

การเรียนรู้โครงสร้างเชิงสาเหตุในข้อมูลการทำงานของสมอง
Learning Causal Structures of Brain Connectivity in fMRI time series
นายอานันท์ พงษ์รัตนกุล อาจารย์ที่ปรึกษา อ.ดร.จิตโกมุท ส่งศิริ
กลุ่มวิจัย การควบคุมและการหาค่าเหมาะสมที่สุดขั้นสูง
ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2555

บทคัดย่อ

โครงงานฉบับนี้ นำเสนอวิธีการเรียนรู้โครงสร้างเชิงสาเหตุในข้อมูลการทำงานของสมองที่วัดได้จากระดับออกซิเจนในเลือดที่จุดต่างๆในสมอง ปริมาณออกซิเจนในแต่ละจุดนี้จะสะท้อนถึงปริมาณการใช้งานของสมองในบริเวณนั้นๆ การเรียนรู้โครงสร้างเชิงสาเหตุ คือการหาความสัมพันธ์ของการทำงานของสมองที่เกิดขึ้นระหว่างสองจุดใดๆ เป้าหมายของเราคือการสร้างแบบจำลองเชิงกราฟที่ใช้อธิบายความสัมพันธ์นี้ ความสัมพันธ์ของการทำงานของสมองที่เกิดขึ้นระหว่างสองจุดใดๆ สามารถนิยามเป็นเงื่อนไขทางคณิตศาสตร์ได้โดยหลักความเป็นเหตุเป็นผลแบบเกรนเจอร์ เราจะสร้างแบบจำลองถดถอยตัวเองหลายตัวแปร ที่สามารถอธิบายการทำงานของสมอง และอธิบายความสัมพันธ์แบบเกรนเจอร์ได้ผ่านทางรูปแบบของศูนย์ในพารามิเตอร์ของแบบจำลอง และแสดงออกมาเป็นทอพอโลยีของกราฟ ปัญหาการประมาณแบบจำลองดังกล่าวจะถูกจัดรูปเป็นปัญหาการถดถอยแบบกลุ่มแลซโซ่ ซึ่งเป็นปัญหาค่าเหมาะที่สุดแบบคอนเวกซ์ จากนั้นเราประยุกต์ขั้นตอนวิธี ADMM (Alternating Direction Method of Multipliers) ซึ่งเป็นวิธีเชิงเลขที่เหมาะสมกับปัญหาที่มีตัวแปรจำนวนมากอย่างเช่นปัญหานี้ เราเสนอวิธีการเลือกแบบจำลองนั้นคือการเลือกรูปแบบของศูนย์ในพารามิเตอร์หรือทอพอโลยี และทดสอบความถูกต้องของวิธีที่เสนอกับข้อมูลชุดที่สังเคราะห์ขึ้นเอง จากนั้นประยุกต์วิธีดังกล่าวกับชุดข้อมูลจริงที่เป็นภาพสแกนสมองจากเครื่อง fMRI (functional Magnetic Resonance Imaging)

คำสำคัญ : fMRI, ปัญหาค่าเหมาะที่สุดแบบคอนเวกซ์, ความสัมพันธ์ระหว่างเหตุและผลแบบเกรนเจอร์, การถดถอยแบบกลุ่มแลซโซ่ , การเลือกแบบจำลอง

Abstract

This paper presents a method for finding directional relationships between any pair of variables in the brain obtained by blood oxygen-level dependent (BOLD) which reflects an activity level at that node. Learning a causal structure is to find the dynamics of any pair in the brain. Our goal is to construct a graphical model that describes these relationships. The dynamics of any pair in the brain can be defined by the Granger causality principle. We construct a multivariate autoregressive (MAR) model that describes the functional activities in the brain and explains Granger causality relationships via a sparsity pattern of MAR parameter or topology. The problem of estimating MAR models is formulated as a group lasso regression which falls into a convex optimization problem. Then, we apply a numerical technique called ADMM (the alternating direction method of multipliers) which is suitable for large-scale problems. We develop our method for selecting the best topology from a candidate set (see section model selection). We verify the effectiveness of this approach on a synthetic data set and then apply the method to fMRI time series.

Keywords : fMRI, Convex optimization, Granger causality, Group Lasso regression, Model selection

รายงานฉบับสมบูรณ์ วิชาโครงงานวิศวกรรมไฟฟ้า ปีการศึกษา 2555

การเรียนรู้โครงสร้างเชิงสาเหตุในข้อมูลการทำงานของสมอง

Learning Causal Structures of Brain Connectivity in fMRI time series

นายอานันท์ พงษ์รัตนกุล 5230619021 กลุ่มวิจัย การควบคุมและการหาค่าเหมาะสมที่สุดขั้นสูง

อาจารย์ที่ปรึกษา อ.ดร.จิตโกมุท สงศิริ ห้องปฏิบัติการวิจัยระบบควบคุม

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

1 บทนำ

โครงงานฉบับนี้จะหาแบบจำลองที่สามารถอธิบายพลวัตของการทำงานในสมองได้ผ่านข้อมูลของระดับออกซิเจนในเลือดในสมอง แบบจำลองที่เลือกใช้คือ แบบจำลองถดถอยด้วยตัวเองหลายตัวแปร (Multivariate Autoregressive Model หรือ MAR) โดยแบบจำลอง MAR ที่มี n ตัวแปร อันดับ p สามารถแสดงได้ดังนี้

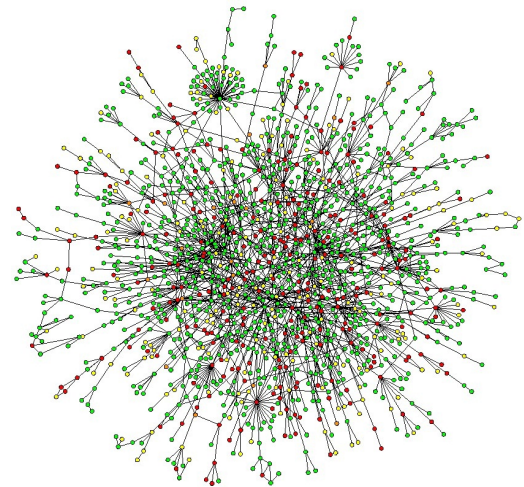
$$y(t) = A_1y(t-1) + A_2y(t-2) + \dots + A_p y(t-p) + u(t) \quad (1)$$

โดยที่ $y \in \mathbb{R}^n$, $A_k \in \mathbb{R}^{n \times n}$, $k = 1, 2, \dots, p$ และ $u(t)$ คือสัญญาณรบกวน ค่าในแต่ละ y_i ในเวกเตอร์หมายถึง ระดับออกซิเจนที่วัดได้จากจุดที่ i ในสมอง (ซึ่งเราเรียกแต่ละจุดนี้ว่า voxel) โดยข้อมูลสมองที่ใช้ในโครงงานนี้มีทั้งหมด 6004 ตัว ($n = 6004$)

