายที่สมนัยกับโปรโทไทพ์ $\mathbf{w}_i$ และ $\mathbf{w}_{i^*}$ ซึ่งถือเป็นการใช้ข้อมูลของโปรโทไทพ์ข้างเคียงร่วมด้วย จึงทำให้ SOM มีความทนทาน (Robust) ต่อการกำหนดตำแหน่งเริ่มต้นของโปรโทไทพ์ กล่าวคือ แม้ว่าจะกำหนดตำแหน่งเริ่มต้นของโปรโทไทพ์ไม่เหมาะสม แต่ด้วยสมการปรับตำแหน่งโปรโทไทพ์ ตามสมการที่ (8) จะทำให้โปรโทไทพ์แต่ละตัวลู่เข้าสู่จุดศูนย์กลางของกลุ่มข้อมูล และเมื่อนำคำตอบสุดท้ายคือ $\mathbf{w}_i$ ไปแทนในสมการวัตถุประสงค์ตามสมการที่ (5) จะได้ค่าที่น้อยสุด ข้อสังเกตอีกอย่างหนึ่งของ SOM คือตำแหน่งของโปรโทไทพ์บนปริภูมินำเข้าจะถูกปรับตำแหน่งให้สอดคล้องกับตำแหน่งการเรียงตัวบนโครงร่างตาข่าย ซึ่งถือเป็นการรักษาคุณสมบัติโครงสร้าง โปรโทไทพ์ที่อยู่ใกล้กันบนปริภูมินำเข้า ก็จะมีตำแหน่งประสาทเทียมที่ใกล้กันบนโครงร่างตาข่ายด้วย ด้วยเหตุนี้รูปแบบนำเข้า (Input pattern) บนมิติสูง สามารถถูกแปลงให้มาอยู่บนโครงร่างตาข่ายสองมิติได้ และโครงสร้างข้อมูลบนปริภูมินำเข้า จะปรากฏอยู่บนโครงร่างตาข่าย จึงถือเป็นวิธีของ data visualization อีกวิธีหนึ่งเช่นกัน

NG ถูกเสนอโดย Martinetz et al. [3] ได้ใช้หลักการเดียวกับ SOM คือการรักษาคุณสมบัติโครงสร้าง แต่ทว่าไม่ได้ทำผ่านโครงร่างตาข่ายเหมือน SOM แต่อาศัยการปรับตำแหน่งโปรโทไทพ์ทั้งหมดในสัดส่วนที่เหมาะสม กล่าวคือ กลุ่มของโปรโทไทพ์ที่อยู่ใกล้เวกเตอร์นำเข้า (Input vector) จะถูกลำดับความสำคัญให้เป็นโปรโทไทพ์ลำดับแรกๆ และจะถูกปรับตำแหน่งมาก ในขณะที่กลุ่มของโปรโทไทพ์ที่อยู่ไกลจากเวกเตอร์นำเข้า จะถือว่าได้รับอิทธิพลจากเวกเตอร์นำเข้าค่อนข้างน้อย จึงปรับตำแหน่งน้อยตามสัดส่วน ด้วยเหตุนี้ จึงจำเป็นต้องเรียงลำดับของโปรโทไทพ์ (Prototype ranking) ที่ใกล้กับเวกเตอร์นำเข้า $\mathbf{x}_j$ ดังสมการ

$$d(\mathbf{x}_j, \mathbf{w}_{i_0}) \le d(\mathbf{x}_j, \mathbf{w}_{i_1}) \le \ldots \le d(\mathbf{x}_j, \mathbf{w}_{i_{n-1}}) \quad (9)$$

แล้วแต่ละโปรโทไทพ์จะต้องมีค่าที่บอกถึงลำดับความใกล้ ซึ่งใช้สัญลักษณ์คือ $rk(\mathbf{x}_j, \mathbf{w}_i) \in \{0,1,\ldots,n-1\}$ โดยที่โปรโทไทพ์ $\mathbf{w}_{i_0}$ ที่อยู่ใกล้ $\mathbf{x}_j$ มากที่สุด จะมีค่า $rk(\mathbf{x}_j, \mathbf{w}_{i_0}) = 0$ ในขณะที่โปรโทไทพ์ $\mathbf{w}_{i_1}$ ที่อยู่ใกล้ $\mathbf{x}_j$ รองลงมาจะมีค่า $rk(\mathbf{x}_j, \mathbf{w}_{i_1}) = 1$ เป็นเช่นนี้ไปเรื่อยๆ จนถึงโปรโทไทพ์ $\mathbf{w}_{i_{n-1}}$ ที่อยู่ไกลจาก $\mathbf{x}_j$ มากที่สุด จะมีค่า $rk(\mathbf{x}_j, \mathbf{w}_{i_{n-1}}) = n-1$ และเนื่องจาก NG ใช้ค่าลำดับความใกล้ ทดแทนค่าของ Neighborhood function ทำให้สมการวัตถุประสงค์ของ NG เปลี่ยนเป็น

$$E_{NG}(\mathbf{w}_i) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\exp\left(\frac{-rk(\mathbf{x}_j, \mathbf{w}_i)}{\sigma(t)}\right)\cdot d(\mathbf{x}_j, \mathbf{w}_i) \quad (10)$$

เมื่อ $\sigma(t)$ เป็นค่าของฟังก์ชันลด ณ ขั้นเวลา t ใดๆ และเมื่อนำสมการวัตถุประสงค์ของ NG มาทำการหาค่าเหมาะสมที่สุดแบบ Batch optimization จะได้รับสมการในการปรับตำแหน่งโปรโทไทพ์คือ

$$\mathbf{w}_i = \frac{\sum_j \exp\left(-rk(\mathbf{x}_j, \mathbf{w}_i)/\sigma(t)\right)\mathbf{x}_j}{\sum_j \exp\left(-rk(\mathbf{x}_j, \mathbf{w}_i)/\sigma(t)\right)} \quad (11)$$

จากสมการที่ (11) จะเห็นว่า วิธี NG จะปรับตำแหน่งของโปรโทไทพ์ทุกตัว ในทุกรอบของการคำนวณ โดยโปรโทไทพ์ที่อยู่ใกล้กับเวกเตอร์ $\mathbf{x}_j$ มากที่สุด จะถูกปรับตำแหน่งมากสุด ในขณะที่โปรโทไทพ์ที่อยู่ไกลสุด จะถูกปรับตำแหน่งน้อยสุด ด้วยหลักการเช่นนี้ทำให้ NG สามารถลดความผันผวนอันเนื่องมาจากการกำหนดตำแหน่งเริ่มต้นของโปรโทไทพ์ที่ไม่เหมาะสม เช่นเดียวกับ SOM

ทั้งนี้สมการวัตถุประสงค์ของ NG สามารถใช้การหาค่าเหมาะสมที่สุดแบบ Stochastic gradient descent ได้ และจะได้สมการปรับตำแหน่งโปรโทไทพ์คือ

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \varepsilon(t)\cdot\exp\left(\frac{-rk(\mathbf{x}_j, \mathbf{w}_i(t))}{\sigma(t)}\right)\cdot[\mathbf{x}_j - \mathbf{w}_i(t)] \quad (12)$$

เมื่อ $\varepsilon(t)$ เป็นค่าของฟังก์ชันลดที่ใช้แทน Step size โดยปกติจะกำหนดให้ $\varepsilon(t) \in [0,1]$ ในลักษณะเดียวกัน สมการวัตถุประสงค์ของ SOM ที่ใช้เทคนิค stochastic gradient descent ในการหาค่าเหมาะสมที่สุด ก็จะได้รับสมการปรับตำแหน่งโปรโทไทพ์คือ

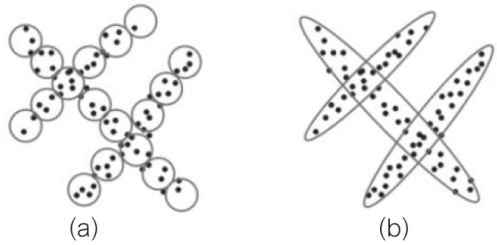
$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \varepsilon(t)\cdot h(i^*, i, t)\cdot[\mathbf{x}_j - \mathbf{w}_i(t)] \quad (13)$$

และในบางบทความวิจัย [4, 5] ได้ลดรูปสมการที่ (13) โดยละ Neighborhood function เอาไว้ เพื่อให้คำนวณได้รวดเร็วขึ้น และลดปัญหาในการกำหนดพารามิเตอร์ให้กับ

