

บทที่ 2

แบบจำลองการกรองข้อมูลสำหรับโครงข่ายประสาทเทียม

ในงานวิจัยชิ้นนี้มุ่งเน้นไปที่การทดสอบเพื่อหาเทคนิคการกรองข้อมูลที่เหมาะสมกับข้อมูลประเภทต่างๆ แบบจำลองการกรองข้อมูลสำหรับโครงข่ายประสาทเทียม (Data Filtering for Neural Network Forecasting: DFNNF) ออกแบบเพื่อเลือกใช้เทคนิคการกรองข้อมูลที่เหมาะสมกับข้อมูลประเภทต่างๆ ในบทนี้จะกล่าวถึงการออกแบบแบบจำลองการกรองข้อมูลสำหรับโครงข่ายประสาทเทียม หรือ DFNNF แสดงได้ดังภาพประกอบที่ 2.1 แบ่งการทำงานออกเป็น 4 ขั้นตอนหลัก นั่นคือ ขั้นตอนวิธีที่ 1 คือ เตรียมข้อมูล (Data Preparation) ขั้นตอนวิธีที่ 2 คือ ขั้นตอนการกรองข้อมูล (Data Filtering) ขั้นตอนวิธีที่ 3 สอนโครงข่ายประสาทเทียม (Training Neural Networks) และ ขั้นตอนวิธีที่ 4 ขั้นตอนพยากรณ์ผลลัพธ์ (Neural Networks Forecasting)

ขั้นตอนวิธีที่ 1 : เตรียมข้อมูล (Data Preparation)
ขั้นตอนวิธีที่ 2 : การกรองข้อมูลและแบ่งช่วงข้อมูล (Data Filtering and Diving)
ขั้นตอนวิธีที่ 3 : สอนโครงข่ายประสาทเทียม (Training Neural Networks)
ขั้นตอนวิธีที่ 4 : พยากรณ์ผลลัพธ์ (Neural Networks Forecasting)

ภาพประกอบที่ 2.1 ขั้นตอนวิธีของแบบจำลองการกรองข้อมูลสำหรับโครงข่ายประสาทเทียม (Data Filtering Technique for Neural Networks Forecasting: DFNNF)

2.1 ขั้นตอนวิธีเตรียมข้อมูล (Data Preparation)

เป็นขั้นตอนในการเตรียมข้อมูลให้พร้อมที่จะนำไปประมวลผล เนื่องจากการจัดเก็บข้อมูลรวบรวมข้อมูลจากหลายแหล่ง ในบางครั้งข้อมูลอาจอยู่ในรูปแบบที่แตกต่างกัน หรือมีความไม่สมบูรณ์ จึงต้องมีการเตรียมข้อมูลก่อนการประมวลผล มีรายละเอียดการทำงาน ดังภาพประกอบที่ 2.2

ขั้นตอนวิธีที่ 1 : เตรียมข้อมูล (Data Preparation)
1.1 แบ่งข้อมูลออกเป็น 2 ชุด คือ ข้อมูลชุดสอน (Training Set) และข้อมูลชุดทดสอบ (Testing Set)
1.2 แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยข้อมูลรอบข้าง ดังสมการที่ (1.1)
1.3 กำหนดให้ $X(t) = \{x_1, x_2, \dots, x_n\}$ คือ เซตของตัวแปรเข้าโครงข่ายประสาทเทียมที่เวลา (t) โดย n คือจำนวนแอดทริบิวต์ของข้อมูลชุดนั้นๆ
1.4 กำหนดข้อมูลผลลัพธ์ (y) ของโครงข่ายประสาทเทียมที่เวลา $t + 1$ ในกรณีที่ข้อมูลอากาศ

ภาพประกอบที่ 2.2 ขั้นตอนวิธีการเตรียมข้อมูล (Data Preparation)

ขั้นตอนที่ 1.1 ทำการแบ่งข้อมูลออกเป็น 2 ชุด คือ ข้อมูลชุดสอน (Training Set) ใช้สำหรับการสอนโครงข่ายประสาทเทียมและข้อมูลชุดทดสอบ (Testing Set) ใช้สำหรับการทดสอบประสิทธิภาพการทำงานของโครงข่ายประสาทเทียม

