

تقدير القيم المفقودة لبيانات السلسلة الزمنية
لتلوث الهواء باستخدام طرائق هجينة للشبكات العصبية
المعاودة والمضبية

رسالة تقدّم بها

مؤمن عامر حسن الصائغ

إلى

مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل وهي جزء من

متطلبات نيل شهادة الماجستير

علوم في الإحصاء

بإشراف

المدرس الدكتور

أسامة بشير شكر

المستخلص

تم في هذه الرسالة تقدير القيم المفقودة باستخدام أنموذج الانحدار الخطي المتعدد $\text{Multiple linear regression (MLR) model}$ كطريقة تقليدية إحصائية شائعة الاستخدام وكذلك الحصول على نمذجة إحصائية لبيانات الدراسة. يعدّ أنموذج MLR أحد النماذج الخطية ولهذا السبب فقد تكون نتائج تقدير القيم المفقودة قليلة الدقة أحياناً. ولتحسين نتائج تقدير القيم المفقودة تم اقتراح أسلوب الشبكات العصبية المعادة $\text{Recurrent neural network (RNN)}$ لاستخدامها في تقدير القيم المفقودة بعد تهجينها مع أنموذج MLR عبر ما يسمى بالطريقة الهجينة MLR-RNN . كذلك تم اقتراح أسلوب النظام العصبي الضبابي الاستدلالي المكيف $\text{Adaptive neuro-fuzzy inference system (ANFIS)}$ واستخدامه في تحسين نتائج تقدير القيم المفقودة بعد تهجينه مع أنموذج MLR عبر ما يسمى بالطريقة الهجينة MLR-ANFIS . وقد استخدمت في الرسالة بيانات تلوث الهواء من خلال ما يسمى بالجسيمات المعلقة $\text{Particular Matter (PM}_{10})$. أثبتت الطرائق المهجنة MLR-RNN و MLR-ANFIS دقة في نتائج تقدير القيم المفقودة مقارنة مع الطريقة التقليدية. وكذلك أثبتت نتائج تقدير القيم المفقودة باستخدام الأسلوب الهجين MLR-ANFIS دقتها العالية مقارنة بالأسلوب الهجين MLR-RNN . ان اهم الاستنتاجات التي توصلت لها الدراسة امكانية استخدام الأساليب الهجينة المقترحة مع بيانات الجسيمات المعلقة والملوثات الجوية للحصول على دقة مثلى في نتائج تقدير القيم المفقودة.

**Missing Values Estimation for Air Pollution
Time Series Datasets by Using Hybrid Methods
for Recurrent Neural Networks and
Neuro-Fuzzy**

A Submitted Thesis

Moamin Amer Hasan AL-Saeegh

To

The Council of the College of Computers Sciences and
Mathematics University of Mosul

In Partial Fulfillment of the Requirements for
a Degree of Master of Science in Statistics

Supervised by

Dr. Osamah Basheer Shukur

Abstract

In this thesis, missing values have been estimated by using multiple linear regression (MLR) model as common classical statistical method and to obtain a statistical modelling of studied datasets. MLR model is one of linear models. For this reason, the results of missing values estimation may be inaccurate. For improving the results of missing values estimation, recurrent neural network (RNN) is proposed after hybridizing with MLR model through MLR-RNN hybrid method. Adaptive neuro-fuzzy inference system (ANFIS) is also proposed for more improving of the results of missing values estimation after hybridizing with MLR through hybrid MLR-ANFIS method. In this thesis, air pollution datasets through particular matter (PM₁₀) have been used. MLR-RNN and MLR-ANFIS confirmed good accuracy in the results of missing values estimation comparing to the classical method. MLR-ANFIS confirmed also higher accuracy comparing MLR-RNN method in missing values estimation. As most important conclusion, the hybrid methods can be used with PM₁₀ and climatic pollutants to obtain the optimal results of missing values estimation.