Neighborhood function แล้วสมการที่ (13) จะถูกลดรูปเหลือเพียง

$$\mathbf{w}_i(t+1)=\begin{cases}\mathbf{w}_i(t)+\varepsilon(t)[\mathbf{x}_j-\mathbf{w}_i(t)], & \text{if } d(\mathbf{x}_j,\mathbf{w}_i)=\min d(\mathbf{x}_j,\mathbf{w}_l),\forall l\\ \mathbf{w}_i(t), & \text{otherwise}\end{cases} \quad (14)$$

ระเบียบวิธี SOM และ NG ที่ได้แนะนำไป เหมาะสมกับการแบ่งกลุ่มข้อมูลที่มีการกระจายตัวแบบหนาแน่น (Dense) เพราะว่าใช้ตัววัดระยะทางแบบยูคลิด รูปร่างของกลุ่มจะมีลักษณะเป็นวงกลม แต่อย่างไรก็ตาม วิธี SOM และ NG ไม่เหมาะกับกลุ่มข้อมูลที่มีลักษณะโครงสร้าง ดังรูปที่ 3(a) ที่เป็นเช่นนี้เพราะว่าตัววัดระยะทางแบบยูคลิด ไม่ได้พิจารณาเรื่องสหสัมพันธ์ (Correlations) ระหว่างมิติ หากต้องการใช้ SOM และ NG กับข้อมูลที่มีลักษณะเชิงโครงสร้างเช่นนี้ จะต้องใช้จำนวนโปรโทไทพ์ค่อนข้างมาก ถึงจะเป็นตัวแทนของข้อมูลได้ดี แต่ทว่าการที่มีกลุ่มของโปรโทไทพ์ขนาดใหญ่ จะทำให้สิ้นเปลืองเวลาในการฝึกให้เรียนรู้ (Training) ดังนั้นเพื่อที่จะปรับปรุงวิธี SOM และ NG ให้สามารถจัดการกับข้อมูลที่มีลักษณะการกระจายตัวแบบต่างๆ ตัววัดระยะทางควรจะเป็นแบบ Non-Euclidean และรูปร่างของกลุ่มควรมีความยืดหยุ่นโดยปรับจากแบบวงกลม (Isotropic cluster shape) ไปเป็นวงรี (Ellipsoidal cluster shape) ดังรูปที่ 3(b) เพื่อให้สอดคล้องกับลักษณะการกระจายตัวของข้อมูล



รูปที่ 3 ข้อมูลที่มีการกระจายตัวแบบโครงสร้าง ที่ต้องการแบ่งกลุ่มโดยใช้ (a) วิธี SOM และ NG ที่ใช้ตัววัดระยะทางแบบยูคลิด จะให้รูปร่างของกลุ่มย่อยเป็นวงกลม ซึ่งไม่เหมาะกับข้อมูลที่มีลักษณะโครงสร้าง เพราะต้องใช้โปรโทไทพ์จำนวนมากในการเป็นตัวแทนของกลุ่มข้อมูล (b) แต่หากใช้ตัววัดระยะทางแบบ Non-Euclidean รูปร่างของกลุ่มย่อยจะเปลี่ยนเป็นวงรีได้ และใช้จำนวนโปรโทไทพ์น้อยกว่ามาก

2.2 การแบ่งกลุ่มโดยใช้ตัววัดระยะทางแบบวงรีที่ปรับตัวได้

การขยายขีดความสามารถของการแบ่งกลุ่มข้อมูลโดยที่ปรับเปลี่ยนรูปร่างของกลุ่มย่อยให้เหมาะสมกับการกระจายตัวของข้อมูล [6] จำเป็นต้องใช้ตัววัดระยะทางแบบที่ปรับตัวเองได้ตามสมการ

$$d(\mathbf{x}_j,\mathbf{w}_i,\Lambda_i)=(\mathbf{x}_j-\mathbf{w}_i)^t\Lambda_i(\mathbf{x}_j-\mathbf{w}_i) \quad (15)$$

เมตริกซ์ $\Lambda_i\in\mathbb{R}^{m\times m}$ ที่ถูกเพิ่มเข้าไป จะทำให้การวัดระยะทางยืดหยุ่นมากขึ้น โดยจะพิจารณาสหสัมพันธ์ (Correlation) ระหว่างมิติร่วมด้วย แล้วสมการวัตถุประสงค์ของ NG จะเปลี่ยนเป็น

$$E_{NG}(\mathbf{w}_i,\Lambda_i)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\exp\left(\frac{-rk(\mathbf{x}_j,\mathbf{w}_i,\Lambda_i)}{\sigma(t)}\right)\cdot d(\mathbf{x}_j,\mathbf{w}_i,\Lambda_i) \quad (16)$$

ซึ่งแต่ละโปรโทไทพ์ $\mathbf{w}_i$ จะมีค่าที่บอกถึงลำดับความใกล้กับเวกเตอร์นำเข้า $\mathbf{x}_j$ ซึ่งใช้สัญลักษณ์คือ $rk(\mathbf{x}_j,\mathbf{w}_i,\Lambda_i)\in\{0,1,\ldots,n-1\}$ โดยที่การหาค่าลำดับของโปรโทไทพ์ที่ใกล้กับเวกเตอร์ $\mathbf{x}_j$ จะต้องใช้การวัดระยะทางแบบวงรีด้วย และสอดคล้องกับสมการ

$$d(\mathbf{x}_j,\mathbf{w}_{i_0},\Lambda_{i_0})\le d(\mathbf{x}_j,\mathbf{w}_{i_1},\Lambda_{i_1})\le\ldots\le d(\mathbf{x}_j,\mathbf{w}_{i_{n-1}},\Lambda_{i_{n-1}}) \quad (17)$$

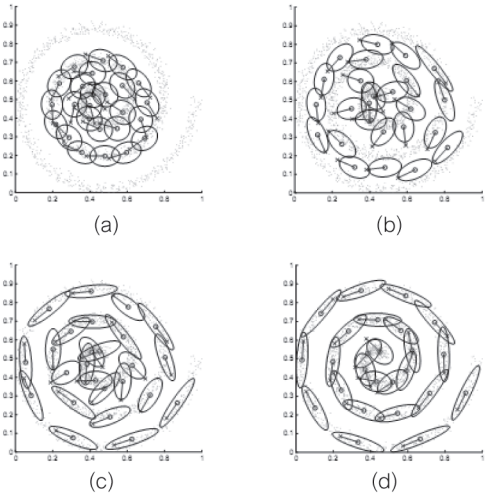
ดังนั้น โปรโทไทพ์ $\mathbf{w}_{i_0}$ ที่อยู่ใกล้ $\mathbf{x}_j$ มากที่สุด จะมีค่า $rk(\mathbf{x}_j,\mathbf{w}_{i_0},\Lambda_{i_0})=0$ ในขณะที่โปรโทไทพ์ $\mathbf{w}_{i_1}$ ที่อยู่ใกล้ $\mathbf{x}_j$ รองลงมาจะมีค่า $rk(\mathbf{x}_j,\mathbf{w}_{i_1},\Lambda_{i_1})=1$ เป็นเช่นนี้ไปเรื่อยๆ และงานวิจัยใน [7] ได้ใช้เทคนิค Batch optimization เพื่อหาคำตอบที่เหมาะสมที่สุดของสมการที่ (16) แล้วจะได้สมการปรับตำแหน่งโปรโทไทพ์ $\mathbf{w}_i$ และสมการปรับค่าในเมตริกซ์ Λ_i ดังนี้

$$\mathbf{w}_i=\frac{\sum_j\exp\left(-rk(\mathbf{x}_j,\mathbf{w}_i,\Lambda_i)/\sigma(t)\right)\mathbf{x}_j}{\sum_j\exp\left(-rk(\mathbf{x}_j,\mathbf{w}_i,\Lambda_i)/\sigma(t)\right)} \quad (18)$$

$$\Lambda_i=S_i^{-1}(\det S_i)^{1/n} \quad (19)$$

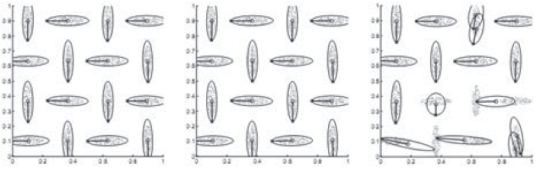
$$S_i=\sum_{j=1}^{p}\exp\left(\frac{-rk(\mathbf{x}_j,\mathbf{w}_i,\Lambda_i)}{\sigma(t)}\right)\cdot(\mathbf{x}_j-\mathbf{w}_i)(\mathbf{x}_j-\mathbf{w}_i)^t \quad (20)$$

