Chapter 12
Fuzzy Integral–Based Kernel Regression Ensemble and Its Application

Yulin He
Shenzhen University, China

Yanxing Hu
The Hong Kong Polytechnic University, Hong Kong

James N. K. Liu
The Hong Kong Polytechnic University, Hong Kong

Xizhao Wang
Shenzhen University, China

ABSTRACT

Similar to ensemble learning for classification, regression ensemble also tries to improve the prediction accuracy through combining several "weak" estimators that are usually high-variance and thus unstable. In this chapter, the authors propose a new scheme of fusing the weak Priestley-Chao Kernel Estimators (PCKEs) based on Choquet fuzzy integral, which differs from all the existing models of regressor fusion. The new scheme uses Choquet fuzzy integral to fuse several target outputs from different PCKEs, in which the optimal bandwidths are obtained with cross-validation criteria. The key of applying fuzzy integral to PCKE fusion is the determination of fuzzy measure. Considering the advantage of Particle Swarm Optimization (PSO) algorithm on convergence rate, the authors use three different PSO algorithms (i.e., Standard PSO [SPSO], Gaussian PSO [GPSO], and GPSO with Gaussian Jump [GPSOGJ]) to determine the general and λ fuzzy measures. The experimental results on the standard testing functions and practical Fourier Transform Infrared Spectroscopy (FTIR) datasets show that the new paradigm for regression ensemble based on fuzzy integral is more accurate and stable in comparison with any individual PCKE and the Basic Ensemble Method (BEM). This demonstrates the feasibility and effectiveness of the proposed regression ensemble model.

INTRODUCTION

Ensemble learning (Zhou, 2012; Zhang & Ma, 2012) is a fusion strategy which tries to make the final decision by integrating the multiple feedbacks from different base-learners so as to reduce the decision maker's variance and improve its robustness and accuracy. That is to say a strong learner will be produced by organizing some weak ones in a proper way. Commonly, these weak learners are integrated through the majority voting for classification and a linear combination for regression (Brown et al., 2005). In recent years, the ensemble learning for classification has been well studied. There are a number of classical works which introduce the ensemble strategies for different classifiers, e.g., boosting or bagging based ensembles for decision trees (Banfield et al., 2007), neural networks (Zhou et al., 2002) and support vector machines (Kim et al., 2003), et c. However, just as said in (Moreiraa et al., 2012), the successful ensemble learning approaches for classification techniques are often not directly applicable to regression. Thus, unlike the sophisticated ensemble methods for classification, the regression ensemble often uses the weighted or ordered weighted average of base-learners to conduct the prediction, where several different methods are developed to determine the weights (Moreiraa et al., 2012).

The weighted average and ordered weighted average operators are good choices to deal with the different importance of individual base-regressor, but these two methods are under an assumption that interaction does not exist among the individual regressors. However, this assumption may not be true in many real problems. If the interaction is considered, fuzzy integrals (Grabisch, 1995(b); Cho & Kim, 1995) may be a better choice. The fuzzy integral as a fusion tool, in which the non-additive measure can clearly express the interaction among regressors and the importance of each individual regressor, has its particular advantages. Motivated by the definition of fuzzy integral which can be considered as a mechanism of maximizing the potential efficiency of base-regressor, we construct a new approach for regression ensemble based on fuzzy integral in this chapter.

One difficulty for applying fuzzy integrals in regressor fusion is how to determine the fuzzy measures. There are some methods to determine fuzzy measures such as Gradient Descent (GD) (Grabisch, 1995(a)), Genetic Algorithm (GA) (Yang et al., 2008; Ganesan et al., 2011(b)), Neural Network (NN) (Wang & Wang, 1997; Ganesan et al., 2011(a)), etc. Although using GD, GA and NN to determine the fuzzy measures is successful to some extent, there exist many limitations in the application process. For example, GD and NN frequently fall into the local minimum, and GA is much slower. It is necessary to find some new computational techniques for determining fuzzy measures. In 2011, Wang et al. (Wang et al., 2011) proposed Particle Swarm Optimization (PSO) based fuzzy measure determination, where PSO is a kind of swarm intelligence optimization algorithm (Ganesan et al., 2012; Ganesan et al., 2013(a); Ganesan et al., 2013(a)). The theoretical analysis and experimental comparison demonstrated the superior performance of PSO based methods. Thus, we use PSO to determine the fuzzy measure in fuzzy integral based regression ensemble in this chapter. The main contributions of this chapter can be summarized as 1) using fuzzy integral to construct the regression ensemble and 2) applying three different PSOs (i.e., Standard PSO-SPSO (Kennedy & Eberhart, 1995), Gaussian PSO-GPSO (Krohling, 2004) and GPSO with Gaussian Jump-GPSOGJ (Krohling, 2005)) to determine the general and λ fuzzy measures. In our study, a kind of kernel regressor, i.e., Priestley-Chao Kernel Estimator-PCKE (Priestley & Chao, 1972), is selected as the base-regressor. We call the Kernel Regression Ensemble model based on Fuzzy Integrals with the general and λ fuzzy measures $KREFI_g$ and $KREFI_\lambda$ respectively.

Related Content

Man in the Browser Attacks

Timothy Douganand Kevin Curran (2012). *International Journal of Ambient Computing and Intelligence (pp. 29-39).*

www.irma-international.org/article/man-browser-attacks/64189

Role of Distance Metric in Goal Geometric Programming Problem (G^2 P^2) Under Imprecise Environment

Payel Ghoshand Tapan Kumar Roy (2019). *International Journal of Fuzzy System Applications (pp. 65-82).*

www.irma-international.org/article/role-of-distance-metric-in-goal-geometric-programming-problem-g2-p2-under-imprecise-environment/214940

Evaluation of Quality of the e-Commerce Service

Manuela Ingaldiand Robert Ulewicz (2018). *International Journal of Ambient Computing and Intelligence (pp. 55-66).*

www.irma-international.org/article/evaluation-of-quality-of-the-e-commerce-service/205576

AI in Market Research: Transformative Customer Insights - A Systematic Review

Manisha Paliwaland Nishita Chatradhi (2024). *Exploring the Intersection of AI and Human Resources Management (pp. 231-255).*

www.irma-international.org/chapter/ai-in-market-research/336269

Quantum Computing: Unveiling the Paradigm Shift and Diverse Applications

R. Siva Subramanian, B. Maheswari, T. Nithya, P. Girija, M. Karthikeyanand T. Saraswathi (2024). *Applications and Principles of Quantum Computing (pp. 95-112).*

www.irma-international.org/chapter/quantum-computing/338285