

ภาคผนวก ก

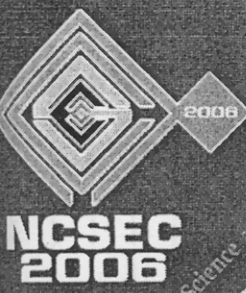
เอกสารงานวิจัยที่ตีพิมพ์ในงานประชุมวิชาการวิทยาการคอมพิวเตอร์

และวิศวกรรมคอมพิวเตอร์แห่งชาติ NCSEC 2006

ณ โรงแรม โซฟีเทล ราชา ออคิด ขอนแก่น

25-27 ตุลาคม 2549

The 10th National Computer Science and Engineering Conference, NCSEC 2006



การประชุมวิชาการ วิทยาการคอมพิวเตอร์และ วิศวกรรมคอมพิวเตอร์แห่งชาติ ครั้งที่ 10

The 10th National Computer Science and Engineering Conference

25-27 ตุลาคม 2549

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
และภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
มหาวิทยาลัยขอนแก่น

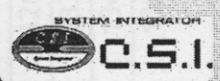
จัดโดย



ร่วมกับ



สนับสนุนโดย



Paper ID	Title
110	On-Line Fundamental Arithmetic Algorithms for Three-Dimensional Vector System
111	การพัฒนาตัวจัดลำดับเว็บเซอร์วิสเพื่อกระบวนการทางธุรกิจ
112	การวิเคราะห์และตัดแยกอักขรพิมพ์ภาษาไทยแบบสัมพันธ์และทับซ้อนกัน
113	Dynamic Routing Algorithm for Trailers in Container Ports
115	Ant Routing for Wireless Sensor Networks
118	Computing Positional Pattern of Multiple Robots from Angular Measures
140	An extended double base number system using different digit sets
120	Location Based Digital Signature on Mobile Internet Devices
121	An improvement of n-way test case generation algorithm
122	Campus Network Traffic Classification and Visualization Tool
123	ตัวแบบการค้นคืนแบบรูปการออกแบบ
124	Decentralized Construction of Active Views for Web Access Control
126	การบริหารความสัมพันธ์กับนักศึกษาในสถาบันระดับอุดมศึกษาโดยการประยุกต์การทำเหมืองข้อมูล
133	Evaluation of MPEG-4 Video Streaming with the Impact of UMTS RLC Parameters Setting
134	A Running Time Analysis for an Ant Colony Optimization Algorithm for Shortest Paths on Directed Acyclic Graphs
136	การประยุกต์ใช้แบบจำลองความน่าจะเป็นสำหรับทำนายผลสำเร็จการศึกษาของนักศึกษาระดับอาชีวศึกษา และระดับอุดมศึกษาโดยใช้เทคนิคการเรียนรู้แบบเบย์
137	การศึกษาหาขนาดเฟรมข้อมูลที่เหมาะสมสำหรับการสื่อสารดาวเทียมวงโคจรต่ำโดยใช้ โปรโตคอล PACSAT
139	GENETIC ALGORITHM & SEGMENTATION TO DISCOVER GRAPH ISOMORPHISM
141	การบีบอัดข้อมูลภาพโดยใช้เวกเตอร์ควอนไทเซชันชนิดปรับเปลี่ยน Haar เวฟเล็ต
143	Fundamental arithmetic operations for the flexible interval representation system
146	Data Filtering Technique for Weather Forecast using Neural Networks
145	Improving efficiency of discover motifs using motif concept
152	Disease-Causing Gene Pathway Discovery via Analysis of Metabolic Networks
153	Age-Based PageRank Computation
154	Hand-written Alpha-Numeric Recognition base-on Fuzzy-Chain Code
159	Formal Modeling of Practical and Secure Mobile Payment Systems
162	Performance Comparisons of MANET Routing Protocols for Multiple Sources Video Streaming Applications
163	Enabling Grid Computing Capabilities in Engineering Applications: Finite Element Calculations

OCTOBER 26, 2006 (Thursday)
At Sofitel Raja Orchid Hotel, Khon Kaen

8.00-9.00	Registration			
9.00-9.15	Welcome speech & Opening ceremony by KKU president			
9.15 – 10.15	Keynote speaker : Professor Dr Jean-Louis Lassez Topic: Computational Biology			
10.15-10.45	Keynote speaker : Dr.Pansak Siriruchatapong, NECTEC Director			
10.45 – 11.00	Break			
11.00 – 11.30	Keynote speaker : Stephen Regelous, CEO, Massive Company			
11.30-12.10	IT: P85	Com Sys: P51	DSP&DIP : P23	AI: P146
	IT: P136	Com Sys: P61	DSP&DIP : P103	AI: P28
12.10-13.30	Lunch			
13.30-14.15	Keynote speaker : Assoc.Prof.Dr.Djitt Laowattana, KMUTT Topic: Intelligent Robotics: Research & Applications			
14.15-14.30	Break			
14.30-16.30	IT: P66	รศ.ดร.วีระ บุญจริง	DSP&DIP: P141	AI: P83
	IT: P47	Com Sys: P170	รศ.ดร.ประกาศ จงสถิตย์วัฒนา	AI: P63
	IT: P91	Com Sys: P163	DSP&DIP: P102	AI: P77
	IT: P126	Com Sys: P162	DSP&DIP: P41	AI: P19
	IT: P40	Com Sys: P10	DSP&DIP: P59	AI: P87
	IT: P58	Com Sys: P134	DSP&DIP: P96	AI: P168
18.30 – 22.00	Reception Dinner			

เทคนิคการกรองข้อมูลเพื่อการพยากรณ์อากาศโดยใช้โครงข่ายประสาทเทียม Data Filtering Technique for Weather Forecast using Neural Networks

วิภาดา เวทย์ประสิทธิ์ และ ณสิทธิ์ เหล่าเสี้ยน

ห้องปฏิบัติการปัญญาประดิษฐ์ ภาควิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ อ. หาดใหญ่ จ. สงขลา 90112

E-mail: wwettayaprasit@yahoo.com nasith@gmail.com

บทคัดย่อ

การพยากรณ์อากาศโดยใช้โครงข่ายประสาทเทียมจะได้อะไรที่ถูกต้องแม่นยำขึ้นเมื่อใช้เทคนิคการกรองข้อมูลที่เหมาะสม บทความนี้ได้นำเสนอขั้นตอนวิธีการสอนโครงข่ายประสาทเทียมในการพยากรณ์ฝนและพายุ ณ เวลาถัดไป ด้วยเทคนิคการกรองข้อมูลแบบต่างๆคือ เทคนิคการกรองข้อมูลแบบเคลื่อนที่ เทคนิคการกรองข้อมูลแบบถดถอยท้องถิ่น เทคนิคการกรองข้อมูลแบบสวิสกี-โกเลย์ และเทคนิคการกรองข้อมูลแบบหน้าต่างแฮมมิง ใช้ข้อมูลอากาศเมืองริโอเดอจาเนโร และเมืองเซาเปาโลจากประเทศบราซิล และข้อมูลอากาศจังหวัดชลบุรี ประเทศไทย ผลการทดลองแสดงให้เห็นว่า เทคนิคการกรองข้อมูลแบบถดถอยท้องถิ่นให้ค่าความถูกต้องสูงสุด

คำสำคัญ: การกรองข้อมูล, โครงข่ายประสาทเทียม, การพยากรณ์อากาศ, สมการถดถอยท้องถิ่น

Abstract

The weather forecast by neural networks would be more precise and accurate when the filtering technique was handled properly. The paper presents a method of neural networks for rainfall forecast and storm forecast by using various techniques of data filtering such as moving average filtering technique, local regression filtering technique, Savitzky-Golay filtering technique and Hamming window filtering technique. The study

used weather data from Rio de Janeiro and Sao Paulo, Brazil and Chonburi, Thailand. The experimental result indicated that the local regression filtering technique gave maximum accuracy.