สำหรับรอบแรกของการคำนวณนั้น เมตริกซ์ Λ_i จะถูกกำหนดให้เป็นเมตริกซ์เอกลักษณ์ (Identity matrix) ซึ่งจะให้รูปร่างของกลุ่มย่อยเป็นวงกลม จากนั้นเมื่อการเรียนรู้ได้ดำเนินไป $\mathbf{w}_i$ และ Λ_i จะถูกปรับค่าด้วยสมการ (18) และ (19) แบบวนซ้ำ แล้วจะนำไปสู่การเปลี่ยนรูปร่างของกลุ่มย่อยจากวงกลมไปเป็นวงรี ดังรูปที่ 4 จะเห็นว่าในช่วงแรกของการเรียนรู้ตำแหน่งโปรโทไทพ์จะเคลื่อนที่เข้าหาข้อมูล และรูปร่างของกลุ่มย่อยจะเริ่มเปลี่ยนเป็นวงรี



รูปที่ 4 การแบ่งกลุ่มของข้อมูลก้นหอย (Spiral data) ด้วยวิธี NG โดยใช้ตัววัดระยะทางแบบวงรีที่ปรับตัวได้ และกำหนดจำนวนรอบของการคำนวณ (Epochs) คือ 90 รอบ จากผลการทดลองจะเห็นว่า (a) แสดงผลการทดลองเมื่อผ่านไป 20 รอบ (b) แสดงผลลัพธ์เมื่อคำนวณไปแล้ว 40 รอบ จะเห็นว่าตำแหน่งโปรโทไทพ์จะเคลื่อนที่เข้าหาจุดศูนย์กลางของกลุ่มข้อมูลย่อย พร้อมทั้งรูปร่างของกลุ่มข้อมูลย่อยเริ่มเปลี่ยนจากวงกลมเป็นวงรี (c) เริ่มสังเกตได้ว่าลู่เข้าสู่คำตอบที่เหมาะสม และ (d) แสดงผลลัพธ์หลังจากคำนวณครบ 90 รอบ จะเห็นว่าโปรโทไทพ์ทุกตัวฝังตัวอยู่ในบริเวณข้อมูล และมีรูปร่างของกลุ่มย่อยสอดคล้องพอดีกับการกระจายตัวของข้อมูล

จากนั้นจะลู่เข้าสู่คำตอบที่เหมาะสมที่สุด ซึ่งรูปร่างของกลุ่มย่อยจะปรับเป็นวงรีอย่างเต็มที่เพื่อให้สอดรับกับการกระจายตัวของข้อมูล นอกจากนี้แล้ว ตัววัดระยะทางแบบวงรีที่ปรับตัวได้ ยังสามารถรวมเข้ากับวิธี SOM และ K-means ได้ในลักษณะเดียวกัน รายละเอียดให้ดูใน [7] ผลลัพธ์ของการใช้ตัววัดระยะทางแบบวงรีถูกแสดงในรูปที่ 5 และจากรูป 5(c) จะเห็นว่าการเพิ่มตัววัดระยะทางแบบวงรีที่ปรับตัวได้ให้กับ K-means แม้ว่าจะได้รูปร่างของกลุ่มย่อยเป็นวงรี แต่กลุ่มของโปรโทไทพ์ก็ไม่ลู่เข้าสู่จุดศูนย์กลางของกลุ่มข้อมูล ซึ่งถือว่าเป็นปัญหาการลู่เข้าสู่ค่าที่น้อยที่สุดสัมพัทธ์ โดยมีสาเหตุมาจากการกำหนดตำแหน่งเริ่มต้นของโปรโทไทพ์ที่ไม่เหมาะสม แต่ถ้าหากนำตัววัดระยะทางแบบวงรีที่ปรับตัวได้ รวมเข้ากับ SOM และ NG ซึ่งมีคุณสมบัติของการรักษาโครงสร้าง จะทำให้กำจัดข้อเสียในจุดนี้ได้ แล้วจะให้คำตอบที่ดีกว่า ดังรูปที่ 5(a) และ 5(b) ตามลำดับ



รูปที่ 5 แสดงผลลัพธ์ของการแบ่งกลุ่มข้อมูลโดยใช้ตัววัดระยะทางแบบวงรีที่ปรับตัวได้ ข้อมูลมีการกระจายตัวแบบ Multimodal (a) และ (b) คือผลลัพธ์ของการรวมตัววัดระยะทางแบบวงรีที่ปรับตัวได้เข้ากับวิธีรักษาโครงสร้างแบบ SOM และ NG จะเห็นว่าโปรโทไทพ์แต่ละตัวลู่เข้าสู่จุดศูนย์กลางของกลุ่มข้อมูลย่อย ในขณะที่ (c) คือผลลัพธ์ของ K-means ที่ใช้ตัววัดระยะทางแบบวงรีเช่นกัน จะเห็นว่าตำแหน่งโปรโทไทพ์ไม่ได้อยู่ที่จุดศูนย์กลางของกลุ่มข้อมูลย่อย ซึ่งมีสาเหตุมาจากการกำหนดตำแหน่งเริ่มต้นของโปรโทไทพ์ไม่เหมาะสม

มีข้อสังเกตประการหนึ่งเกี่ยวกับตัววัดระยะทางที่ปรับเปลี่ยนรูปร่างของกลุ่มย่อยได้ เมื่อการเรียนรู้ลู่เข้าสู่ค่าที่น้อยที่สุดสัมบูรณ์ เมตริกซ์ Λ_i จะลู่เข้าสู่ผกผัน (Inverse) ของเมตริกซ์สหสัมพันธ์ แล้วตัววัดระยะทางแบบวงรีที่ปรับตัวได้ตามสมการที่ (15) จะถือเป็นการวัดระยะทางแบบมาหาลาโนบิส (Mahalanobis distance) รายละเอียดการพิสูจน์อยู่ใน [7]

3. การแบ่งพื้นที่ในภาพ

ในหัวข้อนี้ จะแนะนำการประยุกต์ใช้วิธีการแบ่งกลุ่มข้อมูลกับปัญหาการแบ่งพื้นที่ในภาพ (Image segmentation) หลักการของการแบ่งพื้นที่ในภาพ คือการแบ่งกลุ่มของพิกเซลตามลักษณะความคล้าย ซึ่งความคล้ายจะพิจารณาได้จากการวัดระยะทาง ทั้งนี้จะใช้ตัววัดระยะทางแบบยูคลิดหรือแบบวงรีก็ได้ หากระยะทางระหว่างสองเวกเตอร์น้อยมากแสดงว่าสองเวกเตอร์มีความคล้ายคลึงกันมาก หากพิจารณาข้อมูลภาพทดสอบ ดังรูปที่ 6(a) ซึ่งเป็นภาพระดับสีเทา (Gray scale) 8 บิต ขนาด 256x256 พิกเซล เมื่อสุ่มเลือกกลุ่มของ

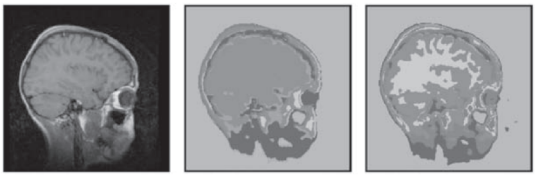
พิกเซลขนาด 7x7 แล้วแปลงให้เป็นเวกเตอร์ $\mathbf{x}_j$ ขนาด m =49 มิติ แล้วทำการสุ่มเลือกเช่นนี้จำนวน p ครั้ง จะได้เซตของเวกเตอร์ $\mathbf{x}_j$; $j = 1,2,...,p$ จากนั้นกำหนดจำนวนโปรโทไทพ์และตำแหน่งของโปรโทไทพ์ตอนเริ่มต้น $\mathbf{w}_i$; $i = 1,2,...,n$ แล้วใช้วิธีการแบ่งกลุ่มข้อมูลเช่น K-means, SOM, และ NG กับเซตข้อมูล $\mathbf{x}_j$ $\mathbf{x}_j$ ผลลัพธ์ที่ได้คือตำแหน่งของโปรโทไทพ์ $\mathbf{w}_i$ ที่เป็นตัวแทนของกลุ่มข้อมูลแต่ละกลุ่ม เนื่องด้วยกำหนดให้จำนวนโปรโทไทพ์คือ n จะทำให้แบ่งกลุ่มของพิกเซลได้เป็น n กลุ่มด้วย จากนั้นแปลงค่าของโปรโทไทพ์ให้เป็นระดับสี ซึ่งอยู่ในระบบสีแบบมุนเซล (Munsell color system) สีของ $\mathbf{w}_i$ จะคำนวณได้จากสมการ

