Expert Systems with Applications 41 (2014) 3955–3964

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification

Jui-Sheng Chou *, Min-Yuan Cheng, Yu-Wei Wu, Anh-Duc Pham

Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan, ROC

ARTICLE INFO

Keywords:
Classification model
Hybrid intelligence
Optimization
Support vector machine
Fast messy genetic algorithm
Dispute prediction
Project management

ABSTRACT

Hybrid system is a potential tool to deal with construction engineering and management problems. This study proposes an optimized hybrid artificial intelligence model to integrate a fast messy genetic algorithm (fmGA) with a support vector machine (SVM). The fmGA-based SVM (GASVM) is used for early prediction of dispute propensity in the initial phase of public–private partnership projects. Particularly, the SVM mainly provides learning and curve fitting while the fmGA optimizes SVM parameters. Measures in term of accuracy, precision, sensitivity, specificity, and area under the curve and synthesis index are used for performance evaluation of proposed hybrid intelligence classification model. Experimental comparisons indicate that GASVM achieves better cross-fold prediction accuracy compared to other baseline models (*i.e.*, CART, CHAID, QUEST, and C5.0) and previous works. The forecasting results provide the proactive-warning and decision-support information needed to manage potential disputes.

1. Introduction

Data mining (DM) has attracted great scientific interest and has become an important research area. Major DM methods include predictive modeling (supervised learning in machine learning, *i.e.*, classification and regression problems), clustering and association (unsupervised learning), evolution, pattern matching, data visualization and meta-rule guided mining (Liao, Chu, & Hsiao, 2012). In particular, classification is one of the important missions in data mining (Ngai, Xiu, & Chau, 2009). The DM- and AI-based approaches are related to computer system programs that attempt to resolve problems by emulating human brain processes (Garg, Rani, & Sharma, 2014; Hajdasz, 2014; Irani & Kamal, 2014; Khashei, Zeinal Hamadani, & Bijari, 2012). Therefore, the use of AI-based models is a potential tool to deal with classification problems in construction engineering and management.

Notably, a global trend has become interested for financial public investment *via* Public Private Partnership (PPP) which is a financial strategy for stimulating private investments in public works. PPP have proved to be a useful and beneficial instrument. However, PPP project involve a variety of organizations and a large number of partnerships as well as exist high risk and uncertainty. Dispute between parties to projects are of great concern to the construction industry (Fenn, Lowe, & Speck, 1997; Tang, Shen, & Cheng, 2010). These problems are also influenced by highly variable and unpredictable factors. Because of these difficulties and the importance of enhancing forecast capability, algorithms with complexity approaching that of integrated models have been developed to improve modeling accuracy, effectiveness and speed (Chou & Lin, 2013; Donis-Díaz, Muro, Bello-Pérez, & Morales, 2014; Pai, Hung, & Lin, 2014; Seera & Lim, 2014; İlhan & Tezel, 2013). However, enhancing the generational capability of advanced DM techniques is still in need for the PPP-related projects.

Taiwan has legally supported PPP projects for more than ten years. The Taiwan Public Construction Commission (TPCC) has actively promoted and encouraged private-sector participation in infrastructure and building construction throughout Taiwan. According to the TPCC, the dispute rate was 23.6% during 2002–2009. The most common processes for handling disputes are mediation or non-mediation (*e.g.*, arbitration, litigation, negotiation, and administrative appeals) procedures. In Taiwan, up to 84% of PPP projects are settled by mediation or negotiation within only 1–9 months after disputes occurs (PCC., 2010). Notably, arbitration or litigation costs all parties considerably more time and money when a mediated agreement cannot be reached. To address these challenges, the proposed classification approaches predict propensity for project claims providing supportive information needed by governmental authority to furnish contact documents in the preparatory and bidding phases of PPPs.

Moreover, the dispute between PPP participantes commonly occur unexpectedly and may involve many issues, including surety bond issue, sub-contractor qualifications, licenses, permits, investment scale, resident rights, government guarantees, excessive

* Corresponding author. Address: Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, 43, Sec. 4, Keelung Rd., Taipei 106, Taiwan, ROC. Tel.: +886 2 2737 6321; fax: +886 2 2737 6606.

E-mail address: jschou@mail.ntust.edu.tw (J.-S. Chou).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.12.035

profits, operating period, taxation, and default loan commitment (Jones, 2006). Numerous studies show that an efficient, effective, and fair dispute resolution process is essential for PPP project success (Cheung, 1999; Cheung, Suen, & Lam, 2002; Cheung, Yiu, & Chan, 2010; Chou & Lin, 2013; Goetz & Gibson, 2009; Menassa & Peña Mora, 2010). Therefore, development of intelligence models can enable early warming of potential dispute resolutions prior to project initiation to be becoming crucial.

Several studies have attempted to develop the hybrid AI models (Hsiao, Wang, Wang, Wen, & Yu, 2012; Irani & Kamal, 2014; Khashei et al., 2012; Lee, Rajkumar, Lo, Wan, & Isa, 2013; Seera & Lim, 2014) by combining one with other techniques to enhance their performance results, such as combination of genetic algorithm and support vector machine (Fei & Zhang, 2009; Huang & Wang, 2006; Jiao & Liu, 2011; Zhao, Fu, Ji, Tang, & Zhou, 2011; İlhan & Tezel, 2013). Particularly, the mechanism for setting tuning parameters in AI models is an important issue that is widely recognized by scholars in many different disciplines (Donis-Díaz et al., 2014; J. Huang, Bo, & Wang, 2011; Pai et al., 2014; İlhan & Tezel, 2013). In practice, identifying the best set of parameters for a model is an optimization problem. Therefore, integrating AI-based models with nature-inspired algorithms has been proposed as a solution to the above problems.