Keyword: Data filtering, Neural Network, Weather forecasting, Local regression

1. บทนำ

ในฐานะข้อมูลขนาดใหญ่ มักมีข้อมูลที่ไม่สมบูรณ์ เช่น มีค่าสุดโต่ง มีการบันทึกข้อมูลผิดพลาด ดังนั้นจึงต้องมีการกรองข้อมูลก่อนนำข้อมูลไปประมวลผล การกรองข้อมูล (Data Filtering) เป็นกระบวนการที่ใช้สำหรับกำจัดสิ่งรบกวน เช่น ค่าสุดโต่ง และค่าข้อมูลผิดพลาดออกจากข้อมูลทำให้ข้อมูลมีคุณภาพและเหมาะสมสำหรับนำไปประมวลผล ส่งผลให้มีความถูกต้องของผลการทดลองมากขึ้น เทคนิคในการกรองข้อมูลมีหลายเทคนิค คือ เทคนิคการกรองข้อมูลแบบเคลื่อนที่ (Moving Average Filtering) [1] เทคนิคการกรองข้อมูลแบบถดถอยท้องถิ่น (Local Regression Filtering) [2] เทคนิคการกรองข้อมูลแบบสวิสกี-โกเลย์ (Savitzky-Golay Filtering) [3,4,5] และเทคนิคการกรองข้อมูลแบบหน้าต่างแฮมมิง (Hamming Window Filtering) [6] เป็นต้น

ในการพยากรณ์อากาศจะต้องมีการเก็บข้อมูลเป็นจำนวนมาก เช่น มีการเก็บข้อมูลทุกวัน หรือทุก 4 ชั่วโมง ทำให้ข้อมูลอาจมีความไม่สม่ำเสมอ ในการนำข้อมูลไปพยากรณ์อากาศจึงต้องมีการกรองข้อมูลเพื่อเพิ่มประสิทธิภาพในการพยากรณ์อากาศ ใน

งานวิจัยแบบจำลองการพยากรณ์อากาศโดยใช้โครงข่ายประสาทเทียมด้วยอัลกอริทึม WFNN (Weather Forecast using Neural Networks) [6] ในการพยากรณ์อากาศแบบอนุกรมเวลา โดยใช้เทคนิคการกรองข้อมูลแบบหน้าต่างแฮมมิง ผลการทดลองแสดงให้เห็นว่าข้อมูลที่ผ่านการกรองข้อมูล จะให้ผลการพยากรณ์ที่ถูกต้องกว่าข้อมูลที่ไม่ได้ผ่านการกรอง

สำหรับบทความนี้จะเปรียบเทียบ และวัดประสิทธิภาพของเทคนิคการกรองข้อมูลแบบต่างๆ เพื่อหาเทคนิคการกรองข้อมูลที่เหมาะสมกับการพยากรณ์ข้อมูลอากาศซึ่งในการทดลองได้ใช้ชุดข้อมูลฝนของจังหวัดชลบุรีประเทศไทย และข้อมูลพายุของเมืองริโอเดอจาเนโรและเมืองเซาเปาโลประเทศบราซิล ในส่วนที่ 2 บทความนี้ได้กล่าวถึงเทคนิคการกรองข้อมูลและโครงข่ายประสาทเทียม ส่วนที่ 3 กล่าวถึงแบบจำลองการกรองข้อมูลเพื่อการพยากรณ์อากาศโดยใช้โครงข่ายประสาทเทียม (DFWNN: Data Filtering for Weather Forecast using Neural Networks Model) ส่วนที่ 4 กล่าวถึงผลการทดลอง และ ส่วนที่ 5 คือบทสรุป

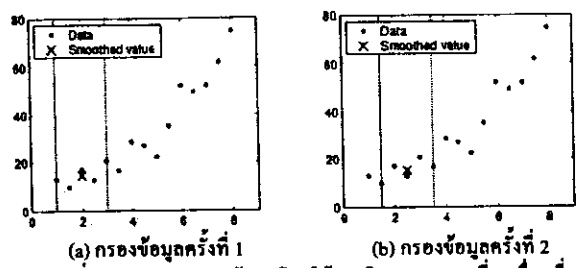
2. เทคนิคการกรองข้อมูลและโครงข่ายประสาทเทียม

2.1 เทคนิคการกรองข้อมูล

ในบทความนี้จะกล่าวถึงเทคนิคการกรองข้อมูล ซึ่งอยู่บนพื้นฐานของการกรองข้อมูลแบบหน้าต่างเคลื่อน (Sliding Window) 4 แบบ คือ การกรองข้อมูลแบบเฉลี่ยเคลื่อนที่ (Moving Average Filtering) การกรองข้อมูลแบบถดถอยท้องถิ่น (LocalRegression Filtering) การกรองข้อมูลแบบสวิสกี-โกเลย์ (Savitzky-Golay Filtering) และ การกรองข้อมูลแบบหน้าต่างแฮมมิง (Hamming Window Filtering) มีรายละเอียดดังต่อไปนี้

2.1.1 เทคนิคการกรองข้อมูลแบบเฉลี่ยเคลื่อนที่ เป็นเทคนิคการกรองข้อมูลอย่างง่ายโดยใช้หลักการหาค่าเฉลี่ยของทุกจุดในหน้าต่างที่ได้กำหนดไว้ การคำนวณหาค่าข้อมูลใหม่ที่ได้ผ่านการกรองแล้ว [1] สามารถพิจารณาได้ดังสมการที่ (1)

$$(y_1)_t = \frac{\sum_{k=-n}^{+n} y_{t+k}}{2n+1} \tag{1}$$



รูปที่ 1. แสดงการกรองข้อมูลโดยใช้เทคนิคการกรองเฉลี่ยเคลื่อนที่ เมื่อ $n = 2$ และขนาดหน้าต่าง คือ $2n+1 = 5$

โดยที่ (y_k) , คือ ข้อมูลค่าแห่งที่ k ที่ผ่านการกรองแล้ว (Smooth Value), y_k คือ ข้อมูลค่าแห่งที่ k และ n คือ จำนวนตำแหน่งของจุดข้อมูลในแต่ละข้าง (ด้านซ้ายหรือด้านขวา) ไปจนถึงขอบหน้าต่าง ขนาดของหน้าต่าง คือ $2n+1$ แสดงดังรูปที่ 1

2.1.2 เทคนิคการกรองข้อมูลแบบถดถอยท้องถิ่น เป็นเทคนิคการกรองข้อมูลที่ใช้วิธีการวิเคราะห์การถดถอย ซึ่งมีการกำหนดน้ำหนักให้ข้อมูลทุกจุด ในหน้าต่างที่กำหนดไว้ โดยใช้ฟังก์ชันน้ำหนักถดถอย (Regression Weight Function) [2] แสดงได้ดังสมการที่ (2)

$$w_i = \left(\frac{1}{2} - \left| \frac{x - x_i}{d(x)} \right|^3 \right)^3 \tag{2}$$

โดยที่ w_i คือ ค่าน้ำหนักถดถอยของจุดที่ i , x คือ ข้อมูลทำนายที่สัมพันธ์กับค่าข้อมูลที่ทำการกรอง, x_i คือ จุดใกล้เคียงที่ใช้ในการกำหนดขนาดของหน้าต่างที่ i และ $d(x)$ คือระยะห่างระหว่างจุด x และจุด x_i

เทคนิคการกรองข้อมูลแบบถดถอยท้องถิ่น สามารถแบ่งเป็น 2 แบบ คือ lowess ใช้การวิเคราะห์การถดถอยแบบเชิงเส้น และ loess ใช้การวิเคราะห์การถดถอยแบบพหุนามกำลังสอง โรบัสต์ฟังก์ชัน (Robust Function) คือ วิธีในการขจัดค่าสุดโต่ง (Outlier) ไม่ให้มีผลกับการกรองข้อมูล สามารถประยุกต์ใช้ร่วมกับ lowess เรียกว่าเทคนิค rlowess และสามารถประยุกต์ใช้กับ loess เรียกว่าเทคนิค rloess ตามลำดับ

2.1.3 เทคนิคการกรองข้อมูลแบบสวิสกี-โกเลย์ เป็นเทคนิคการกรองข้อมูลที่นิยมใช้กับข้อมูลที่มีลักษณะเป็นคลื่นสัญญาณความถี่ โดยใช้เทคนิคการคำนวณแบบพหุนามกำลังสองน้อย