ขั้นตอนที่ 1.2 ทำการแทนค่าข้อมูลสูญหาย (Missing Value) ด้วยค่าเฉลี่ยของข้อมูลรอบข้างดังสมการที่ (1.1)

ขั้นตอนที่ 1.3 กำหนดค่าข้อมูลตัวแปรเข้าสำหรับการสอนโครงข่ายประสาทเทียมที่เวลาปัจจุบัน (t) โดยกำหนดให้ตัวแปร n เป็นจำนวนตัวแปรข้อมูลเข้าทั้งหมด จะได้เซตของตัวแปรเข้าโครงข่ายประสาทเทียมที่เวลา (t) ดังนี้ $X(t) = \{x_1, x_2, \dots, x_n\}$

ขั้นตอนที่ 1.4 ในกรณีที่เป็นข้อมูลอากาศจะกำหนดค่าข้อมูลของตัวแปรผลลัพธ์เป็นเวลาถัดไป ($t+1$) เนื่องจากจะต้องสอนโครงข่ายประสาทเทียมให้พยากรณ์ผลลัพธ์ที่เวลาถัดไปว่าฝนจะตกหรือไม่

2.2 ขั้นตอนวิธีการกรองข้อมูลและแบ่งช่วงข้อมูล (Data Filtering and Diving)

เป็นขั้นตอนในการกรองข้อมูลเพื่อขจัดสิ่งรบกวนออกจากข้อมูล ซึ่งในขั้นตอนนี้จะทดสอบเทคนิคการกรองข้อมูลทั้ง 4 แบบ เพื่อต้องการทราบว่าเทคนิคการกรองข้อมูลแบบใดที่ให้ค่าความถูกต้องมากที่สุด และในขั้นตอนนี้จะมีการแบ่งข้อมูลออกเป็นกลุ่มเพื่อช่วยให้โครงข่ายประสาทเทียมเรียนรู้ได้ดียิ่งขึ้น มีรายละเอียดการทำงานดังภาพประกอบที่ 2.3

ขั้นตอนที่ 2.1 กรองข้อมูล ในขั้นตอนนี้เป็นขั้นตอนในการกรองข้อมูลเพื่อที่จะขจัดสิ่งรบกวนออกไปจากข้อมูล โดยทำการทดลองทั้ง 4 เทคนิคการกรองข้อมูลเพื่อหาว่าเทคนิคการกรองข้อมูลแบบใดให้ค่าความถูกต้องมากที่สุด

ขั้นตอนที่ 2.2 ทำการแบ่งข้อมูลออกเป็นกลุ่มโดยใช้วิธีการแบ่งกลุ่มโดยแบ่งช่วงค่าเท่ากัน เพื่อลดจำนวนค่าที่เป็นไปได้ของข้อมูลเข้าให้อยู่ในขอบเขตที่กำหนด และโครงข่ายประสาทเทียมสามารถเรียนรู้ได้ง่ายขึ้น ในงานวิจัยนี้ได้แบ่งช่วงข้อมูลออกเป็น 5 กลุ่ม คือ เล็กมาก, เล็ก, กลาง, ใหญ่ และ ใหญ่มาก

การแบ่งข้อมูลออกเป็นกลุ่ม โดยแบ่งช่วงค่าเท่ากัน สามารถแบ่งกลุ่ม โดยคำนวณค่าระยะห่างของช่วงในแต่ละกลุ่มได้ดังสมการที่ (2.1)

$$r = \frac{\max - \min}{q} \tag{2.1}$$

กำหนดให้ ตัวแปร r คือระยะห่างของช่วงในแต่ละกลุ่ม
ตัวแปร q คือจำนวนกลุ่มที่ต้องการ