$$color(\mathbf{w}_i) = i/n \times 360^{\circ} \quad (21)$$

จากรูปที่ 6(b) ได้กำหนดให้ $n = 7$ ดังนั้นภาพทดสอบจะถูกแบ่งออกเป็น 7 กลุ่มของพิกเซล ซึ่งแทนด้วย 7 สี แต่ละสีแสดงถึง บริเวณหลังคา บริเวณผนัง บริเวณบานประตู บริเวณท่อน้ำ เป็นต้น แต่อย่างไรก็ตาม วิธีการแบ่งกลุ่มข้อมูลที่ใช้ตัววัดระยะทางแบบยูคลิด และแบบวงรีที่ปรับตัวได้จะให้ผลลัพธ์ที่แตกต่างกัน จากรูปที่ 3 จะเห็นได้ชัดว่าการแบ่งกลุ่มข้อมูลที่ใช้ตัววัดระยะทางที่ยืดหยุ่นกว่า จะส่งผลให้แบ่งกลุ่มข้อมูลได้มีประสิทธิภาพมากกว่า และพอดีกับการกระจายของข้อมูล ดังนั้นหากนำมาใช้กับปัญหาการแบ่งพื้นที่ในภาพ ก็น่าจะให้ผลลัพธ์ที่ดีกว่าด้วย เมื่อทดสอบกับภาพถ่ายแบบ Magnetic resonance imaging (MRI) ดังรูปที่ 7(a) ซึ่งเป็นภาคตัดขวางของกะโหลกศีรษะ เมื่อใช้ตัววัดระยะทางแบบยูคลิด ผลลัพธ์ของการแบ่งพื้นที่ตามรูปที่ 7(b) แสดงให้เห็นว่า พื้นที่บริเวณ Cerebrum ทั้งหมดถูกจำแนกเป็นพื้นที่เดียวกัน ซึ่งแทนด้วยสีม่วง แต่ทว่าจะไม่ได้รายละเอียดปลีกย่อยบนบริเวณ Cerebrum ในขณะที่รูป 7(c) ซึ่งใช้ตัววัดระยะทางแบบวงรีที่ปรับตัวได้ นอกจากจะได้บริเวณ Cerebrum แล้ว ยังได้ผลลัพธ์เป็นบริเวณสีเขียวเพิ่มเติม ซึ่งเป็นส่วนของ Frontal lobe และ Parietal lobe ที่ควบคุมการเคลื่อนไหว การออกเสียง การสัมผัส และการรับรู้รสชาติ

(a) (b)

รูปที่ 6 การแบ่งกลุ่มข้อมูลสามารถนำไปประยุกต์ใช้กับงานแบ่งพื้นที่ในภาพได้ (a) ภาพทดสอบคือภาพบ้าน (house image) และ (b) แสดงผลลัพธ์ของการแบ่งพื้นที่ในภาพ โดยแบ่งกลุ่มของพิกเซลออกเป็น 7 กลุ่ม แต่ละกลุ่มจะมีสีประจำกลุ่ม ดังนั้นภาพที่ผ่านการแบ่งพื้นที่แล้วจะมีจำนวนสีทั้งหมด 7 สี



(a) (b) (c)

รูปที่ 7 การเปรียบเทียบผลลัพธ์ของการแบ่งพื้นที่ในภาพ ระหว่างวิธีแบ่งกลุ่มข้อมูลที่ใช้ตัววัดระยะทางแบบยูคลิดและแบบวงรีที่ปรับตัวได้ (a) แสดงภาพทดสอบคือภาพถ่าย MRI ซึ่งเป็นภาคตัดขวางของกะโหลกศีรษะ (b) และ (c) แสดงผลลัพธ์ของการแบ่งพื้นที่ในภาพออกเป็น 9 กลุ่มของพิกเซล โดยใช้ตัววัดระยะทางแบบยูคลิด และแบบวงรีที่ปรับตัวได้ ตามลำดับ

4. การลดจำนวนมิติของข้อมูล

นอกเหนือไปจากการแบ่งกลุ่มข้อมูลแล้วโครงข่ายประสาทเทียมแบบไม่มีผู้สอน ยังมีหน้าที่หลักอีกอย่างหนึ่ง คือ การค้นหาโครงสร้างที่ซ่อนอยู่ในข้อมูล ผ่านทางการลดจำนวนมิติ ซึ่งใช้หลักการแปลงเซตข้อมูล $X = \{\mathbf{x}_j \in \mathbb{R}^m \mid j = 1,\ldots, p\}$ ไปยังเซตข้อมูลใหม่ $Y = \{\mathbf{y}_j \in \mathbb{R}^k \mid j = 1,\ldots, p\}$ เมื่อ $k < m$ แล้วจะเรียกลักษณะเช่นนี้ว่าข้อมูล $\mathbf{y}_j$ k มิติ ฝังตัวในปริภูมิ m มิติ และจะเรียกปริภูมิที่มีมิติต่ำกว่าว่า ปริภูมิฝังใน (Embedding space) แต่เนื่องจากยังไม่ทราบถึงข้อมูลเกี่ยวกับโครงสร้างและตำแหน่งการเรียงตัวของ $\mathbf{y}_j$ บนปริภูมิฝังใน ดังนั้นการลดจำนวนมิติจะต้องคำนวณหาเซตข้อมูลใหม่ Y บนมิติต่ำ โดยต้องรักษาลักษณะทางกายภาพของข้อมูลให้คล้ายคลึงกับเซตข้อมูล X บนมิติสูง ให้มากที่สุดเท่าที่จะทำได้ ตัวอย่างของลักษณะทางกายภาพที่ต้องรักษาไว้เช่น ระยะ

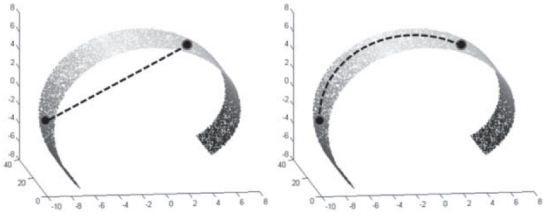
ทางระหว่างจุดเวกเตอร์สองจุดใดๆ ถ้าสมมุติว่าระยะทางระหว่างเวกเตอร์ $\mathbf{x}_i$ และ $\mathbf{x}_j$ บนมิติสูงคือ d หน่วย เราจะต้องคำนวณหาเวกเตอร์ $\mathbf{y}_i$ และ $\mathbf{y}_j$ บนมิติต่ำ ที่มีระยะห่างระหว่าง $\mathbf{y}_i$ และ $\mathbf{y}_j$ เท่ากับ d หน่วยด้วยเช่นกัน แนวความคิดเช่นนี้เรียกว่า การลดจำนวนมิติโดยใช้หลักการรักษาระยะทาง (Distance preservation) ซึ่งจะกล่าวถึงอย่างละเอียดในหัวข้อ 4.1 ต่อไป และยังมีคุณสมบัติของข้อมูลอีกอย่างหนึ่งที่ต้องรักษาไว้คือ แต่ละเวกเตอร์ $\mathbf{x}_j$ ที่อยู่บนมิติสูง จะต้องมี $\mathbf{y}_j$บนมิติต่ำเพียงเวกเตอร์เดียวเท่านั้นที่สมนัยกับ $\mathbf{x}_j$ และถ้ากำหนดให้มีเมตริกซ์การแปลงเชิงเส้น $A_i \in \mathbb{R}^{k\times m}, i = 1,\ldots,n$ แล้วเวกเตอร์ $\mathbf{x}_j \in \mathbb{R}^m$ จะถูกลดจำนวนมิติผ่านทาง $\boldsymbol{A}_i$ โดยลดจำนวนมิติจาก m มิติ เหลือเพียง k มิติ ดังสมการ

$$\mathbf{z}_{j1} = A_1(\mathbf{x}_j - \overline{\mathbf{x}}_1)\ ,\ \ldots\ ,\ \mathbf{z}_{jn} = A_n(\mathbf{x}_j - \overline{\mathbf{x}}_n) \tag{22}$$