ในโครงงานนี้ปัญหาการประมาณแบบจำลอง MAR สำหรับข้อมูลอนุกรมเวลานั้นเป็นปัญหาพื้นฐาน แต่เป้าหมายของเราคือการนำแบบจำลองที่ได้มาแสดงความสัมพันธ์ของตัวแปรในระบบที่ซับซ้อนมาก ผ่านทางแบบจำลองเชิงกราฟ (graphical model) ดังเช่นรูปที่ 1

แบบจำลองเชิงกราฟคือแบบจำลองที่อธิบายความสัมพันธ์ระหว่างตัวแปรที่เราสนใจผ่านทางรูปแบบการเชื่อมโยงของตัวแปร หรือทอพอโลยี (topology) ของกราฟ แบบจำลองเชิงกราฟประกอบไปด้วย เซตของโหนดจำนวน n โหนดซึ่งแทนตัวแปร y_1, y_2, \dots, y_n และเซตของลิงก์ (v) ที่มีสมาชิกเป็นคู่ลำดับ (i, j) ถ้าหากมีลิงก์จากโหนด j ไปยังโหนด i การทราบว่ามีลิงก์เชื่อมระหว่างสอง

โหนดใดๆ จึงเป็นวิธีหนึ่งที่ใช้แสดงความสัมพันธ์ระหว่าง y_i และ y_j คู่หนึ่ง



รูปที่ 1: แบบจำลองเชิงกราฟ : อ้างอิงจาก <http://www.stanford.edu/~montanar/TEACHING/Stat375/stat375.html>

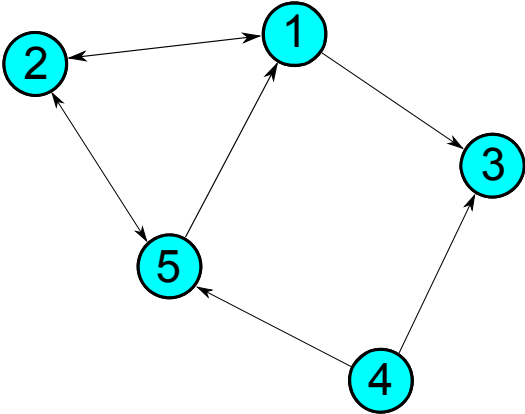
การนิยามการมีลิงก์ระหว่างสองโหนดใดๆ หรือการกล่าวถึงความสัมพันธ์ระหว่างสองตัวแปรใดๆ ในระบบสามารถนิยามเป็นเงื่อนไขทางคณิตศาสตร์ได้หลายแบบ ในโครงงานนี้ เราอาศัยเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล (Granger causality หรือ G-causality) [1, บทที่ 2] โดยหลักการพื้นฐานของ G-causality คือ เหตุห้ามเกิดหลังผลลัพธ์ ถ้าหากตัวแปร x มีผลต่อตัวแปร z แล้ว การที่รู้ค่าตัวแปร x ก่อน ย่อมช่วยในการคาดเดาผลของตัวแปร z ที่จะเกิดขึ้นได้ และใน [1, บทที่ 2] ได้แสดงให้เห็นว่าเงื่อนไข G-causality ของแบบจำลอง MAR นั้นเขียนได้ในรูปสมการเชิงเส้นของสัมประสิทธิ์ใน MAR ซึ่ง

เป็นเงื่อนไขที่ไม่ซับซ้อน กล่าวคือ ถ้า y_j ไม่ใช่สาเหตุแบบ Granger ต่อ y_i แล้ว

$$[A_k]_{ij} = 0 \quad (2)$$

ทุกค่า $k = 1, 2, \dots, p$ โดยที่ $[A_k]_{ij}$ คือ สมาชิกตัวที่ (i, j) ของเมทริกซ์ A_k และเราจะกำหนดให้ทุกคู่อันดับ (i, j) ของเมทริกซ์ A_k ซึ่งมีบางค่า k ที่ $[A_k]_{ij} \neq 0$ เป็นสมาชิกในเซต ν

เนื่องจากเงื่อนไข G-causality เป็นเงื่อนไขที่มีทิศทาง ดังนั้นเซต ν จึงเป็นเซตของลิงก์แบบมีทิศทาง โดยแบบจำลองเชิงกราฟแบบมีทิศทางจะมีลักษณะดังตัวอย่างในรูปที่ 2



รูปที่ 2: แบบจำลองเชิงกราฟแบบมีทิศทางของระบบที่มี 5 ตัวแปร โดยแต่ละโหนดหมายถึงแต่ละตัวแปรที่เราสนใจ และลูกศรหมายถึงลักษณะความสัมพันธ์ เช่น โหนดที่ 5 เป็นสาเหตุของ โหนดที่ 1 ส่วนโหนดที่ 2 กับโหนดที่ 3 ไม่เป็นสาเหตุของกันและกัน

การทำแบบจำลอง fMRI เป็นงานที่เกี่ยวข้องกับอนุกรมเวลา ดังนั้น MAR จึงเป็นแบบจำลองที่ถูกนิยมนำมาใช้ในงานวิจัยอย่างมาก เช่นใน [2, 3, 4, 5, 6] ซึ่งทั้ง 5 งานวิจัยต่างใช้ MAR ที่คำนึงถึงเงื่อนไข G-causality โดยจัดรูปแบบเป็นปัญหาค่าเหมาะที่สุดที่เพิ่มฟังก์ชันลงโทษ (penalty function) ในฟังก์ชันจุดประสงค์ การเลือกฟังก์ชันลงโทษจะมีจุดหมายเพื่อให้ได้ผลเฉลยที่ทำให้พารามิเตอร์ใน A_k มีศูนย์เป็นจำนวนมาก ดังนั้นจาก (2) รูปแบบของศูนย์ใน A_k จะบอกรูปแบบความ

สัมพันธ์ของตัวแปรว่าตัวแปรใดส่งผลต่อตัวแปรใดบ้าง โดยแต่ละงานวิจัยจะใช้ฟังก์ชันลงโทษที่ต่างกัน นอกจากนี้แต่ละงานวิจัยจะจัดรูปแบบปัญหาและวิเคราะห์ด้วยวิธีที่ต่างกัน เช่น [2, 4] จะวิเคราะห์ในโดเมนความถี่ ส่วน [3, 5, 6] จะวิเคราะห์ในโดเมนเวลา จากนั้นแต่ละงานวิจัยจะใช้โปรแกรมสำเร็จรูปในการหาผลเฉลยของปัญหาค่าเหมาะที่สุด หลังจากได้แบบจำลองแล้วจะนำแบบจำลองไปประยุกต์ใช้เช่น [2] นำแบบจำลองที่ได้ไปใช้ในการคัดแยกผู้ป่วย ส่วน [4] ใช้ในการเปรียบเทียบระหว่างบุคคลที่ร่างกายเหนื่อยล้ากับบุคคลที่ได้พักผ่อนอย่างเต็มที่