Advanced AI-based approaches include support vector machine (SVM), which is a machine learning technique with advanced features that enable good generalization and fast computation. The SVM has been demonstrated very powerful in solving classification problems (Lin, 2002). Although the SVM model has proven highly effective for solving classification problems, it has major drawbacks (Jiao & Liu, 2011). The accuracy of the SVM model depends on parameters set in advance. For such need, fast messy genetic algorithms (fmGA)-based method developed by Goldberg, Deb, Kargupta, and Harik (1993) is known for its flexibility in allowing hybridization with other methodologies to obtain enhanced solutions (David E. Goldberg et al., 1993). The primary difference between an fmGA and other genetic algorithms is its ability to manipulate explicit building blocks of genetic material to obtain global optimization (Hettiarachchi, Noman, & Iba, 2013; Wu, 2005). The fmGA can efficiently find the optimal solution of the large-scale permutation problems (Knjazew, 2002). These advantages make fmGA logical candidates for overcoming the disadvantages of SVM.

The aim of this study, thus, is to employ an auxiliary hybrid technique which SVM will call fmGA as subroutine to optimize its structure parameters. The fmGA-based support vector machine (GASVM) enhance their efficacy to early predict PPP dispute likelihood and potential resolutions, thereby alleviating the future adverse effects of disputes on project delivery, operation, and transfer from a governmental prospective. To demonstrate the efficacy of proposed hybrid system, this study used PPP project data collected by TPCC and compared well-known classification and regression-based models (*e.g.*, CART, CHAID, QUEST, and C5.0) using classification performance measure in term of accuracy, precision, sensitivity, specificity, area under the curve and synthesis index. For avoid bias from data (Kohavi, 1995), cross-fold validation was also executed.

The rest of this study is organized as follows. Section 2 thoroughly reviews AI literature and the application of AI to predict construction claims and litigation outcomes. Section 3 characterizes the research methodology, providing a theoretical basic for classification performance models and elaborates on the GASVM model proposed in this study. Section 4 describes and analyzes numerical example and analytical outcomes using classification performance measures. Conclusions are given in Section 5, along with directions for further research.

2. Literature review

The hybrid AI and forecasting techniques are widely used in various engineering and management fields (Hsiao et al., 2012; Huang & Wang, 2006; Kim & Kang, 2012; Kim & Shin, 2007; Moosmayer, Chong, Liu, & Schuppar, 2013), and have been demonstrated to enhance overall performance. Nevertheless, their effectiveness and efficiency are rarely applied in the construction industry, particularly in public–private partnership (PPP) project domain. To support the dispute resolution process, several varieties of tools and systems have been developed (Chong, Mohamad Zin, & Chong, 2012; El-Adaway & Kandil, 2010; Ilter, 2012; Jin & Zhang, 2011; Kassab, Hegazy, & Hipel, 2010; Seifert, 2005). Applying these techniques is useful for both researchers and practitioners to better understand the complex nature of PPP project. Since a dispute always takes in numerous complex and interconnected factors that are difficult to rationalize, using DM techniques is now among the most effective methods for improving prediction accuracy related to some cases such as construction litigation (Chau, 2006; Chau, 2007; Pulket & Arditi, 2009a; Pulket & Arditi, 2009b); construction procurement litigation (Arditi & Pulket, 2010); and change-order-triggered disputes (Chen, 2008; Chen & Hsu, 2007).

Arditi and Tokdemir (1999) have developed several models on the same dataset using AI techniques to enhance prediction result in conventional construction procurement litigation as a prediction accuracy of 83.33% was achieved with a case-based reasoning (Arditi & Tokdemir, 1999), 89.95% was achieved with boosted decision trees (Arditi & Pulket, 2005), and 91.15% was attained with integrated prediction modeling (Arditi & Pulket, 2010). Moreover, their studies have attempted to minimize the number of construction litigation cases by using neural network to predict the likely court rulings and obtained a prediction rate of 67% for litigation outcomes (Arditi, Oksay, & Tokdemir, 1998). They argued that if outcome of construction litigation can be predicted with higher accuracy and reliability by using these approaches, all parties involved in the construction process could save considerable money, time, and aggravation.

Furthermore, Chau (2006) found that excluding the above case studies, AI techniques are rarely applied in the legal field (Chau, 2006). Thus, Chau utilized AI techniques based on particle swarm optimization (PSO) to predict construction litigation outcome, a filed in which relatively new DM techniques are rarely applied. The PSO-based ANN technique developed by Chau obtained the rate of prediction is up to 80%, which is much higher than chance. However, Chau suggested using additional case factors related to cultural, psychological, social, environmental, and political features in the future work.

The other AI techniques were used for construction dispute. In case of changing orders of construction process and design, Chen (2008) developed a *K* nearest neighbor based knowledge sharing model, which obtained 84.38% accuracy in predicting lawsuits based on a nationwide study of US court records (Chen, 2008). Chen and Hsu (2007) further employed hybrid ANN-CBR model with disputed change order dataset to achieve early-warm information. The ANN classifier demonstrated comparable prediction accuracy (84.61%) (Chen & Hsu, 2007). In case of dispute settlement, Cheng, Tsai, and Chiu (2009) refined and improved the conventional CBR approach by combining fuzzy set theory with a new similarity measurement that integrates Euclidean distance and cosine angle distance (Cheng et al., 2009). Their model successfully extracted the knowledge and experience of experts from 153 historical construction dispute cases collected manually from multiple sources.

Several studies have demonstrated that hybrid AI schemes generate promising results in many industries (Chen & Hsu,