ทั้งนี้กลุ่มของ $\mathbf{z}_{j1}, \mathbf{z}_{j2},\ldots,\mathbf{z}_{jn}$ อยู่บนปริภูมิ k มิติเหมือนกัน แต่ทว่ามีพิกัดที่ไม่ต่อเนื่องกัน สาเหตุมาจากการใช้เมตริกซ์การแปลง $\boldsymbol{A}_i$ ที่ต่างกัน แล้วจะเรียก $\mathbf{z}_{j1}, \mathbf{z}_{j2},\ldots,\mathbf{z}_{jn}$ ว่าพิกัดเฉพาะที่ (Local coordination) แต่อย่างไรก็ตามจากเงื่อนไขที่ว่า $\mathbf{x}_j \in \mathbb{R}^m$ จะต้องสมนัยกับ $\mathbf{y}_j \in \mathbb{R}^k$ เพียงเวกเตอร์เดียว ดังนั้นจึงจำเป็นต้องผสานพิกัดเฉพาะที่ $\mathbf{z}_{j1}, \mathbf{z}_{j2},\ldots,\mathbf{z}_{jn}$ ให้เหลือเพียงพิกัดเดียวซึ่งเรียกว่าพิกัดรวม (Global coordination) โดยใช้การแปลงสัมพรรค (Affine mapping)

$$\mathbf{y}_j = \text{Affine\_Mapping}(\mathbf{z}_{j1}, \mathbf{z}_{j2},\ldots,\mathbf{z}_{jn}) \tag{23}$$

และแนวความคิดเช่นนี้ถูกเรียกว่า การลดจำนวนมิติที่ใช้หลักการพิกัดรวม ซึ่งระเบียบวิธีที่ใช้หลักการพิกัดรวมจะถูกกล่าวถึงโดยละเอียดอีกครั้งในหัวข้อที่ 4.2



(a) (b)

รูปที่ 8 ข้อมูลที่มีการกระจายตัวแบบไม่เชิงเส้นบนพื้นผิวโค้งสามมิติ (a) ระยะทางแบบยูคลิดระหว่างจุดเวกเตอร์สองจุดใดๆบนพื้นผิว ซึ่งไม่ได้สะท้อนระยะทางที่แท้จริง (b) ระยะทางแบบ Geodesic distance ระหว่างจุดสองจุดบนพื้นผิว ซึ่งเป็นการวัดระยะทางตามลักษณะการกระจายตัวของข้อมูล

4.1. การลดจำนวนมิติแบบรักษาระยะทาง

หลักการของการรักษาระยะทางมีดังนี้ กำหนดให้เวกเตอร์ $\mathbf{x}_j$ บนมิติสูงและเวกเตอร์ $\mathbf{y}_j$ บนมิติต่ำสมนัยกัน และให้ X_{ij} คือระยะทางแบบยูคลิดระหว่างเวกเตอร์ $\mathbf{x}_i$ และ $\mathbf{x}_j$ และ Y_{ij} แทนระยะทางแบบยูคลิดระหว่างเวกเตอร์ $\mathbf{y}_i$ และ $\mathbf{y}_j$ แล้วต้องการให้ระยะทางระหว่าง $\mathbf{x}_i$ และ $\mathbf{x}_j$ บนมิติสูง กับระยะทางระหว่าง $\mathbf{y}_i$ และ $\mathbf{y}_j$ บนมิติต่ำ มีระยะใกล้เคียงกันมากที่สุด ซึ่งสอดคล้องกับสมการวัตถุประสงค์คือ

$$E = \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1,j\neq i}^{p}(X_{ij} - Y_{ij})^2 \tag{24}$$

แล้วจะเรียกวิธีที่ใช้สมการวัตถุประสงค์ (24) นี้ว่า วิธี Multidimensional scaling (MDS) แต่ทว่าจากรูปที่ 8(a) ตัววัดระยะทางแบบยูคลิด ไม่ได้วัดระยะทางเชิงกายภาพของข้อมูล ขณะที่รูป 8(b) ใช้ตัววัดระยะทางแบบ Geodesic distance ซึ่งเป็นระยะทางที่แท้จริงตามรูปร่างของข้อมูล ดังนั้นในสมการที่ (24) สามารถเปลี่ยนตัววัดระยะทางบนมิติสูงจากแบบยูคลิดเป็นแบบ Geodesic distance ได้ ในขณะที่บนมิติต่ำ ยังคงใช้ระยะทางแบบยูคลิดเช่นเดิม แล้วเปลี่ยนชื่อวิธีเป็น Isometric mapping (ISOMAP) [8] การคำนวณ หาชุดข้อมูลใหม่ของ $\mathbf{y}j$ สามารถทำได้สองแนวทาง แนวทางแรกเป็นการใช้เทคนิคทางพีชคณิตเชิงเส้น เริ่มต้นด้วยการหาเมตริกซ์ระยะทาง $D(i, j) = Xij$ ซึ่งจะใช้ระยะทางแบบยูคลิดหรือแบบ Geodesic distance ก็ได้ จากนั้นทำการ double centering ด้วยสมการ

$$B = -\frac{1}{2}JD^{(2)}J \tag{25}$$

$$J = I - \frac{1}{p}A \tag{26}$$

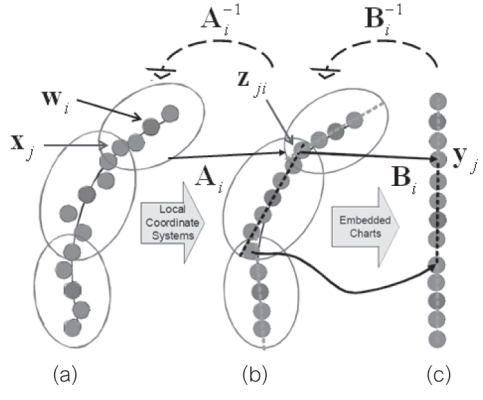
เมื่อ I คือเมตริกซ์เอกลักษณ์ และ A คือเมตริกซ์ซึ่งประกอบไปด้วยสมาชิก 1 และมีมิติ $p \times p$ จากนั้นใช้การกระจายเจาะจง (Eigendecomposition) บน B เพื่อหาค่าเจาะจง (Eigenvalue) ที่มากที่สุด k ตัวแรก $\lambda_1,\ldots,\lambda_k$ และเวกเตอร์เจาะจง (Eigenvector) ที่สมนัยกัน $e_1,\ldots,e_k$ แล้วเซตข้อมูลใหม่ $Y = \{\mathbf{y}_j \in \mathbb{R}^k\}$ คำนวณได้จากสมการ

$$Y = \left[\begin{bmatrix}\vdots\\ e_1\\ \vdots\end{bmatrix} \cdots \begin{bmatrix}\vdots\\ e_k\\ \vdots\end{bmatrix}\right] \begin{bmatrix}\sqrt{\lambda_1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sqrt{\lambda_k}\end{bmatrix} \quad (27)$$

แนวทางที่สองเป็นการนำสมการวัตถุประสงค์ ตามสมการที่ (24) มาหาค่าเหมาะสมที่น้อยสุด (Minimizing an objective function) แล้วจะได้สมการปรับค่า $\mathbf{y}_j$ คือ

$$\mathbf{y}_j(t+1) = \mathbf{y}_j(t) + \alpha(t)(X_{ij} - Y_{ij})\frac{\mathbf{y}_j(t) - \mathbf{y}_i(t)}{Y_{ij}}; \forall j \neq i \quad (28)$$

เมื่อ $\alpha(t)$ คืออัตราการเรียนรู้ (Learning rate) ณ ขั้นเวลา t ใดๆ ทั้งนี้วิธีการลดจำนวนมิติซึ่งใช้หลักการรักษาระยะทางยังมีอีกหลายวิธีเช่น Sammon's nonlinear mapping [9] และ Curvilinear component analysis [10] เป็นต้น



รูปที่ 9 แสดงถึงการลดจำนวนมิติแบบพิกัดรวม เริ่มต้นโดย (a) แบ่งกลุ่มข้อมูลโดยใช้ตัววัดระยะทางแบบวงรีที่ปรับตัวได้ จากนั้นลดจำนวนมิติผ่านทางการฉายแบบเชิงเส้นเฉพาะที่ (Local linear projection) A_i จะได้ (b) ข้อมูลในระบบพิกัดเฉพาะที่ แล้วใช้การแปลงสัมพรรคเฉพาะที่ (Local affine mapping) เพื่อผสานพิกัดเฉพาะที่ให้เป็น (c) ระบบพิกัดรวม