ตัวแปร max คือค่าข้อมูลสูงสุด

ตัวแปร min คือค่าข้อมูลต่ำสุด

ขั้นตอนที่ 2.3 เปลี่ยนรูปข้อมูลให้อยู่ในช่วง $[0,1]$ เพื่อให้โครงข่ายประสาทเทียมสามารถนำข้อมูลไปประมวลผลได้ โดยใช้สมการ (1.2) ในการเปลี่ยนรูปข้อมูลให้อยู่ในช่วงศูนย์ถึงหนึ่ง

ขั้นตอนวิธีที่ 2 : การกรองข้อมูลและแบ่งช่วงข้อมูล (Data Filtering and Diving)

2.1 กรองข้อมูล โดยเลือกใช้เทคนิคการกรองข้อมูลดังต่อไปนี้

2.1.1 Moving Average Filtering

2.1.1.1 กำหนดพารามิเตอร์ขนาดหน้าต่าง (Span) = n

2.1.1.2 คำนวณค่าข้อมูลที่กรองแล้วของจุดข้อมูลตรงกลางของหน้าต่างจากสมการ (1.3)

2.1.1.3 เลื่อนหน้าต่างไปทางขวาหนึ่งจุดข้อมูล

2.1.1.4 ทำซ้ำข้อ 2.1.1.2 จนหมดข้อมูล

2.1.2 Local Regression Filtering

2.1.2.1 กำหนดพารามิเตอร์ขนาดหน้าต่าง (Span) = n

2.1.2.2 กำหนดน้ำหนักให้ข้อมูลทุกจุดในหน้าต่างจากสมการที่ (1.16)

2.1.2.3 ทำการหาค่าข้อมูลที่กรองแล้ว ใช้การวิเคราะห์การถดถอยแบบเชิงเส้น (lowess) หรือแบบพหุนามกำลังสอง (loess)

2.1.2.4 เลื่อนหน้าต่างไปทางขวาหนึ่งจุดข้อมูล

2.1.2.5 ทำซ้ำข้อ 2.1.2.2 จนหมดข้อมูล

2.1.3 Savitzky-Golay Filtering

2.1.3.1 กำหนดพารามิเตอร์ขนาดหน้าต่าง (Span) = n

2.1.3.2 กำหนดโพลีโนเมียลดีกรี = p โดยที่ $p < n$

2.1.3.3 คำนวณค่าสัมประสิทธิ์และค่าข้อมูลที่กรองแล้วโดยใช้หลักการของสมการถดถอยแบบกำลังสองน้อยที่สุด

2.1.3.4 เลื่อนหน้าต่างไปทางขวาหนึ่งจุดข้อมูล

2.1.3.5 ทำซ้ำข้อ 2.1.3.3 จนหมดข้อมูล

2.1.4 Hamming Window Filtering

2.1.4.1 กำหนดพารามิเตอร์ขนาดหน้าต่าง (Span) = n

2.1.4.2 คำนวณค่าข้อมูลที่ผ่านการกรองแล้วจากสมการ (1.18)

2.1.4.3 เลื่อนหน้าต่างไปทางขวาหนึ่งจุดข้อมูล

2.1.4.4 ทำซ้ำข้อ 2.1.4.2 จนหมดข้อมูล

2.2 แบ่งข้อมูลออกเป็นกลุ่มโดยมีช่วงค่าเท่ากันด้วยสมการ (2.1) ซึ่งจะทำได้หรือไม่ก็ได้

2.3 เปลี่ยนรูปข้อมูลให้อยู่ในช่วง $[0,1]$ ด้วยสมการ (1.2)

ภาพประกอบที่ 2.3 ขั้นตอนวิธีการกรองข้อมูลและแบ่งช่วงข้อมูล (Data Filtering and Diving)

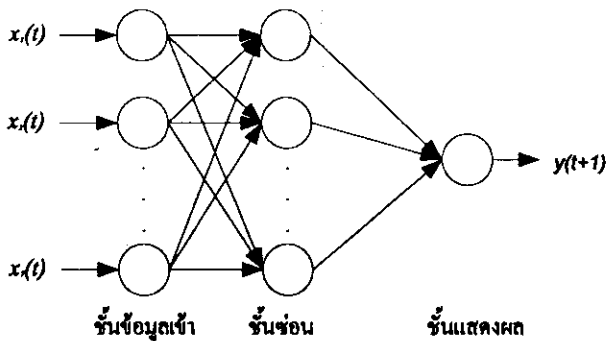
2.3 ขั้นตอนวิธีสอนโครงข่ายประสาทเทียม (Training Neural Networks)