2 ระบบวิธีที่นำเสนอในโครงการ

2.1 ปัญหาการประมาณแบบจำลอง MAR

ในการประมาณพารามิเตอร์ในแบบจำลอง (MAR) นั้น วิธีพื้นฐานที่นิยมใช้คือ วิธีกำลังสองต่ำสุด ซึ่งจะเป็ปัญหาการเลือกค่า A_1, A_2, \dots, A_p ที่ทำให้

$$\sum_{t=p+1}^N \|y(t) - \sum_{k=1}^p A_k y(t-k)\|_2^2 \quad (3)$$

มีค่าต่ำที่สุด โดย $y(1), y(2), \dots, y(N)$ เป็นข้อมูลที่กำหนดให้ และเราอาจเขียนปัญหาดังกล่าวได้ใหม่เป็น

$$\text{minimize } (1/2) \|Y - AH\|_F^2 \quad (4)$$

โดยมีตัวแปรคือ $A_k \in \mathbb{R}^{n \times n}$ สำหรับ $k = 1, 2, \dots, p$ และ

$$A = \begin{bmatrix} A_1 & A_2 & \dots & A_p \end{bmatrix}$$

$$Y = [y(p+1) \ y(p+2) \ \dots \ y(N)]$$

$$H = \begin{bmatrix} y(p) & y(p+1) & \dots & y(N-1) \\ y(p-1) & y(p) & \dots & y(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ y(1) & y(2) & \dots & y(N-p) \end{bmatrix}$$

แต่ปัญหาการทำแบบจำลองที่เราสนใจจะคำนึงถึงรูปแบบความสัมพันธ์ของ G-causality ดังนั้นปัญหาการประมาณแบบจำลอง MAR ที่จะนำรูปแบบศูนย์ร่วม

ของพารามิเตอร์ไปแสดงเป็นแบบจำลองเชิงกราฟ ที่เราพิจารณาในโครงงานนี้จะแบ่งออกเป็น 2 ปัญหาดังนี้

ปัญหาการประมาณที่ทราบทอพอโลยีของกราฟ

ในกรณีที่ทราบทอพอโลยี หรือในที่นี้คือ *รูปแบบร่วม* ของศูนย์ในทุกๆ A_k (common sparsity pattern) เราจะเขียนปัญหาดังกล่าวเป็นปัญหาค่าเหมาะที่สุดแบบมีเงื่อนไขได้ดังนี้

$$\begin{aligned} & \underset{A}{\text{minimize}} \quad (1/2) \|Y - AH\|_F^2 \\ & \text{subject to} \quad (A_1)_{ij} = \dots = (A_p)_{ij} = 0, (i, j) \notin \mathcal{V} \end{aligned} \quad (5)$$

โดย $A = \begin{bmatrix} A_1 & A_2 & \dots & A_p \end{bmatrix} \in \mathbb{R}^{n \times np}$ และ \mathcal{V} เป็นเซตของลิงก์ในกราฟ (เซตของคู่อันดับ (i, j) ใน A_k ที่ไม่เป็นศูนย์)

ในบางปัญหาที่สามารถทราบทอพอโลยีของกราฟนั้น เราจะพบว่าจำนวนพารามิเตอร์ในแบบจำลองที่ใช้ได้ลดลงไป นอกจากนี้ค่าความแปรปรวนของค่าประมาณของ A ก็จะมีค่า **ลดลง** เนื่องจาก $(A_k)_{ij}$ บาง (i, j) ถูกกำหนดให้มีค่าเป็นศูนย์ (ซึ่งเป็นค่าที่แน่นอน)

ปัญหาการประมาณเพื่อเรียนรู้ทอพอโลยีของกราฟ

ในการเรียนรู้โครงสร้างเชิงสาเหตุจากข้อมูล โดยทั่วไปแล้วเราจะไม่ทราบรูปแบบร่วมของศูนย์ แต่เราสามารถอ่านรูปแบบร่วมของศูนย์ (common sparsity pattern) ใน A_1, A_2, \dots, A_p ที่ประมาณได้เพื่อนำไปสรุปโครงสร้างเชิงสาเหตุได้ ดังนั้นหากเราออกแบบปัญหาค่าเหมาะที่สุดเพื่อให้มีผลเฉลยเมทริกซ์ A เป็นเมทริกซ์เบาบาง (เมทริกซ์ ซึ่งมีศูนย์เป็นจำนวนมาก หรือ sparse matrix) แล้ว แบบจำลองที่ได้จะสามารถอธิบายโครงสร้างเชิงสาเหตุได้ ด้วยเหตุนี้เราจะพัฒนาปัญหาค่าเหมาะที่สุดไปในรูปแบบดังนี้

$$\underset{A}{\text{minimize}} \quad (1/2) \|Y - AH\|_F^2 + \gamma g(A) \quad (6)$$

โดย $g(A)$ คือฟังก์ชันลงโทษในตัวแปร $A = \begin{bmatrix} A_1 & A_2 & \dots & A_p \end{bmatrix}$ ที่จะทำให้ได้ผลเฉลยที่สมาชิกภายใน A_k มีศูนย์เป็นจำนวนมาก และ $\gamma > 0$ เป็นพารามิเตอร์ของฟังก์ชันลงโทษ ซึ่งจะเป็นตัวกำหนดว่าเราจะให้น้ำหนักกับฟังก์ชันลงโทษมากเท่าใด

การตีความโครงสร้างเชิงสาเหตุของแบบจำลองที่ได้ตาม (2) นั้นต้องอ่านจาก *รูปแบบร่วม* ของศูนย์ในทุกๆ A_k ด้วยเหตุนี้เราจึงพิจารณา

$$g(A) = \sum_{i \neq j} \left\| \begin{bmatrix} (A_1)_{ij} & (A_2)_{ij} & \dots & (A_p)_{ij} \end{bmatrix} \right\|_2 \quad (7)$$

การใช้ผลบวกของค่าที่ไม่เป็นลบสำหรับทุกๆ (i, j) นั้นเหมือนกับเป็นการใช้ norms-1 ดังนั้นเมื่อ $g(A)$ มีค่าน้อยระดับหนึ่ง แล้วจะมีบางตำแหน่ง (i, j) ที่ทำให้ $\left\| \begin{bmatrix} (A_1)_{ij} & (A_2)_{ij} & \dots & (A_p)_{ij} \end{bmatrix} \right\|_2$ เป็นศูนย์ ซึ่งจะได้ว่าที่ตำแหน่ง (i, j) นี้ $(A_k)_{ij}$ จะเป็นศูนย์ทุกๆ k หากเราเลือกค่า γ ให้มีค่ามาก จะได้ว่าเราลงโทษ $g(A)$ ให้มีค่าน้อยลง นั่นคือจะทำให้ A_k มีศูนย์มากขึ้น กราฟที่ประมาณได้ก็จะเป็นกราฟที่เบาบางลง (sparse graph) ในขณะเดียวกัน ถ้าเลือก γ ให้มีค่าน้อย เราจะได้กราฟที่หนาแน่นขึ้น (dense graph) ดังนั้นการเลือกค่า γ ที่เหมาะสมจึงเป็นหัวข้อที่สำคัญดังจะกล่าวถึงต่อไปในส่วนที่ 2.3