4.2 การลดจำนวนมิติแบบพิกัดรวม

วิธีการลดจำนวนมิติแบบพิกัดรวมที่จะแนะนำในหัวข้อนี้คือวิธี Manifold charting [11,12] ซึ่งแบ่งขั้นตอนการทำงานเป็นสองส่วนคือ ส่วนของการสร้างพิกัดเฉพาะที่ (Local coordination) และส่วนของการนำพิกัดเฉพาะที่มารวมกันเป็นพิกัดรวม (Global coordination) ส่วนที่หนึ่งเริ่มด้วยการแบ่งกลุ่มข้อมูลออกเป็นกลุ่มย่อย ดังรูปที่ 9(a) ทั้งนี้ต้องใช้การแบ่งกลุ่มที่มีตัววัดระยะทางแบบวงรีที่ปรับตัวได้เพราะจะให้ตำแหน่งโปรโทไทพ์ $\mathbf{w}_i$ และเมตริกซ์ Λ_i ของแต่ละกลุ่มย่อยออกมาด้วย และจากสมการที่ (19) Λ_i จะสู่เข้าสู่ผกผันของเมตริกซ์สหสัมพันธ์ ถ้านำ Λ_i ไปปรับขนาดให้เป็นมาตรฐาน (Normailization) และหาผกผันซ้ำอีกครั้ง จะได้ C_i เป็นเมตริกซ์ความแปรปรวนร่วม (Covariance matrix) ของกลุ่มย่อย จากนั้นทำการหาค่าความน่าจะเป็นที่เวกเตอร์ $\mathbf{x}_j$ เป็นของ $\mathbf{w}_i$ (Responsibilities) ด้วยสมการ

$$R_{ji} = \frac{p_i}{p} \cdot \frac{1}{(2\pi)^{m/2}\sqrt{\det C_i}} \cdot F(\mathbf{x}_j, \mathbf{w}_i) \quad (29)$$

$$F(\mathbf{x}_j, \mathbf{w}_i) = \exp\left(-\frac{1}{2} \cdot (\mathbf{x}_j - \mathbf{w}_i)^t \cdot C_i^{-1} \cdot (\mathbf{x}_j - \mathbf{w}_i)\right) \quad (30)$$

เมื่อ p_i คือจำนวนเวกเตอร์ในกลุ่มย่อยที่ i ทั้งนี้การหา $\mathbf{w}_i$ และ C_i สามารถใช้วิธี Mixture of Probabilistic PCA [13] ได้เช่นกัน นอกจากนี้แล้ว แต่ละกลุ่มย่อยจะมีการฉายแบบเชิงเส้นเฉพาะที่ (Local linear projections) ซึ่งคำนวณมาจาก การกระจายเจาะจงบน Λ_i เพื่อหาค่าเจาะจงที่น้อยที่สุด k ตัวแรก $\lambda_1,...,\lambda_k$ และเวกเตอร์เจาะจงที่สมนัยกัน $e_1,...,e_k$ แล้วการฉายแบบเชิงเส้นเฉพาะที่คือ

$$A_i = \left(\left[\begin{bmatrix}\vdots\\ e_1\\ \vdots\end{bmatrix} \cdots \begin{bmatrix}\vdots\\ e_k\\ \vdots\end{bmatrix}\right]_{m\times k}\right)^t \quad (31)$$

จะเห็นว่า A_i อยู่ในรูปแบบของเมตริกซ์การแปลงเชิงเส้น (The matrix of a linear transformation) ที่มีมิติ k x m เวกเตอร์ $\mathbf{x}_j$ ซึ่งมีมิติ m x 1 จะถูกปรับเปลี่ยนขนาด ผ่านทางการแปลง

$$\mathbf{z}_{ji} = A_i(\mathbf{x}_j) = A_i \cdot (\mathbf{x}_j - \mathbf{w}_i) \quad (32)$$

จะได้ว่าเวกเตอร์ $\mathbf{z}_{ji}$ มีมิติลดเหลือ k x 1 ดังนั้นเมตริกซ์ $A_i : \mathbb{R}^m \to \mathbb{R}^k;\ k < m$ จึงถือเป็นวิธีลดจำนวนมิติของข้อมูลแบบเชิงเส้น และจะเรียก $\mathbf{z}_{ji}$ ว่าเป็นเวกเตอร์ในพิกัดเฉพาะที่ เมื่อนำเวกเตอร์ $\mathbf{x}_j$ $\mathbf{x}_j$ เวกเตอร์เดิมไปผ่านการแปลงด้วยเมตริกซ์ A_l จะได้เวกเตอร์ในพิกัดเฉพาะที่คือ $\mathbf{z}_{jl}$ จากรูปที่ 9(b) จะเห็นว่าข้อมูลบนพิกัดเฉพาะที่ไม่ได้เรียงตัวอย่างต่อเนื่อง (Smooth) เท่าที่ควร เนื่องจาก $\mathbf{z}_{ji}$ และ $\mathbf{z}_{jl}$ อยู่บนพิกัดที่ต่างกัน สาเหตุมาจาก A_i และ A_l ไม่ได้ส่ง $\mathbf{x}_j$ ไปยังตำแหน่งเดียวกัน ส่งผลให้โครงสร้างที่แท้จริงของข้อมูลยังไม่ได้ปรากฏออกมา เพื่อแก้ปัญหาเหล่านี้ ในส่วนที่สองจึงจำเป็นต้องผสานพิกัดเฉพาะที่เข้าไว้ด้วยกัน จนเป็นพิกัดรวมซึ่งเป็น

ระบบพิกัดเดียว ดังรูปที่ 9(c) โดยใช้หลักการที่ว่า $\mathbf{x}_j$ บนมิติสูง ต้องถูกส่งผ่านไปยัง $\mathbf{y}_j$ บนมิติต่ำเพียงเวกเตอร์เดียวเท่านั้น ดังนั้นจึงต้องใช้การแปลงสัมพรรคเฉพาะที่ (Local affine mapping) $B_i : \mathbb{R}^k \rightarrow \mathbb{R}^k$ โดยที่ B_i และ B_l จะทำหน้าที่ผสานพิกัดเฉพาะที่ $\mathbf{z}_{ji}$ และ $\mathbf{z}_{jl}$ ให้เป็นพิกัดรวม $\mathbf{y}_j$ ตัวเดียวกัน ซึ่งเขียนเป็นสมการวัตถุประสงค์ได้ดังนี้

$$E = \sum_{j=1}^{p}\sum_{i=1}^{n}\sum_{l=1}^{n} R_{ji}R_{jl}\left\|B_i(\mathbf{z}_{ji}) - B_l(\mathbf{z}_{jl})\right\|^2 \quad (33)$$

เมื่อ R_{ji} และ R_{jl} คือค่า Responsibilities ดังสมการที่ (29) แล้วเวกเตอร์ $\mathbf{y}_j$ สามารถคำนวณได้จากสมการ

$$\mathbf{y}_j = \sum_{i=1}^{n} R_{ji} \cdot B_i(A_i(\mathbf{x}_j)) \quad (34)$$

จะเห็นว่ามีฟังก์ชันส่งผ่านจาก $\mathbf{x}_j$ มิติสูงไปยัง $\mathbf{y}_j$ มิติต่ำแบบชัดเจน (Explicit forward mapping) และเนื่องจาก A_i และ B_i สามารถหาผกผันได้เป็น A_i^{-1} และ B_i^{-1} ตามลำดับ ดังนั้นจึงมีการประมาณค่าฟังก์ชันส่งผ่านแบบผกผัน (Inverse mapping) โดยส่ง $\mathbf{y}_j$ ในมิติต่ำกลับไปยัง $\mathbf{x}_j$ ในมิติสูง ด้วยสมการ

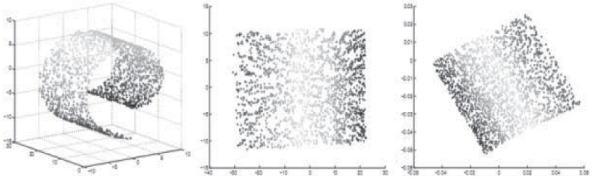
$$\mathbf{x}_j \simeq A_i^{-1} \circ B_i^{-1}(\mathbf{y}_j) \quad (35)$$

การที่มีฟังก์ชันส่งผ่านระหว่างปริภูมิมิติสูงกับปริภูมิมิติต่ำ ถือเป็นจุดเด่นของการลดจำนวนมิติแบบพิกัดรวม ขณะที่การลดจำนวนมิติแบบรักษาระยะทาง ไม่สามารถหาฟังก์ชันส่งผ่านในลักษณะนี้ได้