เป็นขั้นตอนสอนให้โครงข่ายประสาทเทียมเรียนรู้ข้อมูล โดยใช้ข้อมูลชุดสอนที่ผ่านขั้นตอนที่ 1 การเตรียมข้อมูล และขั้นตอนที่ 2 การกรองข้อมูลและแบ่งช่วงข้อมูลแล้ว แสดงรายละเอียดดังภาพประกอบที่ 2.4 ดังนี้

ขั้นตอนวิธีที่ 3 : สอนโครงข่ายประสาทเทียม (Training Neural Networks)
3.1 กำหนดสถาปัตยกรรม โดยชั้นข้อมูลเข้าและชั้นซ่อนมีจำนวนโหนดเท่ากับจำนวนตัวแปรเข้าของข้อมูลและชั้นแสดงผลมี 1 โหนด
3.2 ใช้ข้อมูลชุดสอนสอนโครงข่ายประสาทเทียม

ภาพประกอบที่ 2.4 ขั้นตอนวิธีสอนโครงข่ายประสาทเทียม (Training Neural Networks)

ขั้นตอนที่ 3.1 กำหนดสถาปัตยกรรมของโครงข่ายประสาทเทียม กำหนดให้ชั้นข้อมูลเข้าและชั้นซ่อนมีจำนวนโหนดเท่ากับจำนวนตัวแปรเข้าของข้อมูล และชั้นแสดงผลมี 1 โหนด ซึ่งสถาปัตยกรรมแบบนี้จะให้ค่าความถูกต้องในการพยากรณ์ข้อมูลอากาศที่สูง [17] เช่น จำนวนตัวแปรเข้ามี 7 ตัวแปร ในชั้นข้อมูลเข้าจะมีและชั้นซ่อนจะมี 7 โหนด และชั้นแสดงผลมี 1 โหนด แสดงตัวอย่างสถาปัตยกรรมโครงข่ายประสาทเทียมที่ใช้ดังภาพประกอบที่ 2.5



ภาพประกอบที่ 2.5 สถาปัตยกรรมของโครงข่ายประสาทเทียมที่ใช้

ขั้นตอนที่ 3.2 ใช้ข้อมูลชุดสอนสอนโครงข่ายประสาทเทียม สำหรับข้อมูลที่เป็นข้อมูลอากาศนั้นในการสอนโครงข่ายประสาทเทียมจะต้องสอนที่เวลาถัดไป นั่นคือ กำหนดค่าข้อมูลของตัวแปรผลลัพธ์ที่เวลาถัดไป ($t+1$) เนื่องจากต้องพยากรณ์ที่เวลาต่อไปเสมอฝนจะตกหรือไม่ แต่สำหรับข้อมูลทางการแพทย์ไม่ต้องพยากรณ์ที่เวลาถัดไป ตัวอย่างข้อมูลสำหรับการสอนโครงข่ายประสาทเทียมแสดงได้ดังตารางที่ 2.1 จะใช้ข้อมูล x_1 ถึง x_7 ณ เวลา (t) เช่น 01.00น. เป็นข้อมูลเข้า และใช้ข้อมูลปริมาณฝนที่เวลาถัดไป ($t+1$) เช่น 04.00น. เป็นข้อมูลผลลัพธ์ ตัวอย่างเช่น จะได้ค่า x_1 ถึง x_7 คือ [5, 17.1, 1012.3, 41, 32, 0, 0] (บรรทัดแรก) และค่า y คือ 5.5 (บรรทัดที่ 2) เป็นต้น

