การเติมข้อมูลฝนที่หายไปจำนวนมากโดยใช้วิธีทางระบบประสาทเทียม การวัดความคล้ายและการประสานเวลาก่อนการประมาณความสัมพันธ์ $Z_e - R$

นางรัชนีวรรณ ตาพุมาศสวัสดิ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2548

ISBN 974-53-2853-7

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

I 2519981x

IMPUTING LARGE MISSING RAIN DATA BY NEURAL NETWORKS, SIMILARITY MEASURE, AND TIME SYNCHRONIZATION PRIOR TO ESTIMATION OF $Z_e - R$ RELATION

Mrs. Rachaneewan Talumassawatdi

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Science
Department of Mathematics
Faculty of Science
Chulalongkorn University
Academic year 2005

ISBN 974-53-2853-7

481996

Thesis Title	IMPUTING LARGE MISSING RAIN DATA BY NEURAL NETWORKS, SIMILARITY MEASURE, AND TIME SYNCHRONIZATION PRIOR TO ESTIMATION OF $Z_e - R$ RELATION
By	Mrs. Rachaneewan Talumassawatdi
Filed of study	Computer Science
Thesis Advisor	Professor Chidchanok Lursinsap, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctor's Degree

.....Dean of the Faculty of Science

(Professor Piamsak Menasveta, Ph.D.)

THESIS COMMITTEE

..... Chairman

(Associate Professor Jack Asavanant, Ph.D.)

..... Thesis Advisor

(Professor Chidchanok Lursinsap, Ph.D.)

..... Member

(Associate Professor Boonserm Kijsirikul, Ph.D.)

..... Member

(Associate Professor Punya Charusiri, Ph.D.)

..... Member

(Waewchark Kongpolporm, Ph.D.)

..... Member

(Siripun Sanguansintukul, Ph.D.)

นางรัชนีวรรณ ตาพุมาศสวัสดิ์ : การเติมข้อมูลฝนที่หายไปจำนวนมากโดยวิธีทางระบบประสาทเทียม, การวัดความคล้าย และการประสานเวลา ก่อนการประมาณความสัมพันธ์ $Z_e - R$ (IMPUTING LARGE MISSING RAIN DATA BY NEURAL NETWORKS, SIMILARITY MEASURE, AND TIME SYNCHRONIZATION PRIOR TO ESTIMATION OF $Z_e - R$ RELATION) อาจารย์ที่ปรึกษา: ศาสตราจารย์ ดร. ชิดชนก เหลือสินทรัพย์, 62 หน้า. ISBN 974-53-2853-7.

ความแม่นยำของการประมาณความเข้มของฝนขึ้นอยู่กับข้อมูลที่สมบูรณ์จากกำลังสะท้อนกลับที่ตรวจวัดด้วยเครื่องตรวจอากาศเรดาร์ แต่อย่างไรก็ตามข้อมูลที่ได้รวบรวมไว้สำหรับการประมาณความสัมพันธ์ของกำลังสะท้อนกลับที่ตรวจวัดได้กับการประมาณความเข้มของฝนในประเทศไทยกำลังประสบปัญหาเกี่ยวกับการมีค่าศูนย์กระจายเป็นจำนวนมาก ซึ่งเรียกว่าข้อมูลที่มีค่าไม่เป็นศูนย์มีน้อย (Sparse Data) ปัญหานี้รุนแรงมากกว่าปัญหาของข้อมูลที่หายไป (Missing Data) ซึ่งมีบางส่วนของข้อมูลที่สมมุติว่าหายไป ในกรณีศึกษานี้ไม่ใช่ว่ามีแค่ข้อมูลที่มีค่าศูนย์เป็นจำนวนมาก แต่ข้อมูลที่มีค่าเป็นศูนย์ยังคงเกิดต่อเนื่องกันเป็นระยะยาว ในกรณีนี้ทำให้วิธีที่ใช้จำลองข้อมูลที่หายไปไม่เหมาะสมที่จะประมาณค่าของข้อมูลที่หายไปได้อย่างมีศักยภาพ เราจึงได้นำเสนอวิธีใหม่เพื่อแก้ปัญหานี้โดยการเติมค่าข้อมูลในข้อมูลที่มีศูนย์จำนวนมากด้วยวิธีการนำความคล้าย (Similarity) ของข้อมูลมาเติมให้สมบูรณ์ แต่ถ้าข้อมูลมีค่าศูนย์ตรงกัน เราจึงใช้วิธีประมาณด้วยค่าคาดหวังมากที่สุด (Expectation Maximization: EM) และใช้ก่อนที่จะประมาณความสัมพันธ์ $Z_e - R$ เพื่อเพิ่มความถูกต้องและความน่าเชื่อถือของผลลัพธ์ที่ได้ การกำหนดเงื่อนไขว่าข้อมูลหายหรือไม่มีฝนได้ถูกเปิดเผยอย่างประสบความสำเร็จก่อนที่จะทำการจำลองข้อมูลที่หายไป ผลการทดลองได้ยืนยันว่าวิธีที่นำเสนอเป็นไปได้ในการประมาณความสัมพันธ์ $Z_e - R$.