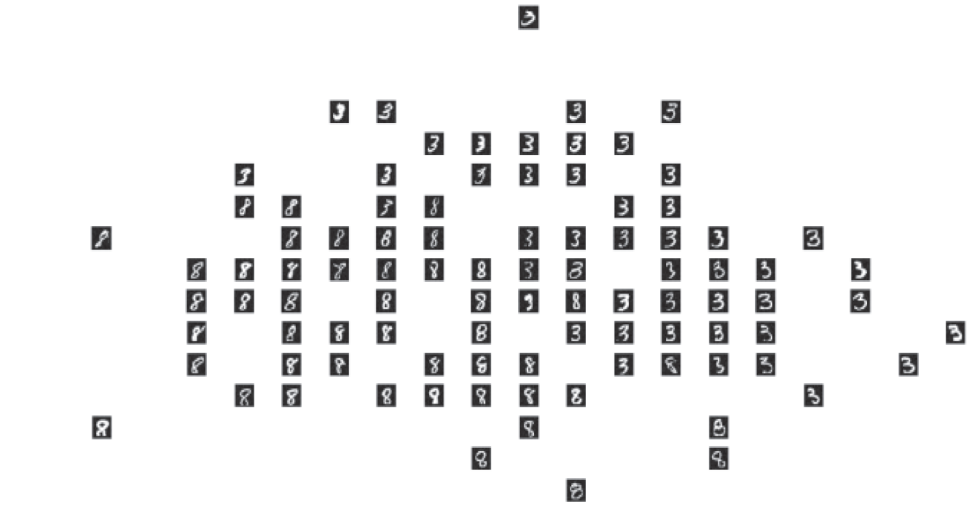
การที่มีฟังก์ชันส่งผ่านระหว่างปริภูมิมิติสูงกับปริภูมิมิติต่ำ ถือเป็นจุดเด่นของการลดจำนวนมิติแบบพิกัดรวม ขณะที่การลดจำนวนมิติแบบรักษาระยะทาง ไม่สามารถหาฟังก์ชันส่งผ่านในลักษณะนี้ได้

5. การสกัดโครงสร้างภายในจากกลุ่มข้อมูล

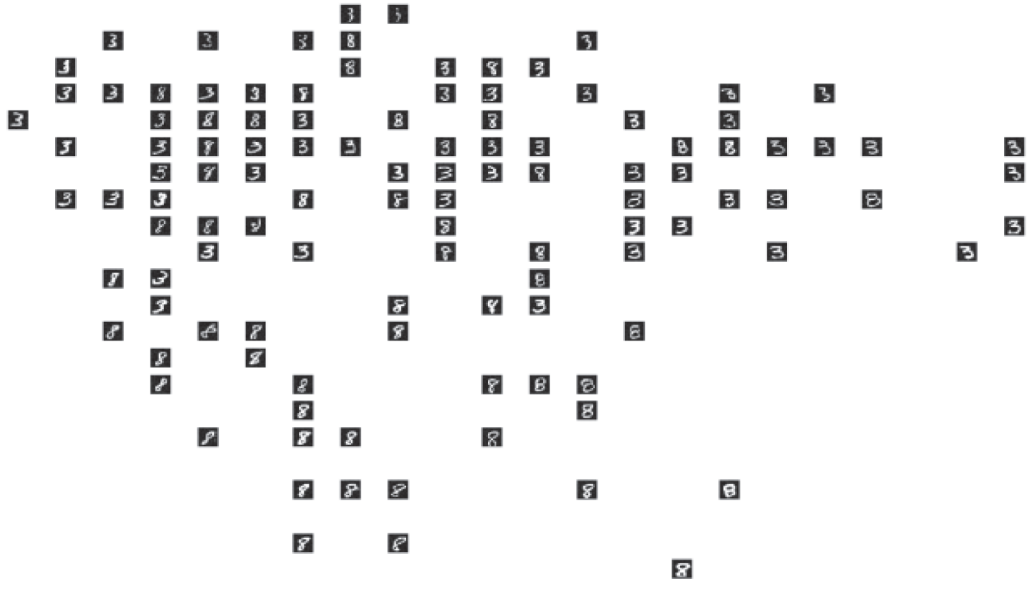
การประมวลผลข้อมูลในปัจจุบันนี้ ต้องเผชิญกับปัญหาที่สำคัญอย่างหนึ่งคือ ข้อมูลมีจำนวนมิติที่สูงมากจนทำให้เครื่องคอมพิวเตอร์ไม่มีหน่วยความจำเพียงพอสำหรับการประมวลผล ประกอบกับข้อมูลที่มีมิติสูงจะแฝงไปด้วยข้อมูลที่ไม่จำเป็น (Redundancy) อยู่เป็นจำนวนมาก ดังนั้นการลดจำนวนมิติจะช่วยให้จัดเก็บข้อมูลให้กระชับมากขึ้น (Compactness) และกำจัดข้อมูลที่ไม่จำเป็นหรือสำคัญน้อยทิ้งไปจนเหลือแต่โครงสร้างภายในของข้อมูล (Internal structure) นอกจากนี้แล้ว ควรจะลดจำนวนมิติลงเหลือ 2 หรือ 3 มิติ เพื่อที่มนุษย์จะมองเห็นโครงสร้างและตีความหมายได้ (Perceivable) และเรียกงานในลักษณะนี้ว่า Data visualization จากรูปที่ 10(a) ข้อมูลต้นแบบอยู่บน 3 มิติ มีการกระจายตัวเป็นแบบไม่เชิงเส้น (Nonlinear) และมีลักษณะคล้ายแผ่นกระดาษม้วน ซึ่งโครงสร้างภายในที่ฝังตัวอยู่ ควรมีลักษณะเป็นแผ่นกระดาษเรียบที่อยู่บนปริภูมิยูคลิด 2 มิติ (2D Euclidean space) รูปที่ 10(b) และ 10(c) แสดงถึงโครงสร้างภายในที่ได้จากวิธี ISOMAP และ Manifold charting ตามลำดับ จะเห็นว่าโครงสร้างภายในสะท้อนถึงโครงสร้างที่แท้จริงของข้อมูล และยังคงรักษาลักษณะทางกายภาพของข้อมูลต้นแบบไว้ได้



รูปที่ 10 แสดงการลดจำนวนมิติของข้อมูลเพื่อหาโครงสร้างที่ฝังตัวอยู่ในชุดข้อมูล (a) ข้อมูลสามมิติที่มีลักษณะคล้ายกระดาษม้วน (b) และ (c) คือโครงสร้างภายในที่เกิดจากการลดจำนวนมิติเหลือสองมิติด้วยวิธี ISOMAP และ Manifold Charting ตามลำดับ



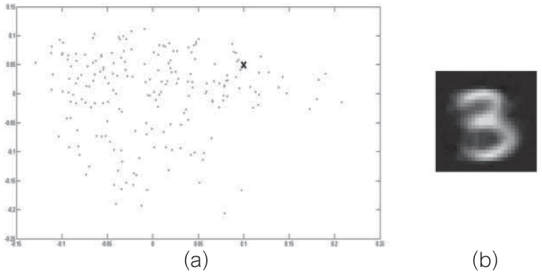
(a)



(b)

รูปที่ 11 ข้อมูลภาพตัวเลข 3 และ 8 ที่สุ่มเลือกมาอย่างละ 100 ภาพ จากฐานข้อมูล MNIST แต่ละภาพจะมีมิติ 28x28 = 784 มิติ ซึ่งมนุษย์ไม่สามารถสังเกตุการกระจายตัวบนปริภูมิ 784 มิติได้ จึงได้ใช้วิธีการลดจำนวนมิติของข้อมูลลงเหลือ 2 มิติ โดยใช้วิธี ISOMAP และ Manifold charting เพื่อหาโครงสร้างภายในที่ฝังตัวอยู่ในกลุ่มข้อมูลชุดนี้ (a) แสดงถึงข้อมูลบนปริภูมิฝังใน 2 มิติ ที่ได้จากวิธี ISOMAP จะเห็นว่าตัวเลขที่มีลักษณะเอียงขวาจะกระจายตัวอยู่บริเวณด้านซ้าย ขณะที่ตัวเลขที่มีลักษณะเอียงซ้ายจะกระจายตัวอยู่บริเวณด้านขวา (b) ผลลัพธ์ของวิธี Manifold charting แต่ละรูปภาพสอดคล้องกับข้อมูลที่ผ่านการลดจำนวนมิติแล้วซึ่งแสดงในรูปที่ 12(a) จะเห็นว่าภาพตัวเลขที่เอียงไปทางขวาอยู่บริเวณฝั่งซ้าย ขณะที่ภาพตัวเลขที่เอียงไปทางซ้าย กระจุกตัวบริเวณฝั่งขวาบน และกลุ่มของตัวเลข 8 จะกระจายตัวเกาะกลุ่มกันตรงบริเวณครึ่งล่าง ดังนั้นการสกัดโครงสร้างภายในจากกลุ่มข้อมูลที่มีมิติสูงมาก จึงเป็นประโยชน์ที่เห็นได้ชัดจากการลดจำนวนมิติ