ที่ถูกต้องอย่างง่าย (Simple Polynomial Least-Square Calculations) ซึ่งเทคนิคการกรองข้อมูลแบบนี้สามารถเก็บรายละเอียดเดิมของข้อมูลเอาไว้ได้ดี และจะไม่กรองรายละเอียดของข้อมูลออกมากเกินไป [3,4,5] มีพารามิเตอร์ควบคุมสองค่า คือ ขนาดหน้าต่างและโพลีโนเมียลดีกรี โดยขนาดหน้าต่างต้องมีค่ามากกว่าโพลีโนเมียลดีกรีเสมอ สูตรในการคำนวณข้อมูลที่ผ่านมาการกรองแสดงได้ดังสมการที่ (3)

$$(y_k)_e = \frac{\sum_{i=-n}^n A_i y_{k+i}}{\sum_{i=-n}^n A_i} \quad (3)$$

โดยที่ $(y_k)_e$ คือ ข้อมูลตำแหน่งที่ k ที่ผ่านการกรองแล้ว, y_k คือ ข้อมูลตำแหน่งที่ k , A_i คือ น้ำหนักค่าสัมประสิทธิ์ตำแหน่งที่ i โดยที่ $-n \leq i \leq n$ และ n คือ จำนวนตำแหน่งของจุดข้อมูลในแต่ละข้าง (ด้านซ้ายหรือด้านขวา) ไปจนถึงขอบหน้าต่าง เทคนิคการกรองแบบนี้ได้ถูกนำไปประยุกต์ใช้ในหลายด้านเช่น การลดสิ่งรบกวนในภาพอัลตราซาวด์ และภาพถ่ายเรดาร์ [3,5] เป็นต้น

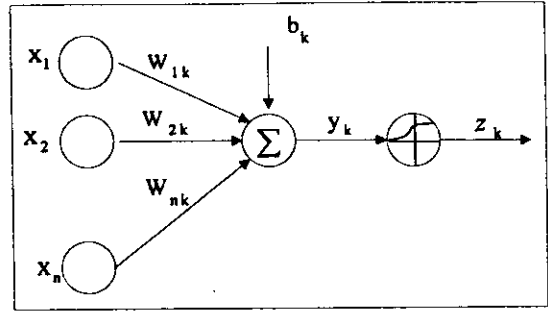
2.1.4 เทคนิคการกรองข้อมูลแบบหน้าต่างแอมมิง เป็นเทคนิคการกรองข้อมูลโดยใช้เทคนิคของการดำเนินการกับสัญญาณ (Signal Processing Technique) ใช้การกรองความถี่แบบอิมพัลส์จำกัด (Finite Impulse Response) [6] แสดงได้ดังสมการที่ (4)

$$\hat{h}(k) = h(k) \cdot w(k) \quad (4)$$

โดยที่ $\hat{h}(k)$ คือ ข้อมูลใหม่ที่ผ่านการกรองแล้ว, $h(k)$ คือ ข้อมูลเดิม และ $w(k)$ คือ ฟังก์ชันหน้าต่างซึ่งเป็นฟังก์ชันที่ใช้ในการลดการกระเพื่อมและปรับปรุงช่วงการตอบสนองการเปลี่ยนแปลง (Transition) ดังสมการที่ (5)

$$w(k) = \begin{cases} 0.54 - 0.5 \cos \frac{2\pi k}{K-1} & ; 0 \leq k \leq K-1 \\ 0 & ; \text{อื่น ๆ} \end{cases} \quad (5)$$

โดยที่ k คือ จำนวนจุดข้อมูล และ K คือ จำนวนจุดข้อมูลของฟังก์ชันหน้าต่าง



รูปที่ 2. แสดงฟังก์ชันการทำงานของหน่วยประมวลผลย่อย

2.2 โครงข่ายประสาทเทียม

เป็นเทคนิคหนึ่งของการทำเหมืองข้อมูล [8,9,10] เพื่อค้นหาความรู้ที่แฝงอยู่ในฐานข้อมูล มีรูปแบบการประมวลผลที่เลียนแบบการทำงานของเซลล์ประสาทของมนุษย์ประกอบด้วยหน่วยประมวลผลย่อยหลายหน่วยเชื่อมต่อกัน รูปแบบการเรียนรู้ของโครงข่ายประสาทเทียม จะเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งจะต้องมีการสอนโครงข่ายประสาทเทียมก่อนนำไปใช้งานจริง มีการแบ่งข้อมูลออกเป็น 2 ชุด คือ ชุดสอนและชุดทดสอบ รูปที่ 2 แสดงฟังก์ชันการทำงานของหน่วยประมวลผลย่อยของโครงข่ายประสาทเทียม ประกอบด้วยฟังก์ชันผลรวม (Summation Function) ในการหาผลรวมของผลคูณระหว่างค่าน้ำหนักข้อมูลเข้ากับค่าข้อมูลเข้า และฟังก์ชันกระตุ้น (Activation Function) เช่น ฟังก์ชันสเตป ฟังก์ชันซายน์ ฟังก์ชันซิกมอยด์ และฟังก์ชันแทนซิกมอยด์ เป็นต้น โครงสร้างของโครงข่ายประสาทเทียมแบบหน่วยประมวลผลย่อยหลายชั้น (Multilayer Perceptron) มี 3 ระดับ คือ ชั้นข้อมูลเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นผลลัพธ์ (Output Layer)

3. Data Filtering for Weather Forecast using Neural Networks Model (DFWFNN)

แบบจำลองการกรองข้อมูลสำหรับการพยากรณ์อากาศ โดยใช้โครงข่ายประสาทเทียมเป็นแบบจำลองเพื่อการพยากรณ์อากาศที่เวลา $t+1$ แบ่งเป็น 4 ชั้นคอน แสดงได้ดังรูปที่ 3

ชั้นคอนที่ 1 เตรียมข้อมูล มีการแบ่งข้อมูลออกเป็นชุดสอนและชุดทดสอบและแทนค่าข้อมูลสูญหาย (Missing Value) ด้วย

ขั้นตอนที่ 1 เตรียมข้อมูล

- 1.1 แบ่งข้อมูลเป็นชุดสอน และชุดทดสอบ
- 1.2 แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยข้อมูลรอบข้าง
- 1.3 กำหนดให้ $X(t) = \{x_1, x_2, \dots, x_n\}$ คือ เซตของตัวแปรเข้าโครงข่ายประสาทเทียมที่เวลา (t) โดย n คือจำนวนแอดทริบิวต์ ของข้อมูลอากาศชุดนั้นๆ
- 1.4 กำหนดข้อมูลผลลัพธ์ (Y) ของโครงข่ายประสาทเทียมที่เวลา t+1

ขั้นตอนที่ 2 การกรองข้อมูลและปรับโดยผล

เทคนิคการกรอง

2.1 กรองข้อมูลโดยเลือกใช้เทคนิคการกรองข้อมูลดังต่อไปนี้

กรณีที่ 1 แบบถลันเคลื่อนที่

1. กำหนดพารามิเตอร์ขนาดหน้าต่าง (Span) = n
2. คำนวณค่าข้อมูลที่กรองแล้วของจุดข้อมูลตรงกลางของหน้าต่างจากสมการ

$$(y_k)_s = \frac{\sum_{i=-n}^{i=n} y_{k+i}}{2n+1}$$

3. เลื่อนหน้าต่างไปทางขวาหนึ่งจุดข้อมูล
4. ทำซ้ำข้อ 2 จนหมดข้อมูล

กรณีที่ 2 แบบถลันถอยท้องถิม

1. กำหนดพารามิเตอร์ขนาดหน้าต่าง (Span) = n
2. กำหนดน้ำหนักให้ข้อมูลทุกจุดในหน้าต่าง จวงสมการ

$$w_i = \left(1 - \left| \frac{x - x_i}{d(x)} \right|^3 \right)^3$$

3. ทำการหาค่าข้อมูลที่กรองแล้ว ใช้การวิเคราะห์การถดถอยแบบเชิงเส้น (lowess) หรือ แบบพหุนามกำลังสอง (loess)
4. เลื่อนหน้าต่างไปทางขวาหนึ่งจุดข้อมูล
5. ทำซ้ำข้อ 2 จนหมดข้อมูล

กรณีที่ 3 แบบสวิตช์-โคเลย์

1. กำหนดพารามิเตอร์ขนาดหน้าต่าง (Span) = n
2. กำหนดโพลีโนเมียลดีกรี = p โดยที่ $p < n$
3. คำนวณค่าน้ำหนักสัมประสิทธิ์จากพารามิเตอร์ทั้งสองตัว
4. คำนวณค่าข้อมูลที่กรองแล้วจากสมการ

$$(y_k)_s = \frac{\sum_{i=-n}^n A_i y_{k+i}}{\sum_{i=-n}^n A_i}$$

5. เลื่อนหน้าต่างไปทางขวาหนึ่งจุดข้อมูล
6. ทำซ้ำข้อ 3 จนหมดข้อมูล

กรณีที่ 4 แบบหน้าต่างแฮมมิง

1. กำหนดพารามิเตอร์ขนาดหน้าต่าง (Span) = n
2. คำนวณค่าข้อมูลที่ผ่านการกรองแล้วจากสมการ

$$\hat{h}(k) = h(k) \cdot w(k)$$

โดยที่

$$w(k) = \begin{cases} 0.54 - 0.5 \cos \frac{2\pi k}{K-1} & ; 0 \leq k \leq K-1 \\ 0 & ; \text{อื่นๆ} \end{cases}$$

