Chapter 12
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.

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 regres-
The problem of fitting a crisp linear function to interval-valued input-output data has been studied extensively. The model consists of minimizing the sum of squared interval distances between observed and predicted values. This can be achieved by interval arithmetic (or the extension principle).

However, as far as we know, solving this problem in high dimensional cases 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:
Related Content

Learning Fuzzy Network Using Sequence Bound Global Particle Swarm Optimizer
www.igi-global.com/article/learning-fuzzy-network-using-sequence/63355?camid=4v1a

Improving Trustworthiness in E-Market Using Attack Resilient Reputation Modeling
www.igi-global.com/article/improving-trustworthiness-in-e-market-using-attack-resilient-reputation-modeling/116743?camid=4v1a

Extracting and Customizing Information Using Multi-Agents
www.igi-global.com/chapter/extracting-customizing-information-using-multi/24346?camid=4v1a

Multi-Agent Negotiation in B2C E-Commerce Based on Data Mining Methods
www.igi-global.com/chapter/multi-agent-negotiation-b2c-commerce/64382?camid=4v1a