เมื่อทดสอบกับข้อมูลภาพตัวเลขที่เขียนด้วยลายมือ (Handwritten) ซึ่งเป็นภาพระดับสีเทาจากฐานข้อมูล MNIST ขนาดของภาพคือ 28x28 พิกเซล ดังนั้นมิติของข้อมูลคือ 784 มิติ และทำการสุ่มเลือกภาพตัวเลข 3 และ 8 อย่างละ 100 ภาพ รวมเป็น 200 ภาพ และยังไม่ทราบถึงการกระจายตัวของทั้ง 200 ภาพบนปริภูมิ 784 มิติ และไม่รู้ถึงโครงสร้างภายในของข้อมูลกลุ่มนี้ด้วย แต่เมื่อผ่านการลดจำนวนมิติเหลือเพียง 2 มิติแล้ว โครงสร้างภายในของข้อมูลบนปริภูมิฝังใน 2 มิติจึงได้ปรากฏออกมา ดังแสดงในรูปที่ 11(a) และ 11(b) จากรูปที่ 11(a) จะเห็นว่าภาพตัวเลข 3 จะกระจุกตัวอยู่ฝั่งขวา ในขณะที่ภาพตัวเลข 8 กระจุกตัวทางฝั่งซ้าย และภาพตัวเลขที่มีลักษณะเอียงขวาก็จะเรียงตัวอยู่ฝากซ้าย ขณะที่ภาพตัวเลขที่เอียงซ้ายก็จะเรียงตัวอยู่บริเวณด้านขวาล่าง เช่นเดียวกับรูปที่ 11(b) ภาพตัวเลขที่เอียงไปทางขวาอยู่บริเวณฝั่งซ้าย ขณะที่ภาพตัวเลขที่เอียงไปทางซ้าย กระจุกตัวบริเวณฝั่งขวาบน และบริเวณครึ่งล่างจะมีตัวเลข 8 เกาะกลุ่มกัน จะเห็นว่าการลดจำนวนมิติสามารถนำไปใช้ประโยชน์ในการวิเคราะห์โครงสร้างภายในที่ฝังตัวอยู่ในกลุ่มข้อมูลที่มีจำนวนมิติสูงมาก นอกจากนี้แล้ว ยังได้ทดสอบเกี่ยวกับฟังก์ชันส่งผ่านระหว่างมิติสูงและมิติต่ำ โดยสมมุติเวกเตอร์ $\hat{\mathbf{y}}$ บนมิติต่ำแล้วคำนวณหาภาพ $\hat{\mathbf{x}}$ ที่สมนัยกันบนมิติสูง จากรูปที่ 12(a) แสดงกลุ่มข้อมูลที่ลดจำนวนมิติแล้ว และอยู่บนปริภูมิฝังใน 2 มิติ ซึ่งได้จากการลดจำนวนมิติด้วยหลักการพิกัดรวม โดยใช้วิธี Manifold Charting ตามที่ได้อธิบายแล้วในหัวข้อที่ 4.2 กำหนดให้เวกเตอร์ $\hat{\mathbf{y}}$ อยู่ที่ตำแหน่งกากบาท จากนั้นใช้ฟังก์ชันส่งผ่านแบบผกผันตามสมการที่ (35) เพื่อแปลงเวกเตอร์ $\hat{\mathbf{y}}$ 2 มิติ กลับไปเป็นเวกเตอร์ $\hat{\mathbf{x}}$ 784 มิติ แล้วเปลี่ยนให้เป็นภาพขนาด 28x28 พิกเซล ดังรูปที่ 12(b) ทั้งนี้ $\hat{\mathbf{x}}$ เป็นภาพที่ถูกสร้างขึ้นมาใหม่ ดังนั้นการมีโครงสร้างภายในของกลุ่มข้อมูล และมีฟังก์ชันส่งผ่านระหว่างมิติสูงและมิติต่ำจึงมีความสำคัญมากต่อการสร้างข้อมูลชุดใหม่ด้วย



รูปที่ 12 ข้อมูลกลุ่มตัวเลข 3 และ 8 รวม 200 ภาพ ได้ถูกลดจำนวนมิติจาก 784 มิติเหลือเพียง 2 มิติ โดยใช้วิธี Manifold Charting (a) จุดข้อมูลแสดงถึงการกระจายตัวของข้อมูล 200 ภาพ บนปริภูมิฝังใน 2 มิติ และภาพที่สอดคล้องกับแต่ละจุดข้อมูลได้แสดงแล้วในรูปที่ 11(b) ตำแหน่งกากบาทคือจุดข้อมูลใหม่ $\hat{\mathbf{y}}$ ที่ต้องการสร้างขึ้นมา เมื่อใช้ฟังก์ชันส่งผ่านแบบผกผัน ดังสมการที่ (35) ส่ง $\hat{\mathbf{y}} \in \mathbb{R}^2$ กลับไปยัง $\hat{\mathbf{x}} \in \mathbb{R}^{784}$ แล้วแปลงให้เป็นรูปภาพขนาด 28x28 พิกเซล จะได้รูป 12(b) จึงถือเป็นการสร้างข้อมูลภาพใหม่บนมิติสูง โดยใช้การกำหนดข้อมูลใหม่บนมิติต่ำ

6. สรุป

บทความนี้ได้แนะนำโครงข่ายประสาทเทียมแบบไม่มีผู้สอนเพื่อค้นหาโครงสร้างที่ฝังตัวอยู่ในกลุ่มข้อมูล โครงข่ายประสาทเทียมแบบไม่มีผู้สอนสามารถถูกแบ่งได้เป็นสองแบบตามลักษณะของระเบียบวิธีคือวิธีการแบ่งกลุ่มข้อมูล และวิธีการลดจำนวนมิติของข้อมูล วิธีการแบ่งกลุ่มจะให้โครงสร้างข้อมูลในลักษณะกลุ่มของโปรโทไทพ์ที่ใช้เป็นตัวแทนของข้อมูล และสามารถประยุกต์ใช้กับปัญหาการแบ่งพื้นที่ในภาพได้ ในส่วนของวิธีการลดจำนวนมิติของข้อมูล จะให้โครงสร้างข้อมูลในลักษณะ โครงสร้างภายในที่ฝังตัวอยู่ในกลุ่มข้อมูลที่มีจำนวนมิติสูงมาก และสามารถประยุกต์ใช้กับงานด้าน Data visualization และช่วยในการสร้างข้อมูลใหม่อีกด้วย

7. กิตติกรรมประกาศ

เนื้อหาในส่วนการแบ่งพื้นที่ในภาพ ได้รับทุนส่งเสริมการวิจัย จากกองทุนพัฒนาและส่งเสริมด้านวิชาการของคณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น พ.ศ. 2554 ประเภททุนนักวิจัยหน้าใหม่ ในหัวข้อ Lmage segmentation using matrix neural maps

8. เอกสารอ้างอิง

[1] Kohonen T. The self-organizing maps. Proceedings of the IEEE. 1990; 78(9): 1464-1480.

[2] Cottrell M, Hammer B, Hasenfuss A, Villmann T. Batch and median neural gas. Neural Networks. 2006; 19: 762-771.

[3] Martinetz T, Berkovich S, Schulten K. Neural gas network for vector quantization and its application to time series prediction. IEEE Transactions on Neural Networks. 1993; 4(4): 558-569.

[4] Kusumoto H. $O(\log_2 M)$ Self-Organizing Map Algorithm Without Learning of Neighborhood Vectors. IEEE Transactions on Neural Networks. 2006; 17(6): 1656-1661.

[5] Arnonkijpanich B, Lursinsap C. Adaptive Second Order Self-Organizing Mapping for 2D Pattern Representation. IEEE International Joint Conference on Neural Networks. 2004; 1: 775-780.

[6] Mochihashi D, Kikui G, Kita K. Learning an optimal distance metric in a linguistic vectorspace. Systems and computers in Japan. 2006; 37(9): 12-21.

[7] Arnonkijpanich B, Hasenfuss A, Hammer B. Local Matrix Adaptation in Topographic Neural Maps. Neurocomputing. 2011; 74(4): 522-539.

[8] Tenenbaum J, Silva V, Langford J. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science. 2000; 290(5500): 2319-2323.

[9] Sammon J. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers. 1969; 18: 401-409.

[10] Dermartines P, Herault J. Curvilinear Component Analysis: A Self Organizing Neural Network for Nonlinear Mapping. IEEE Transactions on Neural Networks. 1997; 8(1): 148-154.

[11] Brand M. Charting a manifold. Advances in Neural Information Processing Systems. 2003; 15: 961-968.

[12] Arnonkijpanich B, Hasenfuss A, Hammer B. Local Matrix Learning in Clustering and Applications for Manifold Visualization. Neural Networks. 2010; 23(4): 476-486.

[13] Tipping M, Bishop C. Mixtures of probabilistic principal component analyzers. Neural Computation. 1999; 11: 443-482.