3. เลื่อนหน้าต่างไปทางขวาหนึ่งจุดข้อมูล
4. ทำซ้ำข้อ 2 จนหมดข้อมูล

- 2.2 คำนวณเวลาที่ใช้ในการกรองข้อมูล (Optional)
- 2.3 แบ่งข้อมูลออกเป็นกลุ่ม โดยมีช่วงค่าเท่ากัน (Optional)
- 2.4 เปลี่ยนรูปข้อมูลให้อยู่ในช่วง [0,1]

ขั้นตอนที่ 3 การปรับโครงข่ายประสาทเทียม

- 3.1 กำหนดสถาปัตยกรรม โดยชั้นข้อมูลเข้า และชั้นซ่อนมีจำนวนโหนดเท่ากับจำนวนตัวแปรเข้าของข้อมูล และชั้นแสดงผลมี 1 โหนด
- 3.2 ใช้ข้อมูลชุดสอนสอนโครงข่ายประสาทเทียม

ขั้นตอนที่ 4 ทดสอบโครงข่าย

- 4.1 พยากรณ์อากาศที่เวลา t+1 โดยใช้ข้อมูลชุดทดสอบจากการใช้โครงข่ายประสาทเทียมที่ได้จากขั้นตอนที่ 3

รูปที่ 3. แบบจำลองการกรองข้อมูลสำหรับการพยากรณ์อากาศโดยใช้โครงข่ายประสาทเทียม

ค่าเฉลี่ยของข้อมูลรอบข้าง โดยนำค่าก่อนหน้าข้อมูลสูญหาย
บวกกับค่าหลังค่าข้อมูลสูญหายแล้วนำมาหารด้วยสอง

ขั้นตอนที่ 2 วิธีการของข้อมูลและแบ่งข้อมูลออกเป็นช่วง ทั้ง
ข้อมูลชุดสอนและชุดทดสอบ สูตรคำนวณระยะห่างของช่วง
ข้อมูลในแต่ละช่วง สามารถคำนวณได้ดังสมการที่ (6)

$$r = \frac{\max - \min}{q} \quad (6)$$

โดยที่ r คือ ระยะห่างของช่วงในแต่ละกลุ่ม, q คือ จำนวน
กลุ่มที่ต้องการ, \max คือ ค่าข้อมูลสูงสุด และ \min คือ ค่า
ข้อมูลต่ำสุด

ในการเปลี่ยนรูปข้อมูลให้อยู่ในช่วง $[0,1]$ เพื่อให้พร้อม
สำหรับการประมวลผลด้วยโครงข่ายประสาทเทียม สามารถ
คำนวณได้ดังสมการที่ (7)

$$\text{ค่าข้อมูลใหม่} = \frac{\text{ค่าข้อมูลเดิม} - \text{ค่าต่ำสุดของช่วง}}{\text{ค่าสูงสุดช่วง} - \text{ค่าต่ำสุดช่วง}} \quad (7)$$

ขั้นตอนที่ 3 สอน โครงข่ายประสาทเทียม ใช้ข้อมูลชุดสอน
และโครงข่ายประสาทเทียมแบบเพอร์เซพตรอนหลายชั้น และ
ฟังก์ชันกระตุ้นซิกมอยด์ กำหนดสถาปัตยกรรมของโครงข่าย
ประสาทเทียม ชั้นข้อมูลเข้า และชั้นซ่อนมีจำนวน โหนดเท่ากับ
จำนวนตัวแปรเข้าของข้อมูลและชั้นแสดงผลมี 1 โหนด

ขั้นตอนที่ 4 พยากรณ์อากาศ เป็นขั้นตอนที่ทำการพยากรณ์
อากาศที่เวลา $t+1$ โดยใช้ข้อมูลชุดทดสอบจากโครงข่าย
ประสาทเทียมที่ได้จากขั้นตอนที่ 3

4. ผลการทดลอง

ข้อมูลที่ใช้ในการทดลองมี 3 ชุด คือ ข้อมูลฝนจากกรมอุตุนิยมวิทยาประเทศไทยจังหวัดชลบุรี ข้อมูลพายุเมืองริโอเดอเจเนโรและเมืองเซาเปาโล ประเทศบราซิล

4.1 ชุดข้อมูลอากาศประเทศไทยจังหวัดชลบุรี เป็นข้อมูลจากกรมอุตุนิยมวิทยาประเทศไทย [11] จำนวน 17,520 เรคอร์ด (6 ปี) เก็บข้อมูลเป็นราย 3 ชั่วโมง คือ เวลา 01.00น. 04.00น. 07.00น. 10.00น. 13.00น. 16.00น. 19.00น. และ 22.00น. ของแต่ละวัน มีตัวแปรเข้า 7 ตัวแปร คือ x_1 : ปริมาณเมฆ (0 ถึง 10 ส่วนในท้องฟ้า) x_2 : อุณหภูมิจุดน้ำค้าง (องศาเซลเซียส) x_3 : ความกดอากาศ (เฮกโตปาสกาล) x_4 : ความชื้นสัมพัทธ์

ตารางที่ 1. ตัวอย่างชุดข้อมูลฝนของจังหวัดชลบุรี

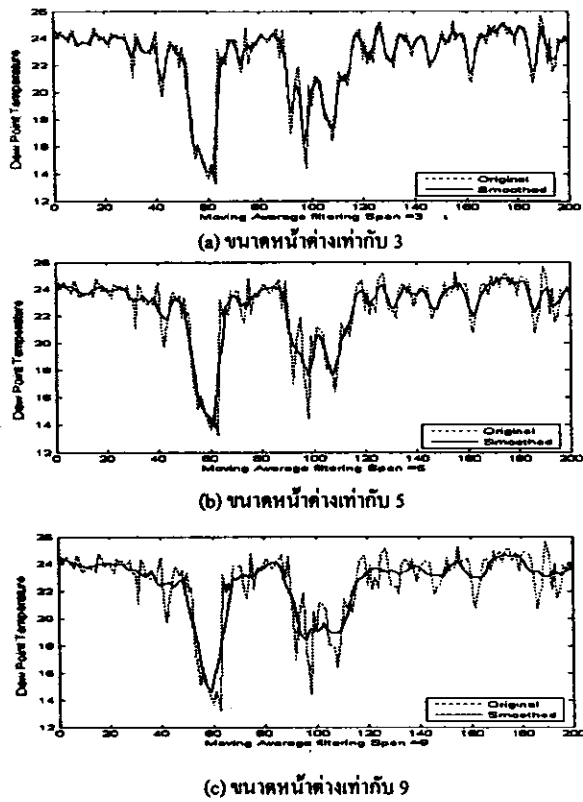
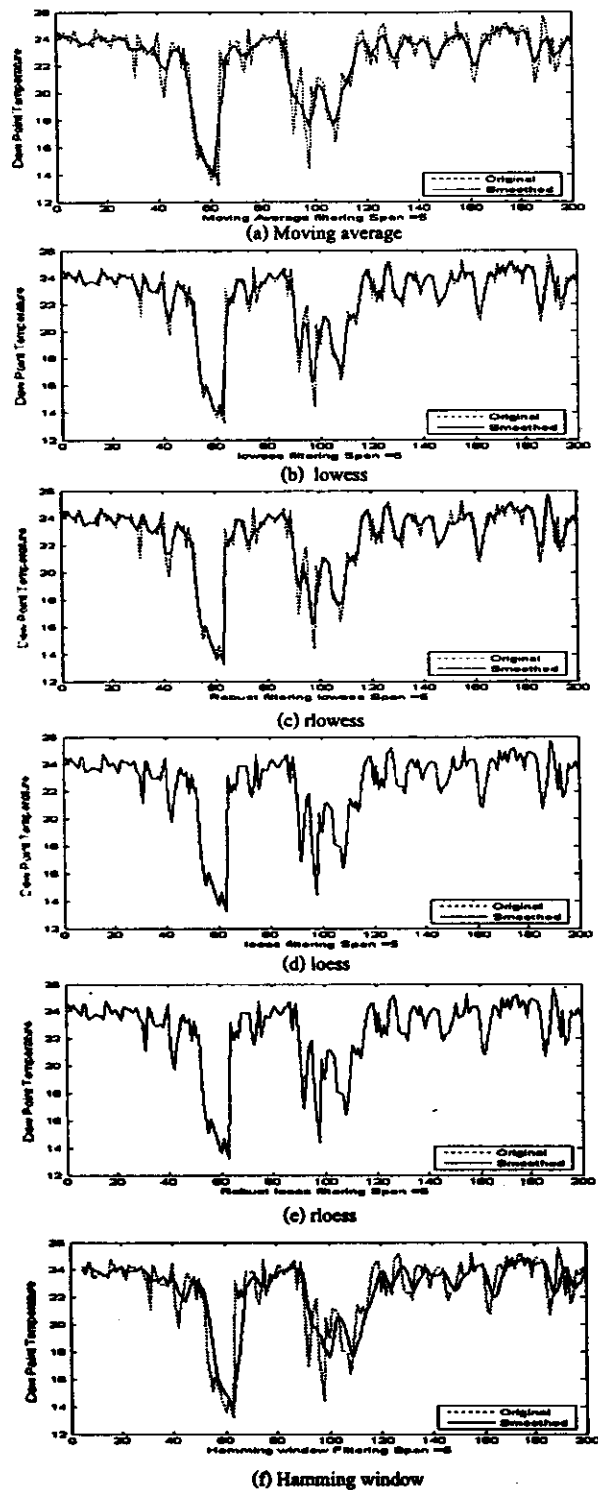
เวลา	x_1 ปริมาณเมฆ	x_2 อุณหภูมิจุดน้ำค้าง	x_3 ความกดอากาศ	x_4 ความชื้นสัมพัทธ์	x_5 อุณหภูมิ	x_6 ความเร็วลม	x_7 ทิศทางลม	Y ปริมาณฝน
01.00	5	17.1	1012.3	41	32	0	0	0
04.00	4	18.7	1011.1	43	33	4	60	5.5
07.00	5	17.7	1010.6	38	34	0	50	0
10.00	2	19.2	1010.4	41	35	2	0	0.1
13.00	2	15.1	1014.7	43	29	6	20	0.7
16.00	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