หลังจากทราบ *รูปแบบร่วม* ของศูนย์ในทุกๆ A_k แล้ว เราจะนำรูปแบบร่วมที่ได้เป็นเงื่อนไขบังคับในปัญหาค่าเหมาะที่สุดแบบมีเงื่อนไขตาม (5) เพื่อหาแบบจำลองที่ดีที่สุดที่สอดคล้อง *รูปแบบร่วม* ของศูนย์ในทุกๆ A_k นั้น

2.2 การลดจำนวนตัวแปร

ข้อมูลที่วัดได้จาก fMRI จะมีมิติขึ้นอยู่กับจำนวน voxel โดยทั่วไปจะมีจำนวนมาก (ระดับหลายๆ พัน) ดัง

นั่นในทางปฏิบัติเราจะทำการลดมิติของข้อมูลลง โดยการจับกลุ่ม voxel ออกเป็นส่วนย่อยๆ (Region of interest หรือ ROI) ในกรณีที่มีความรู้ทางด้านประสาทวิทยา เราอาจใช้ความรู้ทางด้านประสาทวิทยาช่วยในการแบ่ง ROI ออกเป็นส่วนๆได้ เช่นใน [4] แต่ในโครงการนี้เราจะใช้วิธี Principal Components Analysis (PCA) [7]

หลักการของ PCA [7] คือ การแปลงเชิงเส้นของตัวแปร x จากมิติเดิมไปเป็นตัวแปร y ในมิติใหม่ที่มีมิติน้อยลงโดยเลือกการแปลงที่ทำให้ความแปรปรวนของตัวแปรสูญหายน้อยที่สุด

ถ้าให้ตัวแปร $x \in R^q$ ที่มีเมทริกซ์ความแปรปรวนเป็น C_x จะได้ว่า C_x สามารถเขียนได้ในรูป $C_x = \sum_{k=1}^q \lambda_k \phi_k \phi_k^T$ โดยที่ (λ_k, ϕ_k) คือคู่ค่าเฉพาะ และเวกเตอร์เฉพาะของ C_x และสมมติให้ $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$ ถ้าให้ n (ซึ่ง $n < q$) เป็นมิติของข้อมูลชุดใหม่ จะเห็นได้ว่าการแปลงเชิงเส้น

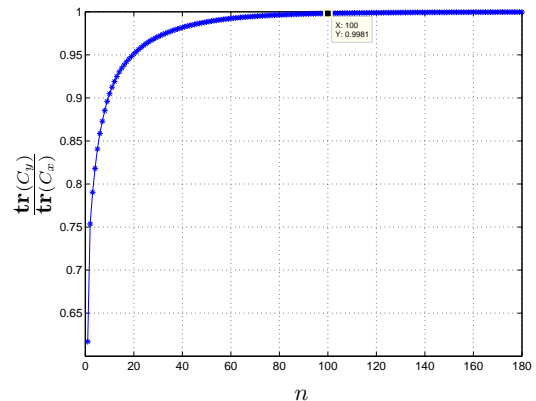
$$y = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_n \end{bmatrix}^T (x - \bar{x}) \quad (8)$$

โดยที่ \bar{x} คือเวกเตอร์ค่าเฉลี่ยของ x จะทำให้ $y \in R^n$ และมีเมทริกซ์ความแปรปรวนเป็น $C_y = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ [8]

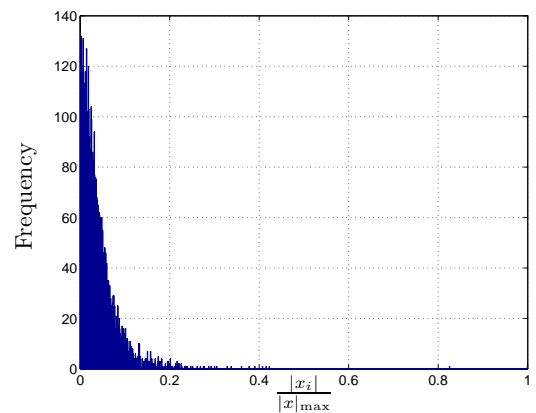
การเลือกขนาดของมิติ n ตามหลักของ PCA คือเราจะเลือกค่าที่ทำให้ $\text{tr}(C_y)$ กับ $\text{tr}(C_x)$ ไม่แตกต่างกันมากเกินไปดังในรูปที่ 3

เราสามารถมองการทำ PCA ได้ว่าเป็นการหาค่าเฉลี่ยแบบถ่วงน้ำหนักในแต่ละกลุ่ม ROI โดยแต่ละค่าในเวกเตอร์เฉพาะจะเป็นน้ำหนักที่ใช้กับแต่ละค่าในเวกเตอร์ x อีกทั้งหากพิจารณารูปที่ 4 จะเห็นได้ว่าการกระจายตัวของแต่ละค่าในเวกเตอร์เฉพาะจะมีลักษณะเบาบาง (มีเพียงส่วนน้อยที่ถูกถ่วงน้ำหนักด้วยค่าที่มาก) ดังนั้นเราอาจตีความว่าการทำ PCA เป็นการแบ่งกลุ่มข้อมูล ROI ออก

เป็นส่วนๆ โดยการใช้การหาค่าเฉลี่ยแบบถ่วงน้ำหนักแทนการหาค่าเฉลี่ยแบบปกติ



รูปที่ 3: กราฟแสดงอัตราส่วนของรอยของความแปรปรวน $\frac{\text{tr}(C_y)}{\text{tr}(C_x)}$ กับ n (จำนวนเวกเตอร์เฉพาะของ C_x ที่ใช้ในการแปลงเชิงเส้นของ x) จากรูปจะเห็นได้ว่าเมื่อเพิ่มจำนวนค่าเวกเตอร์เฉพาะในการแปลงเชิงเส้นถึงจุดหนึ่ง การเพิ่มดังกล่าวไม่ช่วยทำให้อัตราส่วนเปลี่ยนแปลงมากนัก ในที่นี้เราเลือกที่จะใช้ $n = 100$ ทำให้ลดจำนวนมิติลงจากข้อมูลเดิม $x \in R^{6004}$ ไปเป็น $y \in R^{100}$



รูปที่ 4: ฮิสโทแกรมของ $\frac{|\phi_7|}{|\phi_7|_{\max}}$ เราจะเห็นได้ว่ามีเพียงส่วนน้อยเท่านั้นที่มีค่าเกิน 10% ของค่าที่มีขนาดสูงสุดในเวกเตอร์เฉพาะนั้น

2.3 การเลือกแบบจำลอง

ในกรณีทั่วไปของปัญหาการเรียนรู้ของทอพอโลยีของกราฟที่มี n โหนดนั้น เราจะได้ว่ามีจำนวนทอพอโลยีที่เป็นไปได้ทั้งหมด 2^{n^2-n} แบบ ซึ่งเป็นจำนวนที่เยอะมาก (ที่ $n = 20$ มีทอพอโลยีทั้งหมด 2^{380} แบบ !) ดังนั้นโครงการนี้เราจะเลือกแบบจำลองมาเพียงกลุ่มเล็กๆ ใน

