Semidefinite Programming-Based Method for Implementing Linear Fitting to Interval-Valued Data

Minghuang Li, Beijing Normal University, China

Fusheng Yu, Beijing Normal University, China

ABSTRACT

Building a linear fitting model for a given interval-valued data set is challenging since the minimization of the residue function leads to a huge combinatorial problem. To overcome such a difficulty, this article proposes a new semidefinite programming-based method for implementing linear fitting to interval-valued data. First, the fitting model is cast to a problem of quadratically constrained quadratic programming (QCQP), and then two formulae are derived to develop the lower bound on the optimal value of the nonconvex QCQP by semidefinite relaxation and Lagrangian relaxation. In many cases, this method can solve the fitting problem by giving the exact optimal solution. Even though the lower bound is not the optimal value, it is still a good approximation of the global optimal solution. Experimental studies on different fitting problems of different scales demonstrate the good performance and stability of our method. Furthermore, the proposed method performs very well in solving relatively large-scale interval-fitting problems.

Keywords: Fuzzy Regression, Interval Least Squares, Interval Regression, Semidefinite Programming, Semidefinite Relaxations

INTRODUCTION

The objective of linear regression is to adjust the parameters of a linear model function so as to best fit a given data set. In the real world, the observations and measurements are usually subject to imprecision and vagueness. In order to assess such uncertainty one ideal way is to represent the data as intervals or fuzzy sets.

The linear regression analysis involving fuzzy data is the central concern of fuzzy regression. Tanaka (1982) first proposed a fuzzy linear regression model by minimizing the index of fuzziness of the system. Diamond (1990, 1994, 1997) introduced a metric on the set of fuzzy numbers and used this metric to define a least-sum-of-squares criterion function as in the usual sense. Further discussion can be found in (Chang, 2001; Coppi et al., 2004; D'Urso, 2003, 2006; Tanaka et al., 1982, 1987, 1996; Sakawa

et al., 1992; Ma et al., 1997; Chang et al., 2001; Wu, 2003) and other references therein.

At the interval level, the similar topic has also been discussed extensively in the field of symbolic data analysis (Bock et al., 2000). Bertrand and Goupil (2000) and Billard and Diday (2000, 2003) introduced statistics methods based on covariance and correlation function suitable for interval-valued data. A centre and range method was developed in Lima Neto et al. (2008) for fitting interval data. Various forms of support vector interval regression approaches have been discussed (An et al., 2008; Jeng et al., 2003; Hong et al., 2005, 2006). Additionally, in practical context, this model has shown its potential in forecasting interval-valued time series (Maia et al., 2008).

The model studied in this article is fitting a crisp linear function to interval-valued input-output data. More precisely, the model consists of minimizing the sum of squared interval distances between observed and predicted values. The predicted values are calculated by interval arithmetic (or the extension principle).

However, as far as we know, solving this fitting problem in high dimensional case is still an open problem. The main difficulty is that 2^n different combinations of the signs of n regression coefficients have to be treated separately. Hence the complexity of the problem will increase extremely.

Such combinatorial difficulty was also mentioned in Diamond et al. (1997), where the author discussed the fitting model involving minimizing the sum of squared L_2 distances between observed and predicted values represented by fuzzy numbers.

Theoretically, the problem involving minimization of interval distance may lead to a hard problem, but we have found that this problem, in many cases, can be exactly solved in polynomial time. It thus makes it possible for this type of 'distance-based' fitting model to handle high dimensional problems.

The method proposed in this paper consists of two main steps: casting the interval fitting problem as a quadratically constrained quadratic programming (QCQP) problem, and using the relaxation theory to obtain a lower bound of the optimal value by solving a semidefinite programming problem. It is worthy of emphasizing that in many situations, the lower bound is exactly equal to the global minimum.

Considering the fact that the lower bound may not always be the global minimum, we also present a correcting step to enhance the accuracy of the solution by solving a quadratic programming problem. The combination of the two procedures performs very well in practical setting, especially in dealing with relatively large scale problems.

The remainder of this paper is organized as follows. First we provide the preliminary knowledge and notations used throughout this paper. Next, we give the QCQP form of an interval fitting problem. Two formulas for solving the optimization problem based on two types of relaxations are presented. Finally, we show the experiment results of solving small scale $(n = 2)$ and large scale $(n = 200)$ fitting problems and conclude this study.

PRELIMINARIES

In this section, we give some notations used in this paper and briefly introduce the related preliminary knowledge.

Notation

Throughout this paper, the following notations are used:

For a vector x, $x \geq 0$ $(x > 0)$ means element-wise $x_k \geq 0$ $(x_k > 0)(k = 1, 2, ..., n)$.

For matrix A, its range, pseudo-inverse, transpose, and trace are denoted by:

$\mathcal{R}(A)$, $A^{\dagger}$, A^t and $\mathrm{tr}(A)$ respectively.

For two matrices A and B, $A \geq B$ $(A > B)$ means $A - B$ is positive semidefinite (positive definite).

Related Content

Subtractive Clustering and Particle Swarm Optimization Based Fuzzy Classifier

Halima Salah, Mohamed Nemissi, Hamid Seridiand Herman Akdag (2019). *International Journal of Fuzzy System Applications (pp. 108-122).*

www.irma-international.org/article/subtractive-clustering-and-particle-swarm-optimization-based-fuzzy-classifier/233589

A Fuzzy Logic-Based Method for Incorporating Ethics in the Internet of Things

Sahil Sholla, Roohie Naaz Mirand Mohammad Ahsan Chishti (2021). *International Journal of Ambient Computing and Intelligence (pp. 98-122).*

www.irma-international.org/article/a-fuzzy-logic-based-method-for-incorporating-ethics-in-the-internet-of-things/279587

Policy Guidelines Post ChatGPT Era in Education, Research, and Public Administration: A Document Analysis

(2023). *Artificial Intelligence Applications Using ChatGPT in Education: Case Studies and Practices (pp. 149-157).*

www.irma-international.org/chapter/policy-guidelines-post-chatgpt-era-in-education-research-and-public-administration/329839

Construction of an Ensemble Scheme for Stock Price Prediction Using Deep Learning Techniques

Justice Kwame Appati, Ismail Wafaa Denwar, Ebenezer Owusuand Michael Agbo Tettey Soli (2021). *International Journal of Intelligent Information Technologies (pp. 1-24).*

www.irma-international.org/article/construction-of-an-ensemble-scheme-for-stock-price-prediction-using-deep-learning-techniques/277073

Behavioral Implicit Communication (BIC): Communicating with Smart
Environments via our Practical Behavior and Its Traces