(เปอร์เซ็นต์) x_5 : อุณหภูมิ (องศาเซลเซียส) x_6 : ความเร็วลม (มอด) x_7 : ทิศทางลม และผลลัพธ์ 1 ตัวแปร คือ y : ปริมาณน้ำฝน (มิลลิเมตร) ตัวอย่างของข้อมูลแสดงได้ดังตารางที่ 1

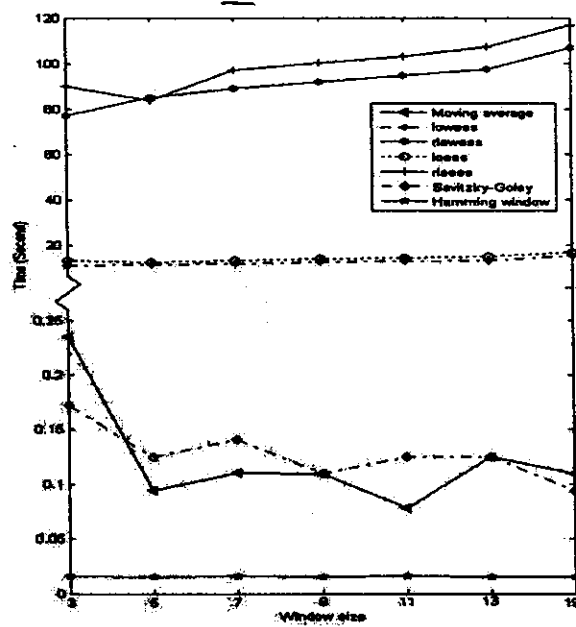
ขั้นตอนที่ 1 แบ่งข้อมูลเป็นชุดสอนจำนวน 14,600 เรคอร์ด (5 ปี) และชุดทดสอบจำนวน 2,920 เรคอร์ด (1 ปี) จากนั้นแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยของข้อมูลรอบข้าง

ขั้นตอนที่ 2 กรองข้อมูลทั้งชุดสอนและชุดทดสอบโดยใช้เทคนิคการกรองทั้ง 4 แบบ รูปที่ 4 แสดงตัวอย่างลักษณะการกรองข้อมูลอุณหภูมิจุดน้ำค้างด้วยเทคนิคต่างๆ โดยกำหนดขนาดหน้าต่างเท่ากับ 5 โดยที่แกน x คือจำนวนข้อมูลอุณหภูมิจุดน้ำค้าง และแกน y คือปริมาณอุณหภูมิจุดน้ำค้าง จะเห็นได้ว่าผลการกรองข้อมูลที่ได้มีลักษณะที่แตกต่างกัน รูปที่ 5 แสดงการกรองข้อมูลแบบเฉลี่ยเคลื่อนที่ที่กำหนดขนาดหน้าต่างในขนาดต่างๆ กัน จะเห็นได้ว่าถ้าขนาดของหน้าต่างมีค่ามาก (เท่ากับ 9) ลักษณะการกรองข้อมูลจะหยาบกว่าขนาดของหน้าต่างที่น้อยกว่า (เท่ากับ 3) รูปที่ 6 แสดงการเปรียบเทียบการใช้เวลาการกรองข้อมูลของแต่ละเทคนิค จะเห็นว่าเทคนิคที่ใช้เวลามากคือ เทคนิคการกรองข้อมูลแบบดลคอดท้องถิ่นแบบ floess และ flowess และเทคนิคที่ใช้เวลาน้อยคือ แบบหน้าต่างสามมิ่ง เป็นต้น รูปที่ 7 แสดงลักษณะการกรองข้อมูลแบบสวิทช์-โกธีย์ ที่โพลีโนเมียลดีกรีต่างกัน

ขั้นตอนที่ 3 เนื่องจากมีตัวแปรเข้า 7 ตัวแปร ดังนั้นโครงข่ายประสาทเทียมจะใช้สถาปัตยกรรมแบบ 7:7:1 คือ มีชั้นข้อมูลเข้า 7 โหนด ชั้นซ่อน 7 โหนด และชั้นแสดงผล 1 โหนดแสดงดังรูปที่ 8 จากนั้นสอนโครงข่ายประสาทเทียมโดยใช้ข้อมูลชุดสอนที่ผ่านการกรองแล้ว

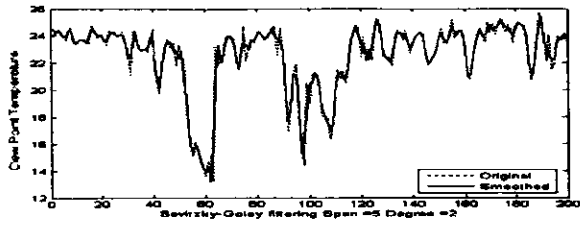


รูปที่ 5. แสดงการกรองข้อมูลที่หน้าต่างขนาดแตกต่างกัน

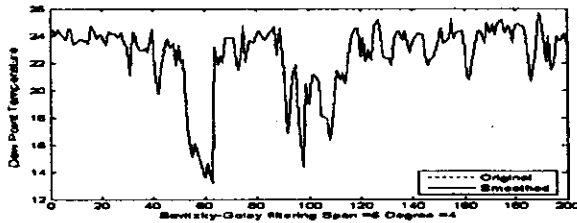


รูปที่ 6. แสดงการเปรียบเทียบเวลาการกรองข้อมูลของแต่ละเทคนิคข้อมูลจังหวัดชลบุรี

รูปที่ 4. แสดงลักษณะการกรองข้อมูลด้วยเทคนิคต่างๆ โดยใช้ขนาดหน้าต่างเท่ากับ 5 ของข้อมูลอุณหภูมิจุดน้ำค้าง จังหวัดชลบุรี