การพิจารณา โดยการเปลี่ยนค่าพารามิเตอร์ของฟังก์ชัน
 ลงโทษ γ ในปัญหา (6) ไปทั้งหมด M ค่า ซึ่งสอดคล้องกับ
 ทอพอโลยีของกราฟทั้งหมด M แบบ โดยที่ $M \ll 2^{n^2-n}$
 ทอพอโลยีที่ประมาณได้จะมาจากการแก้ปัญหา (6) ที่ค่า
 พารามิเตอร์ γ และ p ค่าหนึ่ง โดยสรุปเป็นขั้นตอนได้ดัง
 ต่อไปนี้

ขั้นตอนการเรียนรู้ทอพอโลยีของกราฟ

เราจะกำหนด γ และ p ค่าหนึ่ง สำหรับปัญหา (6)
 โดยปัญหา (6) จะให้ผลเฉลยที่สอดคล้องทอพอโลยีหนึ่ง
 แบบ จากนั้นนำทอพอโลยีที่ประมาณได้มาใช้เป็นเงื่อนไข
 ในปัญหา (5) ที่มีผลเฉลยคือ \hat{A}

ในโครงการนี้เราจะใช้ค่า p ตั้งแต่ 1 ถึง 5 และค่า
 γ ทั้งหมด M ค่าจากช่วง $(0, \gamma_{\max})$ โดยค่า γ_{\max} คือค่า
 ต่ำที่สุดที่ใช้ใน (6) ที่ทำให้ $\hat{A}_1, \hat{A}_2, \dots, \hat{A}_p$ เป็นเมทริกซ์
 ทแยงมุม [9] ดังนั้นเราจะได้ว่ามีทอพอโลยีทั้งหมด $5M$
 แบบ ปัญหาถัดมาคือ เราจะเลือกทอพอโลยีอย่างไรให้ได้
 ผลลัพธ์ที่สอดคล้องกับข้อมูลที่มีมากที่สุด เราจะพิจารณา
 เกณฑ์การเลือกแบบจำลอง 2 วิธีดังต่อไปนี้

Bayesian Information Criterion (BIC) [10]

วิธีนี้เป็นหนึ่งในเกณฑ์การเลือกแบบจำลองที่นิยมใช้
 เช่นใน [5, 11, 12] หลักการของ BIC คือเราจะเลือกแบบ
 จำลองที่ให้ค่า BIC score ต่ำที่สุด นิยามโดย

$$\text{BIC score} = -2\mathcal{L} + d \log(N) \quad (9)$$

โดย \mathcal{L} คือ loglikelihood function (ฟังก์ชัน
 ลอการิทึมความเป็นไปได้), d คือจำนวนตัวแปรอิสระใน
 แบบจำลอง และ N คือจำนวนข้อมูล

ใน (1) หากตั้งสมมติฐานให้ $u(t)$ มีการแจกแจงแบบ
 ปกติแล้ว [10] จะได้ว่า

$$\mathcal{L} = -(N/2) \log \det (R(\theta))$$

โดยที่ $R(\theta) = \sum_{t=1}^N e(t, \theta)e(t, \theta)^T$ เป็นเมทริกซ์ความ
 แปรปรวนของตัวอย่าง โดย $e(t, \theta)$ คือ ค่าความคลาด
 เคลื่อนระหว่างค่าที่ประมาณได้จากแบบจำลองกับข้อมูล
 ค่าจริงที่เวลา t และ $e(t, \theta) \in \mathbf{R}^n$ โดยที่ n คือจำนวน
 voxel ในสมอง

จากปัญหา (4) เราสามารถแสดงได้ว่า

$$R(\theta) = (1/N)(Y - \hat{A}H)(Y - \hat{A}H)^T$$

โดย \hat{A} คือพารามิเตอร์ของแบบจำลองที่ได้จากการ
 ประมาณ และ d คือจำนวนตัวแปรที่ไม่เป็นศูนย์ของ \hat{A}

ดังนั้นจากแบบจำลองที่มีพารามิเตอร์ \hat{A} ทั้งหมด $5M$
 แบบ เราจะเลือกแบบจำลองที่ทำให้ BIC ใน (9) มีค่าต่ำ
 สุด

Validation Method

วิธีนี้เป็นที่นิยมใช้ในสายการเรียนรู้ด้วยเครื่องจักร
 (Machine Learning) [7] ในการเลือกแบบจำลอง หลัก
 การของ validation คือการแบ่งข้อมูลออกเป็น 2 ส่วน
 เป็นชุดที่ใช้สำหรับประมาณค่าแบบจำลอง (training set)
 และชุดที่ใช้ในการคัดเลือกแบบจำลอง (validation set)
 ในอัตราส่วน 2 ต่อ 1 ของข้อมูลทั้งหมดที่มี

ในโครงการนี้ training set จะใช้ในการเรียน
 รู้ทอพอโลยีของกราฟตามขั้นตอนข้างต้น กำหนดให้
 พารามิเตอร์ของแบบจำลองที่สอดคล้องกับทอพอโลยี
 หนึ่งๆ คือ \hat{A} ส่วน validation set ใช้ในการคำนวณค่า
 ความผิดพลาดของแบบจำลองนั้น เราจะเลือกแบบจำลอง
 ที่ให้

$$\text{relative error} = \frac{\|Y - \hat{A}H\|_F}{\|Y\|_F} \times 100\%$$

ต่ำที่สุด

3 ผลการทดลอง

ในส่วนนี้เราทำการทดลองบนข้อมูลสังเคราะห์ จากนั้นนำวิธีที่ใช้ไปประยุกต์ใช้วิธีในการเรียนรู้โครงสร้างเชิงสาเหตุในข้อมูล fMRI

การสังเคราะห์ข้อมูล

ในการสร้างแบบจำลองที่มีเสถียรภาพและได้เมทริกซ์เบาบางนั้น เราเริ่มจากการสุ่มค่าให้แนวทแยงมุมของเมทริกซ์ A_1, \dots, A_p จากนั้นในขั้นตอนที่สอง เราจะเพิ่มบางตำแหน่งแบบสุ่มให้เมทริกซ์ A_1, \dots, A_p โดยจำนวนที่เพิ่มนั้นจะขึ้นอยู่กับปริมาณความเบาบางที่ต้องการ จากนั้นเราจะทำการตรวจสอบเสถียรภาพของแบบจำลองว่ามีเสถียรภาพหรือไม่ หากไม่เราจะทำขั้นตอนที่สองซ้ำ

หลังจากได้เมทริกซ์แล้ว สังเคราะห์ข้อมูลตาม (1) โดยกำหนดให้ $u(t)$ มีการแจกแจงแบบปกติ

โดยเราจะสร้างข้อมูลมา 3 ชุดซึ่งแบ่งตามลักษณะความหนาแน่นของกราฟนั้นคือ 5%, 35%, 70% โดยค่าตัวแปรอื่นๆ กำหนดดังนี้ $y \in \mathbf{R}^{20}$, $p = 3$ จำนวนข้อมูล $N = 1500$