ตารางที่ 2.1 ตารางแสดงตัวอย่างข้อมูลการพยากรณ์ฝนแบ่งเป็นราย 3 ชั่วโมง

เวลา	x_1 เมฆ	x_2 อุณหภูมิ จุดน้ำค้าง	x_3 ความกด อากาศ	x_4 ความชื้น สัมพัทธ์	x_5 อุณหภูมิ	x_6 ความเร็ว ลม	x_7 ทิศทางลม	Y ปริมาณ ฝน
01.00								0
04.00	4	18.7	1011.1	43	33	4	60	5.5
07.00	5	17.7	1010.6	38	34	0	50	0
10.00	2	19.2	1010.4	41	35	2	0	0.1
13.00	2	15.1	1014.7	43	29	6	20	0.7
16.00	1	11.9	1015.9	37	28	7	40	0
19.00	0	12.7	1015.9	40	28	6	40	4.2
22.00	3	13.3	1014.3	35	30	4	190	0

หมายเหตุ ในการสอนโครงข่ายประสาทเทียม มีการกำหนดจำนวนรอบในการสอนโครงข่ายประสาทเทียมเป็นจำนวน 100 รอบ (epoch = 100) นั่นคือ โครงข่ายประสาทเทียมจะหยุดการเรียนรู้เมื่อค่าความผิดพลาดที่เกิดขึ้นเท่ากับ 0 หรือจำนวนรอบในการทำงานเท่ากับ 100

2.4 ขั้นตอนวิธีพยากรณ์ผลลัพธ์ (Neural Networks Forecasting)

ขั้นตอนวิธีพยากรณ์ผลลัพธ์ โดยใช้ข้อมูลชุดทดสอบที่ผ่านขั้นตอนที่ 1 การเตรียมข้อมูล และขั้นตอนที่ 2 การกรองข้อมูลและแบ่งช่วงข้อมูลแล้ว แสดงรายละเอียดคังภาพประกอบที่ 2.6 ดังนี้

ขั้นตอนที่ 4 : พยากรณ์ผลลัพธ์ (Neural Networks Forecasting)
4.1 เตรียมข้อมูลทดสอบ โดยใช้ข้อมูลชุดทดสอบและสำหรับข้อมูลอากาศจะพยากรณ์ผลลัพธ์ที่เวลา $t+1$
4.2 พยากรณ์ผลลัพธ์ โดยใช้โครงข่ายประสาทเทียมที่ได้จากขั้นตอนที่ 3

ภาพประกอบที่ 2.6 ขั้นตอนวิธีพยากรณ์ผลลัพธ์ (Neural Networks Forecasting)

ขั้นตอนที่ 4.1 เตรียมข้อมูลทดสอบ โดยใช้ข้อมูลชุดทดสอบที่ผ่านขั้นตอนที่ 1 การเตรียมข้อมูล และขั้นตอนที่ 2 การกรองข้อมูลและแบ่งช่วงข้อมูลแล้ว และสำหรับข้อมูลอากาศจะพยากรณ์ผลลัพธ์ที่เวลา $(t+1)$ เช่นเดียวกับการสอนโครงข่ายประสาทเทียม

ขั้นตอนที่ 4.2 พยากรณ์ผลลัพธ์ โดยใช้โครงข่ายประสาทเทียมที่ใช้สอนแล้วในขั้นตอนที่ 3

แบบจำลอง DFNNF มีการทำงานตามลำดับการทำงานตามขั้นตอนดังนี้ 1) การเตรียมข้อมูล เพื่อทำความสะอาดข้อมูลและพร้อมที่จะนำไปประมวลผลต่อ 2) การกรองข้อมูลและแบ่งช่วงข้อมูล เพื่อขจัดสิ่งรบกวนออกจากข้อมูลและเปรียบเทียบเทคนิคการกรองข้อมูลแต่ละแบบว่าเทคนิคการกรองข้อมูลแบบใดที่ให้ค่าความถูกต้องมากที่สุด และแบ่งข้อมูลออกเป็นช่วง เพื่อให้โครงข่ายประสาทเทียมเรียนรู้ข้อมูลได้ง่ายขึ้น 3) สอน โครงข่ายประสาทเทียม และ 4) พยากรณ์ผลลัพธ์