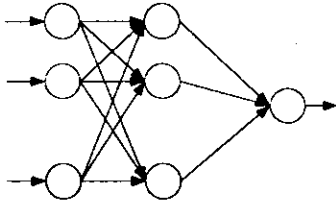


(a) ข้อมูลจากการกรองขนาดหน้าต่าเท่ากับ 5 โพลีโนเมียลดีกรีเท่ากับ 2



(b) ข้อมูลจากการกรองขนาดหน้าต่าเท่ากับ 5 โพลีโนเมียลดีกรีเท่ากับ 4

รูปที่ 7. แสดงรูปการกรองข้อมูลแบบสวิทซ์-โกเลย์ ที่โพลีโนเมียลดีกรีต่างกัน



รูปที่ 8. แสดงโครงข่ายประสาทเทียมสำหรับข้อมูลจังหวัดชลบุรี

ขั้นตอนที่ 4 ทดสอบโครงข่ายประสาทเทียมโดยใช้ข้อมูลชุดทดสอบ เนื่องจากเทคนิคสวิทซ์-โกเลย์ ต้องกำหนดพารามิเตอร์โพลีโนเมียลดีกรี ดังนั้นจึงต้องหาค่าโพลีโนเมียลดีกรีที่เหมาะสม ตารางที่ 2 แสดงค่าความถูกต้องที่เวลา $t+1$ จากการใช้เทคนิคแบบสวิทซ์-โกเลย์ ที่ค่าโพลีโนเมียลดีกรีต่างกัน จะได้ว่าเมื่อค่าโพลีโนเมียลดีกรีสูงขึ้น ทำให้ค่าความถูกต้องสูงขึ้น ตารางที่ 3 แสดงค่าความถูกต้องที่เวลา $t+1$ จากการกรองข้อมูลเพียงอย่างเดียว (Filtering) เปรียบเทียบขนาดหน้าต่าต่างแตกต่างกัน สำหรับเทคนิคสวิทซ์-โกเลย์มีค่าโพลีโนเมียลดีกรีเท่ากับ 4 ตารางที่ 4 แสดงค่าความถูกต้องที่เวลา $t+1$ จากการกรองข้อมูลและแบ่งช่วงข้อมูล (Filtering and Dividing) เปรียบเทียบขนาดหน้าต่าต่างแตกต่างกัน จากผลการทดลองจะพบว่ากรกรองข้อมูลอย่างเดียว เทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องสูงสุด คือ

ตารางที่ 2. ค่าความถูกต้องการทำนายที่เวลา $t+1$ เปรียบเทียบค่าโพลีโนเมียลดีกรีของเทคนิคสวิทซ์-โกเลย์

ขนาดหน้าต่า	โพลีโนเมียลดีกรี						
	2	4	6	8	10	14	18
5	93.44	94.88	-	-	-	-	-
9	94.74	94.98	95.22	95.49	-	-	-
19	93.44	93.44	94.74	94.95	95.05	94.98	95.46

ตารางที่ 3. ค่าความถูกต้องการทำนายที่เวลา $t+1$ จากการกรอง

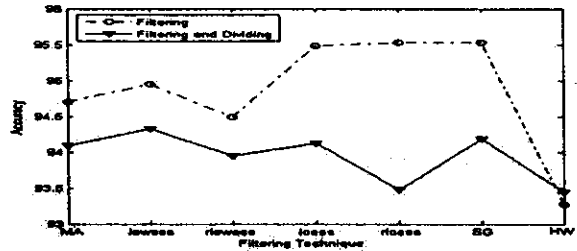
ข้อมูลอย่างเดียว

เทคนิค	ขนาดหน้าต่า						
	3	5	7	9	11	13	19
MA	94.91	94.71	93.95	93.72	93.55	93.48	93.44
lowess	95.46	94.95	95.01	94.84	94.13	93.99	93.72
rlowess	95.46	94.50	94.57	93.95	93.51	93.31	93.41
loess	95.63	95.49	95.08	94.81	94.84	94.84	94.09
rloess	95.39	95.53	94.67	94.64	94.36	94.50	93.44
SG	-	95.53	95.01	94.91	95.15	94.88	93.65
HW	93.44	93.27	93.1	93.44	93.44	93.27	93.27

ตารางที่ 4. ค่าความถูกต้องการทำนายที่เวลา $t+1$ จากการกรอง

ข้อมูลและแบ่งกลุ่มข้อมูล

เทคนิค	ขนาดหน้าต่า						
	3	5	7	9	11	13	19
MA	93.95	94.09	93.44	93.48	93.41	93.41	93.44
lowess	94.36	94.33	93.92	94.23	93.89	93.61	93.44
rlowess	94.50	93.95	93.85	93.44	93.17	93.44	93.44
loess	94.54	94.13	94.30	94.23	94.16	94.02	93.48
rloess	94.33	93.48	93.82	93.61	93.37	93.58	94.13
SG	-	94.19	94.26	94.06	93.68	93.92	93.41
HW	93.44	93.44	93.41	93.41	93.44	93.41	93.44



รูปที่ 9. เปรียบเทียบค่าความถูกต้องเมื่อกำหนดขนาดหน้าต่าเท่ากับ 5 กำหนดโพลีโนเมียลดีกรีเท่ากับ 4 ข้อมูลจากจังหวัดชลบุรี

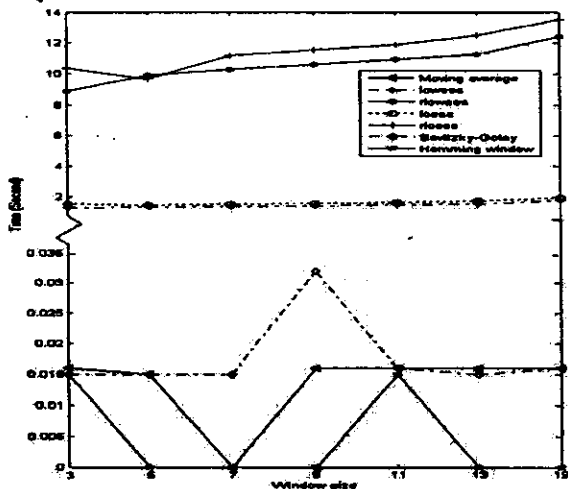
เทคนิคการกรองข้อมูลแบบลดรอยที่อิงกันแบบ loess ที่ขนาดหน้าต่าเท่ากับ 3 ให้ค่าความถูกต้อง 95.63% ส่วนการทดลองที่มีการกรองข้อมูลและแบ่งช่วงข้อมูลพบว่า เทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องสูงสุด คือ เทคนิคการกรองข้อมูลแบบลดรอยที่อิงกันแบบ loess ที่ขนาดหน้าต่าเท่ากับ 5 ให้ค่าความถูกต้อง 94.54% รูปที่ 9 แสดงการเปรียบเทียบค่าความถูกต้องของเทคนิคต่างๆ ระหว่างกรกรองข้อมูลอย่างเดียวกับ

กรองข้อมูลและแบ่งช่วงข้อมูลจากขนาดหน้าต่างเท่ากับ 5 โพลีโนเมียลดีกรีเท่ากับ 4 จะเห็นได้ว่าแบบการกรองข้อมูลอย่างเดียจะได้ค่าความถูกต้องที่สูงกว่า

4.2 ข้อมูลพายุเมืองริโอเดอจาเนโร มีตัวแปรเข้า 10 ตัวแปร ดังนี้ [12] x_1 : อุณหภูมิสูงสุด(ฟาเรนไฮต์) x_2 : อุณหภูมิต่ำสุด(ฟาเรนไฮต์) x_3 : อุณหภูมิจุดน้ำค้างสูงสุด(ฟาเรนไฮต์) x_4 : อุณหภูมิจุดน้ำค้างต่ำสุด(ฟาเรนไฮต์) x_5 : ความชื้นสัมพัทธ์สูงสุด(เปอร์เซ็นต์) x_6 : ความชื้นสัมพัทธ์ต่ำสุด(เปอร์เซ็นต์) x_7 : ความกดอากาศสูงสุด(นิ้ว) x_8 : ความกดอากาศต่ำสุด(นิ้ว) x_9 : ความเร็วลม(ไมล์ต่อชั่วโมง) x_{10} : ปริมาณเมฆ(0 ถึง 10 ส่วนในท้องฟ้า) และผลลัพธ์ 1 ตัวแปร คือ y : ข้อมูลพายุ ซึ่งแทนด้วย 1 (เกิดพายุ) และ 0 (ไม่เกิดพายุ) ตามลำดับ

ขั้นตอนที่ 1 แบ่งข้อมูลเป็นชุดสอนและชุดทดสอบ และแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยของข้อมูลรอบข้าง

ขั้นตอนที่ 2 กรองข้อมูลทั้งชุดสอนและชุดทดสอบ โดยใช้เทคนิคการกรองทั้ง 4 แบบ รูปที่ 10 แสดงเวลาที่ใช้ในการกรองข้อมูลของแต่ละเทคนิคซึ่งผลการทดลองแสดงให้เห็นว่าเทคนิคการกรองข้อมูลที่ใช้เวลาในการกรองสูงคือเทคนิคการกรองข้อมูลแบบดลยทอ้งถิ่นแบบ rloess และ rlowess