การแก้ปัญหาการประมาณแบบจำลอง

จากส่วนที่ 2.3 จะพบว่าต้องอาศัยการคำนวณเพื่อให้ได้แบบจำลองเป็นจำนวน $5M$ แบบ ถ้า $M = 400$ จะได้ว่าต้องใช้เวลามากในการแก้ปัญหา 2000 ครั้ง ดังนั้นในโครงการนี้ เราจะใช้ขั้นตอนวิธี ADMM (Alternating Direction Method of Multipliers) [13] ในการหาผลเฉลยของปัญหา (6) เนื่องจาก ADMM เป็นขั้นตอนวิธีที่อาศัยเพียงเกรเดียนต์ของฟังก์ชันจุดประสงค์ ทำให้มีความรวดเร็วพอสมควรภายใต้ค่าความคลาดเคลื่อนที่ยอมรับได้ จากนั้นนำรูปแบบร่วมของศูนย์ที่ได้มาเป็นเงื่อนไขรูปแบบศูนย์ในปัญหา (5) ซึ่งสามารถลดรูปได้เป็นการแก้สมการเชิงเส้นเท่านั้น [13]

การตรวจสอบความถูกต้องของทอพอโลยี

หลังจากได้ทอพอโลยีมาแล้ว เราจะตรวจสอบความถูกต้องของทอพอโลยีที่ประมาณได้ โดยการเปรียบเทียบรูปแบบร่วมของศูนย์ที่ประมาณได้กับรูปแบบร่วมของศูนย์ที่สร้างขึ้น การเปรียบเทียบดังกล่าวคือการพิจารณาทีละคู่อันดับ (i, j) ระหว่างทอพอโลยีที่ได้จากแบบจำลองกับทอพอโลยีที่สร้างขึ้น หากทอพอโลยีที่สร้างขึ้นที่คู่อันดับ (i, j) มีค่าเป็นศูนย์ แต่ค่าที่ประมาณได้มีค่าไม่เป็นศูนย์ เราจะเรียกความผิดพลาดนี้ว่า mismatch zeros หรือ false negative (FN) ในทางกลับกันหากทอพอโลยีที่สร้างขึ้นที่คู่อันดับ (i, j) มีค่าไม่เป็นศูนย์ แต่ค่าที่ประมาณได้มีค่าเป็นศูนย์ เราจะเรียกความผิดพลาดนี้ว่า mismatch nonzeros หรือ false positive (FP) ดังนั้นค่าความผิดพลาดของทอพอโลยีสามารถคำนวณได้จาก

$$\text{error} = \frac{\#\text{mismatch zeros} + \#\text{mismatch nonzeros}}{n^2 - n} \quad (10)$$

ในการทดลองกับข้อมูลจริงเราไม่อาจวิเคราะห์ความถูกต้องด้วยวิธีข้างต้น อย่างไรก็ตามมีวิธีที่น่าสนใจในการวิเคราะห์ความถูกต้องอีกหลายวิธี เช่นใน [14] เนื่องจากการทำข้อมูล fMRI แบบเวลาจริง ดังนั้นจึงใช้วิธีวิเคราะห์ความแม่นยำโดยการป้อนกลับ หลักการคือหลังจากหาแบบจำลองเสร็จแล้ว จะวัดค่าความผิดพลาด จากนั้นป้อนกลับโดยการให้บุคคลที่กำลังสแกนสมองอยู่นั้น ทำภารกิจอย่างหนึ่งเพื่อกระตุ้นสมองส่วนที่สนใจ จากนั้นหาแบบจำลองอีกครั้งและดูว่าค่าความผิดพลาดลดลงหรือไม่ เป็นต้น หรือใน [2] ใช้วิธีวิเคราะห์ความแม่นยำโดยดูจากความถูกต้องในการนำแบบจำลองที่ได้ไปใช้ในการคัดแยกผู้ป่วยจริงกับคนปกติ

การเลือกแบบจำลอง

ในส่วนนี้ เราจะทดลองเปรียบเทียบทอพอโลยีที่ได้จากแบบจำลองด้วยวิธีที่ต่างกัน เราทดลองบนข้อมูล 40 ตัวอย่างที่ลักษณะความหนาแน่นของกราฟ คือ 5%,

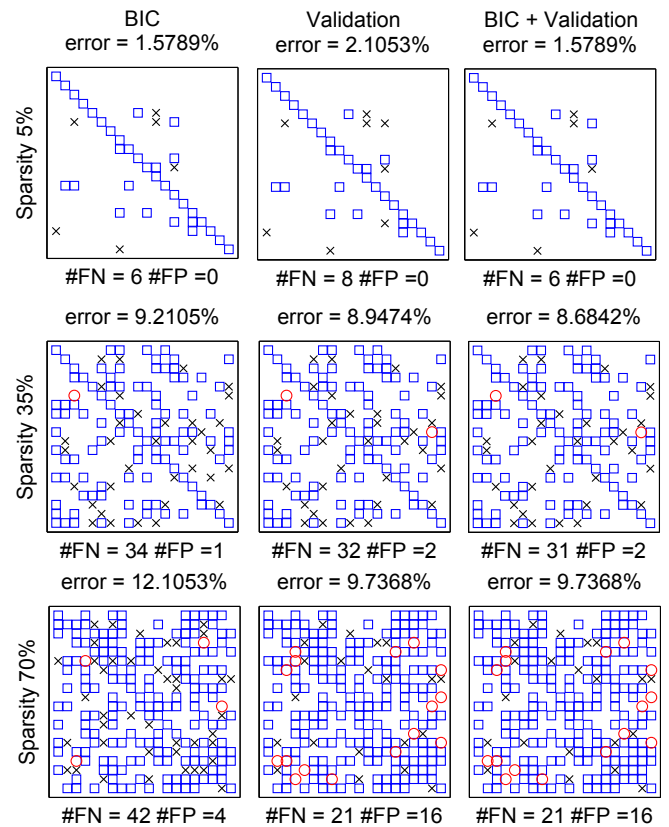
35%, 70% ตามลำดับโดยกำหนดให้ $n = 20, p = 3, N = 1500$ รวมทั้งหมด 120 ครั้ง ตารางที่ 1 แสดงจำนวนครั้งที่วิธีในการสร้างทอพอโลยีนั้นๆ ให้ค่าความผิดพลาดต่ำที่สุดเมื่อเทียบกับวิธีที่เหลือ โดยค่าความผิดพลาดคำนวณตาม (10)

ตารางที่ 1: ผลการทดลองเปรียบเทียบประสิทธิภาพในการหาทอพอโลยีของ 3 วิธี: วิธีสุดท้ายคือการนำทอพอโลยีจาก BIC และ Validation มายูเนียนกัน โดยแต่ละ density เราจะทดลองทั้งหมด 40 ครั้ง ตารางข้างล่างนี้แสดงสัดส่วนจำนวนครั้งที่วิธีนั้นๆ จะเลือกทอพอโลยีที่ผิดพลาดน้อยที่สุด

Density	BIC	Validation	BIC+Validation
5%	0.825	0.375	0.750
35%	0.350	0.450	0.725
70%	0.175	0.650	0.700
Expectation	0.450	0.492	0.725