ภาควิชา คณิตศาสตร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาการคอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2548 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4473830023 : MAJOR COMPUTER SCIENCE

KEY WORD: Similarity Measure / Missing Data / Incomplete Data / Imputation / Neural Networks / "Missing Data" or "No Rain" Condition.

RACHANEEWAN TALUMASSAWATDI: IMPUTING LARGE MISSING RAIN DATA BY NEURAL NETWORKS, SIMILARITY MEASURE, AND TIME SYNCHRONIZATION PRIOR TO ESTIMATION OF $Z_e - R$ RELATION. THESIS ADVISOR: PROFESSOR CHIDCHANOK LURSINSAP, Ph.D., 62 pp. ISBN 974-53-2853-7

The precision of estimating rain intensity heavily depends upon the completeness of the collected data from radar reflectivity. However, the collected data for estimating the relationships of $Z_e - R$ (measured reflectivity-estimated rain intensity) in Thailand encountered a severe problem of having a large amount of distributed time data with zero values, which is called sparse data. This problem is more severe than the problem of missing data, where only some small portions of data are assumed missing. In our case, not only large numbers of data are zeros; they are also occurred in a long sequence of time. This makes the existing imputing technique inappropriate to estimate the potential value of zero data. A new solution to filling in these sparse data using the similarity of time domain behaviors with expectation maximization (EM) is introduced and used prior to the estimation of $Z_e - R$ relation. To increase the accuracy and reliability of the outcome, determination of "missing" or "no rain" condition has successfully been revealed prior to data imputation, so the experimental results have confirmed that our approach is feasible for this estimation $Z_e - R$.

Department Mathematics Student's signature.....
Field of study Computer Science Advisor's signature.....
Academic year 2005 Co-advisor's signature.....

Acknowledgements

Although, it is not possible to name all those who have been extremely supportive, I would like to acknowledge the support and assistance of several special individuals. Most importantly, I am great indebted to my supervisor, Professor Dr. Chidchanok Lursinsap, at The Advanced Virtual and Intelligent Computing (AVIC) Research Center, for his invaluable advice and great patience in guiding me through my Doctorial program. I would like to thank the Thai Government, National Science and Technology Development Agency (NSTDA) of Thailand and National Electronics and Computer Technology Center (NECTEC), who sponsor the research scholarships. I also would like to thank dissertation committee with their advices and guidance.

In addition, I would like to acknowledge the advice of Dr. Royal Chitradon, Head of The High Performance Computing Laboratory at NECTEC, who encourages me to participate in the program. I owe a great deal to Khun Saneh Warit, Dr. Utai Pison, Dr. Waewchark Kongpolporm and Khun Pairoj Limchamroon, Former Director of Bureau of the Royal Rainmaking and Agricultural Aviation (BRRAA) and Khun Apichai Chingprapa, Current Director of BRRAA, who has given me support and encouragement. Also, thanks my superior officers for their permission to take leave, my colleagues at BRRAA. Moreover, I would like to thank Associate Professor Suchada Siripant, Professor Dr. Frank C. Lin, A. J. Wanmai Jodking for many suggestions and my friends at Chulalongkorn University. I specially thanks Dr. Mansour Ioualalen Geosciences Azur,(CNRS-IRD), Villefranche-sur-Mer, France, for many suggestions in this research.

Finally, I adore my parents, who have been waiting for me to graduate Doctorial program and my family, especially Khun Somchob Talumassawatdi, and my darling daughter, Patcharada Talumasswatdi, who is patient with my lack of time for her.