รูปที่ 10. แสดงการเปรียบเทียบเวลาการกรองข้อมูลของแต่ละเทคนิค ข้อมูลเมืองริโอเดอจาเนโร

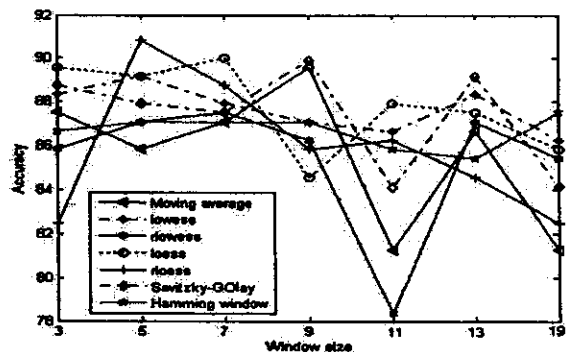
ขั้นตอนที่ 3 เนื่องจากมีตัวแปรเข้า 10 ตัวแปร ดังนั้นโครงข่ายประสาทเทียมจะใช้สถาปัตยกรรมแบบ 10:10:1 คือ มีชั้นข้อมูลเข้า 10 โหนด ชั้นซ่อน 10 โหนด และชั้นแสดงผล 1

โหนด จากนั้นสอนโครงข่ายประสาทเทียมโดยใช้ข้อมูลชุดสอนที่ผ่านการกรองแล้ว

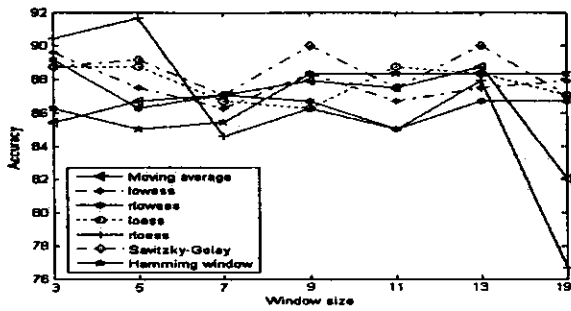
ขั้นตอนที่ 4 นำข้อมูลชุดทดสอบที่ผ่านการกรองแล้วไปทำการพยากรณ์พายุ ตารางที่ 5 แสดงค่าความถูกต้องจากการทำนายที่เวลา $t+1$ จากการใช้การกรองแบบสวิทช์-โกเลย์ที่มีค่าโพลีโนเมียลดีกรีต่างกัน ซึ่งเมื่อค่าโพลีโนเมียลดีกรีสูงขึ้นทำให้ค่าความถูกต้องของข้อมูลสูงขึ้นด้วย รูปที่ 11 แสดงค่าความถูกต้องจากการกรองข้อมูลเพียงอย่างเดียวที่ขนาดหน้าต่างแตกต่างกันกำหนดโพลีโนเมียลดีกรีเท่ากับ 4 (ที่ขนาดหน้าต่างเท่ากับ 3 กำหนดโพลีโนเมียลดีกรีเท่ากับ 2) รูปที่ 12 แสดงค่าความถูกต้องจากการกรองข้อมูลและแบ่งช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน จากผลการทดลองพบว่า การกรองข้อมูลอย่างเดียวนั้นเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องสูงสุด คือเทคนิคการกรองข้อมูลแบบดลยทอ้งถิ่นแบบ rloess ที่ขนาดหน้าต่างเท่ากับ 5 ซึ่งมีค่าความถูกต้องเท่ากับ 90.83% ส่วนการทดลองที่มีการกรองข้อมูลและแบ่งช่วงข้อมูล พบว่าการกรองข้อมูลที่ให้ค่าความถูกต้องสูงสุด คือเทคนิคการกรองข้อมูลแบบดลยทอ้งถิ่นแบบ rloess ที่ขนาดหน้าต่างเท่ากับ 5 มีค่าความถูกต้อง 91.67%

ตารางที่ 5. แสดงการเปรียบเทียบค่าโพลีโนเมียลดีกรีของเทคนิค สวิทช์-โกเลย์ ข้อมูลเมืองริโอเดอจาเนโร

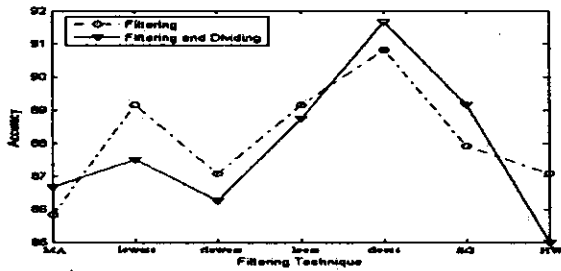
ขนาดหน้าต่าง	โพลีโนเมียลดีกรี					
	2	4	6	8	10	14
5	87.92	88.75	-	-	-	-
9	87.08	87.08	87.92	90.00	-	-
19	86.67	86.67	86.25	86.25	85.42	86.25



รูปที่ 11. ค่าความถูกต้องจากการกรองข้อมูลอย่างเดียว ข้อมูลเมืองริโอเดอจาเนโร



รูปที่ 12. ค่าความถูกต้องจากการกรองข้อมูลและแบ่งช่วงข้อมูล ข้อมูลเมืองริโอเดอจาเนโร



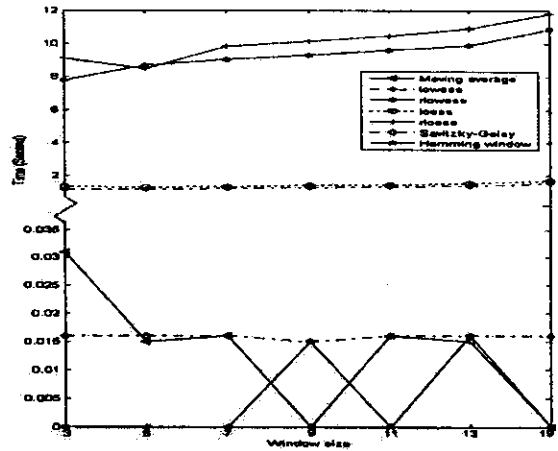
รูปที่ 13. เปรียบเทียบค่าความถูกต้องของการกรองแต่ละเทคนิค ข้อมูลเมืองริโอเดอจาเนโร

รูปที่ 13 แสดงการเปรียบเทียบค่าความถูกต้องของเทคนิคต่างๆ ระหว่างการกรองข้อมูลอย่างเดียวกับการกรองและแบ่งช่วงข้อมูล เมื่อกำหนดขนาดหน้าต่างเท่ากับ 5 และ โพลีโนเมียลดีกรีเท่ากับ 4

4.3 ข้อมูลพายุเมืองเซาเปาโล มีตัวแปรเข้า 10 ตัวแปร [12] ดังนี้ x_1 : อุณหภูมิสูงสุด (ฟาเรนไฮต์) x_2 : อุณหภูมิต่ำสุด (ฟาเรนไฮต์) x_3 : อุณหภูมิจุดน้ำค้างสูงสุด (ฟาเรนไฮต์) x_4 : อุณหภูมิจุดน้ำค้างต่ำสุด (ฟาเรนไฮต์) x_5 : ความชื้นสัมพัทธ์สูงสุด (เปอร์เซ็นต์) x_6 : ความชื้นสัมพัทธ์ต่ำสุด (เปอร์เซ็นต์) x_7 : ความกดอากาศสูงสุด (นิ้ว) x_8 : ความกดอากาศต่ำสุด (นิ้ว) x_9 : ความเร็วลม (ไมล์ต่อชั่วโมง) x_{10} : ปริมาณเมฆ (0 ถึง 10 ส่วนในท้องฟ้า) และผลลัพธ์ 1 ตัวแปร คือ y : ข้อมูลพายุ ซึ่งแทนด้วย 1 (เกิดพายุ) และ 0 (ไม่เกิดพายุ) ตามลำดับ

ขั้นตอนที่ 1 แบ่งข้อมูลเป็นชุดสอนและชุดทดสอบ และแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยของข้อมูลรอบข้าง

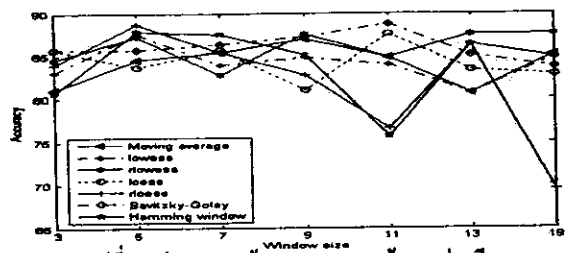
ขั้นตอนที่ 2 กรองข้อมูลโดยใช้เทคนิคการกรองทั้ง 4 แบบ รูปที่ 14 แสดงเวลาการกรองข้อมูลของแต่ละเทคนิค ซึ่งผลการทดลองแสดงให้เห็นว่า เทคนิคการกรองข้อมูลที่ใช้เวลาในการ