จากตารางที่ 1 จะเห็นได้ว่าหากเราใช้ BIC ในการเลือกแบบจำลองและพารามิเตอร์ของแบบจำลองที่ได้มีลักษณะเบาบาง แล้วเราจะได้ว่ามีความเป็นไปได้อย่างมากที่แบบจำลองนั้นจะให้ค่าความผิดพลาดของทอพอโลยีต่ำกว่าการเลือกด้วยวิธีอื่น วิธีการรวมทอพอโลยีจากแบบจำลองที่เลือกมาจากทั้งวิธี BIC และ Validation Method เข้าด้วยกัน จะมีค่าเฉลี่ยของความน่าจะเป็นที่ทอพอโลยีที่ได้จะให้ค่าความผิดพลาดของทอพอโลยีต่ำกว่าวิธีอื่น ดังนั้นกลยุทธ์ที่เหมาะสมในการเลือกแบบจำลองคือการทำ BIC ในรอบแรกจากนั้นดูว่าแบบจำลองที่ถูกเลือกมีลักษณะอย่างไร หากพารามิเตอร์ของแบบจำลองที่ได้มีลักษณะเบาบาง เราสามารถมั่นใจได้ว่าแบบจำลองที่ได้จะให้ค่าความคลาดเคลื่อนของทอพอโลยีต่ำกว่าแบบจำลองที่เลือกด้วยวิธีอื่น แต่ถ้าพารามิเตอร์ของแบบจำลองที่ได้จาก BIC ไม่มีลักษณะเบาบาง เราจะใช้

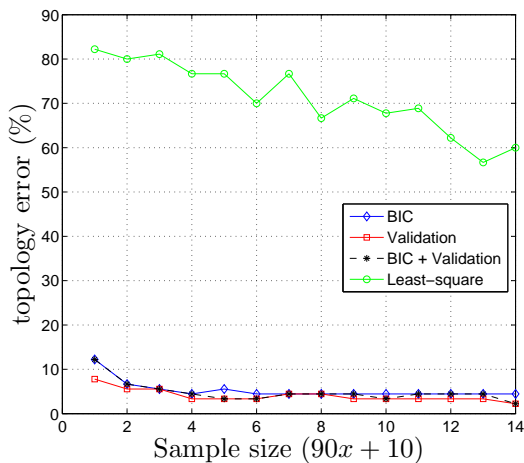
วิธี Validation Method เลือกแบบจำลอง จากนั้นนำทอพอโลยีที่ได้จากทั้งสองวิธีมายูเนียนเพื่อสร้างแบบจำลองเชิงกราฟ รูปที่ 5 แสดงผลลัพธ์บางส่วนในการทดลอง โดยจะเห็นได้ว่ากลยุทธ์ข้างต้นจะทำให้เราเลือกทอพอโลยีที่มีค่าผิดพลาดน้อยที่สุด



รูปที่ 5: กราฟแสดงค่าผิดพลาดของรูปแบบรวมของศูนย์ในการเลือกแบบจำลองด้วยวิธีต่างๆ ที่ลักษณะความหนาแน่น 5%, 35%, 70% ตามลำดับ เราจะเห็นได้ว่าการรวมทอพอโลยีของ BIC และ Validation เข้าด้วยกันให้ผลการเลือกที่ดีที่สุด โดยสีเหลี่ยมสีฟ้าหมายถึงตำแหน่งที่ค่าจริงไม่เป็นศูนย์แต่ประมาณได้ถูกต้อง วงกลมสีแดง (FP) หมายถึงตำแหน่งที่ค่าจริงเป็นศูนย์แต่ประมาณว่าไม่เป็นศูนย์ กากบาทสีดำ (FN) หมายถึงตำแหน่งที่ค่าจริงไม่เป็นศูนย์แต่ประมาณว่าเป็นศูนย์

จากสมมติฐานเกี่ยวกับการเลือกทอพอโลยีข้างต้น เราทดลองตรวจสอบประสิทธิภาพของแต่ละวิธีโดยการเปลี่ยนจำนวนข้อมูลที่ใช้ ได้ผลการทดลองดังรูปที่ 6 ซึ่งจะเห็นได้ว่าทั้งสามวิธีข้างต้นให้ค่าความผิดพลาดของทอพอโลยีที่ต่ำกว่าวิธีกำลังสองต่ำสุดอย่างมากโดยเฉพาะเมื่อจำนวนข้อมูลมีปริมาณน้อยมากๆ (ในกรณีที่ใช้วิธี

กำลังสองต่ำสุด เราจะนับว่าค่าที่น้อยกว่า 10% ของค่าสูงสุดมีค่าเป็นศูนย์) ดังนั้นเราอาจกล่าวได้ว่าวิธีการเลือกทอพอโลยีที่ใช้ในโครงการฉบับนี้เป็นวิธีที่มีประสิทธิภาพในระดับหนึ่ง

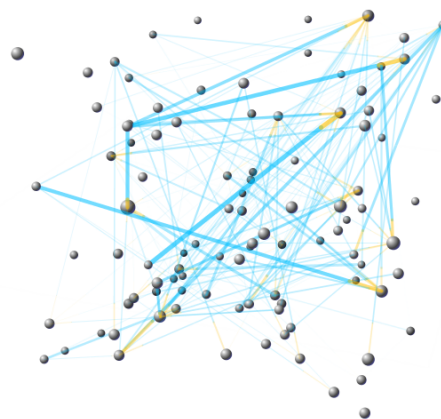


รูปที่ 6: กราฟแสดงค่าผิดพลาดของรูปแบบร่วมของศูนย์ในการเลือกแบบจำลองด้วยวิธีต่างๆ ที่ลักษณะความหนาแน่น 5% ที่ $n = 10, p = 2$ โดยการเพิ่มจำนวนข้อมูลที่ใช้ตั้งสมการ $90x + 10$ เราจะเห็นได้ว่าทั้งสามวิธีให้ผลลัพธ์ที่น่าพอใจเมื่อเทียบกับวิธีกำลังสองต่ำสุด วิธีรวมทอพอโลยีจาก BIC และ Validation จะให้ผลลัพธ์ที่อยู่ระหว่าง BIC และ Validation ซึ่งทำให้โดยเฉลี่ยแล้ว วิธีนี้มีค่าเฉลี่ยของความผิดพลาดต่ำที่สุด

การประยุกต์วิธีที่เสนอกับข้อมูล fMRI

ข้อมูล fMRI ที่ใช้ในครั้งนี่คือค่า BOLD (blood oxygen-level dependent) ซึ่งเป็นค่าที่นิยมในการนำมาวิเคราะห์เพื่อศึกษาการทำงานของสมอง เราใช้ข้อมูลนี้เราการหาความสัมพันธ์เชิงพลวัตระหว่างแต่ละจุดในสมอง ในการทดลองนี้เราจะหาโครงสร้างเชิงสาเหตุผ่านทางข้อมูล BOLD ของจุด 6004 จุด หรือ $n = 6004$ ใน (1) ซึ่งจะได้ว่าจำนวนตัวแปรทั้งหมดคือ n^2p อยู่ในระดับ 10^9 เนื่องด้วยขีดจำกัดในหลายๆ ด้าน (เช่น จำนวนข้อมูล, ประสิทธิภาพของเครื่องคอมพิวเตอร์) ทำให้เราไม่อาจแก้ปัญหาขนาดดังกล่าวได้ ดังนั้นเราจึงใช้วิธี PCA ในการลดจำนวนตัวแปรลงเหลือ $n = 100$ ซึ่งช่วยทำให้การแก้ปัญหาทำได้ง่ายขึ้น จากนั้นแก้ปัญหาค่าประมาณแบบจำลองตามขั้นตอนการเรียนรู้ทอพอโลยีในส่วนที่ 2.3