TABLE OF CONTENTS

Page

Thai Abstract..... iv

English Abstract..... v

Acknowledgements..... vi

List of Tables..... ix

List of Figures..... x

CHAPTER

I INTRODUTION..... 1

1.1 Problem Identification..... 1

1.2 Objectives..... 2

1.3 Scope of Work and Constraints..... 2

II LITERATURE REVIEW..... 4

2.1 Review of Literatures Related to Statistical Inferences of Missing Data..... 4

2.2 Review of Literatures Related to Neural Inferences of Missing Data..... 6

2.3 Review of Literatures Related to Radar Rainfall Estimates..... 7

III THEORETICAL BACKGROUND..... 10

3.1 An Overview of Neural Computing..... 10

3.1.1 A Back-propagation Artificial Neural Network..... 11

3.2 Weather Radar Fundamentals..... 14

3.2.1 Mie and Rayleigh scattering..... 16

3.3 Convective and Stratiform Precipitation..... 16

3.3.1 Precipitation Theories..... 16

3.3.2 Precipitation Processes..... 17

IV PROBLEM DEFINITION..... 19

4.1 Problem Definition..... 19

4.2 Determination of "Missing" or "No Rain" Conditions..... 22

4.2.1 Problem Formulation..... 22

4.2.2 Designing Threshold..... 23

V PROPOSED SOLUTION..... 25

5.1 Overview of Proposed Solution..... 25

5.2 Similarity Measure..... 28

5.3 The Expectation Maximization (EM) Algorithm..... 29

VI EXPERIMENTAL RESULTS..... 32

6.1 Data Collection..... 32

6.2 Preprocessing of Training and Testing Data..... 32

6.3 Imputing Efficiency and Robustness..... 33

6.4 Testing Accuracy..... 37

6.4.1 Testing Determination "Missing" or "No Rain" Condition Accuracy..... 37

6.5 Enhancing Reliability in Radar Rainfall Estimates..... 40

6.5.1 Algorithm for Similarity Alignment..... 41

VII CONCLUSION..... 46

REFERENCES..... 48

CURRICULUM VITAE.....51

LIST OF TABLES

Table **Page**

4.1 A result of a Difference of Radar Reflectivity Thresholds..... 24

6.1 Imputation Efficiency and Robustness among Similarity Measure (SM), Expectation Maximization (EM) and Neural Network (NN) for Gauge no. 071 and 081..... 33

6.2 Imputation Efficiency and Robustness among Similarity Measure (SM), Expectation Maximization (EM) and Neural Network (NN) (Gauge no. 062, 063 & 073)..... 35

6.3 An example of a piece of manifold of imputing large missing rain data..... 36

6.4 Efficiency of Classification of "Missing" or "No Rain" Conditions using Neural Network for Gauge No. 071 and 081..... 38

6.5 Efficiency of Classification of "Missing" or "No Rain" Conditions using Neural Network for Gauge No. 062, 063 and 073..... 39

LIST OF FIGURES

Figure **Page**

3.1 A two-layer feed forward network, showing the notation for units and weights..... 12

4.1 The WSR-88D radar site in Omkoi Chiangmai, Thailand corresponding to rain gauge location..... 20

4.2 The altitude categorization of a difference of rain gauge location corresponding to monthly total gauge rain intensity.....21

5.1 An example of a difference between convective and stratiform rainfall type..... 26

5.2 An example of convective rainfall type having high correlation..... 27

6.1 An example of similarity manifold matching which is having maximum Similarity Measure (SM) *before* applied similarity alignment algorithm.....42

6.2 An example of similarity manifold matching which is having maximum Similarity Measure (SM) *after* applied similarity alignment algorithm.....42

6.3 A result of 16:00-18:00 hours daily monthly of a difference Z-R parameter at cloud base with radar reflectivity threshold value of 24 dB..... 43

6.4 A result of 16:00-18:00 hours daily monthly of a difference Z-R parameter at cloud base with radar reflectivity threshold value of 30 dB..... 43

6.5 A result of 16:00-18:00 hours daily monthly comparison of radar-gauge rainfall accumulation for Gauge no. 071..... 44

6.6 A result of 16:00-18:00 hours daily monthly comparison of radar-gauge rainfall accumulation for Gauge no. 081..... 44