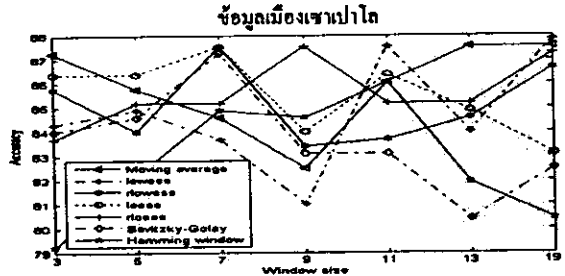
รูปที่ 14. แสดงการเปรียบเทียบเวลาการกรองข้อมูลของแต่ละเทคนิค ข้อมูลเมืองเซาเปาโล กรองสูงคือเทคนิคการกรองข้อมูลแบบลดถอยท้องถิ่นแบบ rloess และ rlowess

ขั้นตอนที่ 3 เนื่องจากมีตัวแปรเข้า 10 ตัวแปร ดังนั้นโครงข่ายประสาทเทียมจะใช้สถาปัตยกรรมแบบ 10:10:1 คือมีชั้นข้อมูลเข้า 10 โหนด ชั้นซ่อน 10 โหนด และชั้นแสดงผล 1 โหนด จากนั้นสอนโครงข่ายประสาทเทียม โดยใช้ข้อมูลชุดสอนที่ผ่านการกรองแล้ว

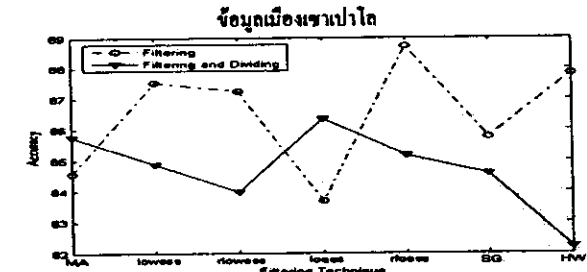
ขั้นตอนที่ 4 นำข้อมูลชุดทดสอบที่ผ่านการกรองแล้วไปพยากรณ์พายุด้วยโครงข่ายประสาทเทียมโดยกำหนดขนาดหน้าต่างเท่ากับ 5 และกำหนดโพลีโนเมียลดีกรีเท่ากับ 4 รูปที่ 15 แสดงผลการทดลองการกรองข้อมูลอย่างเดียวก่อนกำหนดโพลีโนเมียลดีกรีเท่ากับ 4 (ที่ขนาดหน้าต่างเท่ากับ 3 กำหนดโพลีโนเมียลดีกรีเท่ากับ 2) เทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องสูงสุดคือเทคนิคการกรองข้อมูลแบบลดถอยท้องถิ่นแบบ rloess ที่ขนาดหน้าต่างเท่ากับ 5 ซึ่งมีค่าความถูกต้องเท่ากับ 88.72% รูปที่ 16 แสดงผลการทดลองที่มีการกรองข้อมูลและแบ่งช่วงข้อมูล พบว่าการกรองข้อมูลที่ให้ค่าความถูกต้องสูงสุดคือเทคนิคการกรองข้อมูลแบบลดถอยท้องถิ่นแบบ lowess ที่ขนาดหน้าต่างเท่ากับ 19 มีค่าความถูกต้อง 87.83% รูปที่ 17 แสดงการเปรียบเทียบค่าความถูกต้องของเทคนิคต่างๆ ระหว่างการกรองข้อมูลอย่างเดียวกับการกรองและแบ่งช่วงข้อมูล เมื่อกำหนดขนาดหน้าต่างเท่ากับ 5 และ โพลีโนเมียลดีกรีเท่ากับ 4 ของข้อมูลพายุเมืองเซาเปาโล



รูปที่ 15. ค่าความถูกต้องจากการกรองข้อมูลอย่างเคียว



รูปที่ 16. ค่าความถูกต้องจากการกรองข้อมูลและแบ่งช่วงข้อมูล



รูปที่ 17. เปรียบเทียบค่าความถูกต้องของการกรองของแต่ละเทคนิค ข้อมูลเมืองเซาเปาโล

5. บทสรุป

บทความนี้แนะนำเสนอเทคนิคการกรองข้อมูลแบบต่างๆ ที่ใช้กับการกรองข้อมูลอากาศ ผลการทดลองสามารถสรุปได้ดังนี้

- 1) ขนาดหน้าต่างในการกรองชนิดต่างๆ มีผลต่อค่าความถูกต้องของข้อมูล ถ้าขนาดของหน้าต่างมีค่ามากมีผลทำให้ค่าความถูกต้องของการทำนายน้อยลง
- 2) เทคนิคการกรองข้อมูลแบบถดถอยท้องถิ่นให้ค่าความถูกต้องในการทำนายสูงสุด
- 3) เทคนิคการกรองข้อมูลแบบต่างๆ มีเวลาที่ใช้ในการทำงานแตกต่างกัน

เทคนิคที่ใช้เวลาน้อย คือเทคนิคการกรองข้อมูลแบบเฉลี่ยเคลื่อนที่ แบบสวิตซ์-โกเลย์ และแบบหน้าต่างแอมมิง เทคนิคที่ใช้เวลาปานกลาง คือ เทคนิคการกรองข้อมูลแบบถดถอยท้องถิ่นแบบ lowess และ loess เทคนิคที่ใช้เวลามากที่สุด คือ เทคนิคการกรองข้อมูลแบบถดถอยท้องถิ่นแบบ fflowess และ fflowess ดังนั้นในการเลือกใช้การกรองข้อมูล จำเป็นต้องพิจารณา

ถึงขนาดของหน้าต่าง เวลาที่ใช้ และค่าความถูกต้องในการทำนายที่ได้จากเทคนิคต่างๆ

6. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนจากทุนวิจัยคณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์

7. เอกสารอ้างอิง

- [1] Vaiphasa C., "Consideration of Smoothing Techniques for Hyperspectral Remote Sensing", ISPRS Journal of Photogrammetry & Remote Sensing, pp.91-99, 2006.
- [2] Mathwork [online] available : <http://www.mathwork.com/access/helpdesk/help/toolbox/curvefit>
- [3] Chinrungrueng C. and Suvichakorn A., "Fast Edge-Preserving Noise Reduction for Ultrasound Images", IEEE Transactions on Nuclear Science, vol. 48, pp. 849-854, 2001.
- [4] Tarumi T., Gary W. S., Roger J. C., and Robert T. K., "Infinite Impulse Response Filters for Direct Analysis of Interferogram Data from Airborne Passive Fourier Transform Infrared Spectrometry", Vibrational Spectroscopy, vol. 37, Issue 1, pp. 39-52, 2005.
- [5] Chinrungrueng C., "Combining Savitzky-Golay Filters and Median Filters for Reducing Speckle Noise in SAR Images", IEEE International Conference on Man and Cybernetics, vol. 1, pp. 690-696, 2003.
- [6] Wettayaprasit W. and Nanakorn P., "Feature Extraction and Interval Filtering Technique for Time-series Forecasting Using Neural Networks", in Proc. 2006 IEEE International Conferences on Cybernetics and Intelligent Systems (CIS), pp. 635-640, 2006.
- [7] Castellanos J., Pazos A., Rios J. and Zafra J. L., "Sensitivity Analysis on Neural Networks for Meteorological Variable Forecasting", Neural Networks for Signal Processing, 1994.
- [8] Wettayaprasit W. and Lursinsap C., "Neural rule extraction based on activation projection with certainty factor", IEEE International Conference on Neural Networks, pp. 1730-1735, 2002.
- [9] Fu L., "Knowledge Discovery by Inductive Neural Networks", IEEE Transaction on Knowledge and Data Engineering, vol. 11, no. 6, pp. 992-998, 1999.
- [10] Gupta A., Park S. and Lam S. M., "Generalized Analytic Rule Extraction for Feedforward Neural Networks", IEEE Transactions on Knowledge and Data Engineering, vol. 11, no. 6, pp. 985-991, 1999.
- [11] กรมอุตุนิยมวิทยาประเทศไทย. [online]. available:<http://www.tmd.go.th>
- [12] Internet Weather Service. [online]. available : <http://www.wunderground.com>