โดยจะได้ทอพอโลยีทั้งหมด 2000 แบบ เราใช้วิธี BIC และ Validation ในการเลือกทอพอโลยี จากนั้นนำทอพอโลยีจากทั้งสองวิธีมาเทียบกัน แบบจำลองเชิงกราฟที่ได้จะมีลักษณะดังรูป 7 เราจะเห็นได้ว่าแบบจำลองที่ได้สอดคล้องกับสมมติฐานที่ว่า การทำงานของสมองจะสัมพันธ์กันอย่างโดดเด่นกับบางส่วนเท่านั้น



รูปที่ 7: แบบจำลองเชิงกราฟของข้อมูล fMRI จุดสีเทาหมายถึง จุดแต่ละจุดในสมอง ความยาวของเส้นสีส้มหมายถึงปริมาณของความสัมพันธ์โดยปลายที่มีสีส้มหมายถึงเป็นหัวลูกศร หมายเหตุ พิกัดที่ได้นี้เป็นการสมมติขึ้นเนื่องจากเราไม่มีข้อมูลพิกัดที่แท้จริงของทั้ง 6004 จุด ในกรณีที่มีพิกัดนั้น พิกัดของ 100 จุดจะเป็นพิกัดของจุดเซนทรอยด์ของแต่ละองค์ประกอบในเวกเตอร์ค่าเจาะจงที่ใช้ในการทำการแปลงเชิงเส้นด้วยวิธี PCA

4 บทสรุป

เราแสดงวิธีการเรียนรู้โครงสร้างเชิงสาเหตุสำหรับข้อมูลการทำงานในสมองผ่านแบบจำลองเชิงกราฟ โดยอาศัยการแก้ปัญหาค่าเหมาะที่สุดแบบคอนเวกซ์ สำหรับเรียนรู้ทอพอโลยี ซึ่งสมมูลกับการประมาณค่าพารามิเตอร์ในแบบจำลอง MAR ที่กำหนดให้พารามิเตอร์มีลักษณะเบาบาง การแก้ปัญหาค่าเหมาะที่สุดที่มีขนาดใหญ่นี้ เราใช้วิธี PCA เพื่อลดมิติของตัวแปรลง จากนั้นประยุกต์ใช้วิธี ADMM ซึ่งเป็นวิธีที่รวดเร็วและเหมาะสมกับปัญหาขนาดใหญ่ในการหาผลเฉลยการประมาณ ในการเลือกแบบจำลอง เราใช้ทั้งวิธี BIC และ Validation ในการ

เลือกแบบจำลอง ซึ่งแบบจำลองที่เลือกมาทั้งสองแบบนี้ อาจให้ทอพอโลยีที่ต่างกัน ดังนั้นเราจะพิจารณาทอพอโลยีทั้งสองประกอบกันในการสร้างแบบจำลองเชิงกราฟของข้อมูล fMRI

เนื่องจากแบบจำลองเชิงกราฟที่ได้ไม่มีพิกัดอ้างอิง ดังนั้นการตีความจากแบบจำลองเชิงกราฟที่ได้ อาจทำได้ยาก ในอนาคตเราจะสร้างแบบจำลองเชิงกราฟที่อ้างอิงพิกัดที่ถูกต้องเพื่อให้มีความชัดเจนในการตีความมากขึ้น หากเราสามารถเก็บข้อมูลได้มากขึ้น เราอาจใช้แบบจำลองในการเรียนกระบวนการคิดในแต่ละกรณีเพื่อใช้ในการคัดแยกสภาวะบุคคลได้

5 กิตติกรรมประกาศ

ขอขอบคุณ Dr. Pedro Antonio Valdes-Sosa ศาสตราจารย์ประจำ Cuban Neuroscience Center (CNEURO) ที่กรุณาให้ข้อมูล fMRI มาใช้ในโครงการนี้ และขอขอบคุณ นายธนชัย บรรลุทางธรรม นิสิตปี 4 ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย สำหรับโปรแกรมในการสร้างภาพ 3 มิติ

6 เอกสารอ้างอิง

- [1] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer Verlag, 2006.
- [2] O. Demirci, M. C. Stevens, N. C. Andreasen, A. Michael, J. Liu, T. White, G. D. Pearlson, V. P. Clark, and V. D. Calhoun. Investigation of relationships between fMRI brain networks in the spectral domain using ICA and Granger causality reveals distinct differences between schizophrenia patients and healthy controls. *Neuroimage*, 46(2):419–431, 2009.
- [3] R. Garg, G. A. Cecchi, and A. R. Rao. Full-brain autoregressive modeling (FARM) using fMRI. *Neuroimage*, 58(2):416–441, 2011.
- [4] G. Deshpande, S. LaConte, G. A. James, S. Peltier, and X. Hu. Multivariate Granger causality analysis of fMRI data. *Human Brain Mapping*, 30(4):1361–1373, 2009.
- [5] R. Goebel, A. Roebroeck, D. S. Kim, and E. Formisano. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging*, 21(10):1251–1261, 2003.
- [6] W. Tang, S. L. Bressler, C. M. Sylvester, G. L. Shulman, and M. Corbetta. Measuring Granger Causality between Cortical Regions from Voxelwise fMRI BOLD Signals with LASSO. *PLoS Computational Biology*, 8(5):e1002513, 2012.
- [7] E. Alpaydin. *Introduction to Machine Learning*. MIT press, 2004.
- [8] A. Leon-Garcia. *Probability, Statistics, and Random Processes for Electrical Engineering*. Pearson/Prentice Hall, 2009.
- [9] J. Songsiri. Learning Granger Graphical Models via Alternating Direction Method of Multipliers. Technical report, Department of Electrical Engineering, Chulalongkorn University, 2012.
- [10] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International, London, 1989.
- [11] S. Ryali, T. Chen, K. Supekar, and V. Menon. Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *Neuroimage*, 59(4):3852–3861, 2011.
- [12] N. D. Daw, S. J. Gershman, B. Seymour, P. Dayan, and R. J. Dolan. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215, 2011.
- [13] P. Lertkultanon and J. Songsiri. A Comparison of Fast Algorithms for ℓ_1 -Type Penalized Estimation Problems. Technical report, Department of Electrical Engineering, Chulalongkorn University, 2013.
- [14] C. O. Plumptre, L. I. Kuncheva, N. N. Oosterhof, and S. J. Johnston. Naive random subspace ensemble with linear classifiers for real time classification of fMRI data. *Pattern Recognition*, 45(6):2101–2108, 2012.