A multi-objective evolutionary algorithm-based soft computing model for educational data mining

A distance learning experience

Choo Jun Tan, Ting Yee Lim and Chin Wei Bong

School of Science and Technology, Wawasan Open University, Penang, Malaysia, and

Teik Kooi Liew

Department Registry, Directorate of Quality Assurance and External Relations, Wawasan Open University, Penang, Malaysia

Abstract

Purpose – The purpose of this paper is to propose a soft computing model based on multi-objective evolutionary algorithm (MOEA), namely, modified micro genetic algorithm (MmGA) coupled with a decision tree (DT)-based classifier, in classifying and optimising the students' online interaction activities as classifier of student achievement. Subsequently, the results are transformed into useful information that may help educator in designing better learning instructions geared towards higher student achievement.

Design/methodology/approach – A soft computing model based on MOEA is proposed. It is tested on benchmark data pertaining to student activities and achievement obtained from the University of California at Irvine machine learning repository. Additional, a real-world case study in a distance learning institution, namely, Wawasan Open University in Malaysia has been conducted. The case study involves a total of 46 courses collected over 24 consecutive weeks with students across the entire regions in Malaysia and worldwide.

Findings – The proposed model obtains high classification accuracy rates at reduced number of features used. These results are transformed into useful information for the educational institution in our case study in an effort to improve student achievement. Whether benchmark or real-world case study, the proposed model successfully reduced the number features used by at least 48 per cent while achieving higher classification accuracy.

Originality/value – A soft computing model based on MOEA, namely, MmGA coupled with a DT-based classifier, in handling educational data is proposed.

Keywords Classification, Optimization, Student achievement, Educational data mining, Multi-objective evolutionary algorithm

Paper type Research paper

Introduction

For decades, educators have been troubled by questions such as "Does study habits correlate to test scores achievements?", "Why educational attainment, especially on higher educational completion rate or student retention rate is so difficult to achieve?" (Atkinson and Geiser, 2009; Nandeshwar *et al.*, 2011), and "What is needed to improve the quality of education?" (Langstrand *et al.*, 2015; Willman *et al.*, 2015). We argue that by inspecting the variation in observational data on predictors (independent variables) and

outcomes (dependent variables), our understanding of the relationship inter-played among different factors, for instance, student-teacher interactions (Allen *et al.*, 2011) and teacher quality (Kraft, 2015) can lead to intervention instruction design that improves academic outcomes. An educational system is a complex system which comprises the pedagogy, human assets (e.g. students and teachers), supportive tools (e.g. course materials and infrastructure), social and cultural influence, and government policies, among others. At the same time, the advent of technologies, in particular the World Wide Web (as known as the internet) has also given rise to new waves in the educational domain.

In this context, distance learning spurred by the emergent of the internet, has become one of the key components in higher education. It offers adult learners the opportunity to pursue their educational dreams without the barrier of distance. Distance learning practises the delivery of course materials without needing both student and teacher to be in the same physical room. The convenience is usually made possible by computer-mediated learning or two-way interactive videos (Tan, Bong and Natarajan, 2015). As such, pedagogic strategies designed to match the new level of instructions with the level of distance learners are rapidly gaining attention among modern educators.

Additionally, the internet has also opened up new opportunities for open education in the form of open educational resources (OER). OER include books, videos, journals, articles, podcasts, lesson plans, open software, and so on at the discretion of openness and sharing (Smith, 2009). Consequently, the increase in e-learning resources and student databases leads to massive repositories of data. Although there has been considerable research on the use of data mining (DM) techniques in discovering potentially useful information from large sets of data in the fields of healthcare (Wang *et al.*, 2013), manufacturing (Durán *et al.*, 2010), market analysis (Fiol-Roig and Miro-Julia, 2011), and many others; only recently has researchers begun to apply them to issues of educational realm through a discipline of known specifically as the educational data mining (EDM). The plausible revelation comes from the realisation that given any context, additional hidden information may be revealed if one invests in trying. EDM involves the development, research, and application of computerised methods in detecting potential patterns from huge pools of educational data (Romero *et al.*, 2010).

With the aggressive push towards sophisticated technology and boundary-less education, more and more factors interact within the already complex educational systems. Unfortunately, as remarked by the *Science* journal (Smith, 2009), we are to some extent still lacking of the understanding about the effectiveness of value-added components (e.g. OER) introduced to the educational arena. Many educational research works employ EDM techniques that rely on statistical methods (with exception a few exceptions). However, these traditional ad hoc combinations of data management tools and statistical methods are now far from adequate in analysing the vast pool of data. Therefore, it is time to seek new computing approach that can be applied to the highly non-linear, complex, and large volume of data environment. Moreover, the approach ought to be able to handle incompleteness and shortage of data. Fortunately, soft computing lends the solution. Soft computing is a collection of methodologies capable of handling imprecision and uncertainty with the aim of achieving tractable, robust, yet low-cost solutions. Such characteristics of soft computing models make them invaluably attractive when applying to DM applications.

To give an immediate example, soft computing method, in particular the genetic programming complements existing educational research by providing better insights into how student participation affects achievement (Xing *et al.*, 2015). Furthermore, empirical research using DM technique to classify early dropout from selective Mexican high schools have been carried out with success (Márquez-Vera *et al.*, 2016). Yet another recent application of DM technique (i.e. cluster analysis) in exploring the correlation between the students' online interaction patterns and achievement is reported in Cerezo *et al.* (2016). While these approaches certainly have their strengths, the generic applications are

contested when handling multi-objective optimisation problem (MOPs). In principle, MOPs require simultaneous optimisation of conflicting objectives. As in the classification of student achievement model, one would consider improving the classification accuracy rate of the predictive results using the least numbers of available features (e.g. demographic data, social data, past examination grades, and other school-related data) from a pool of data since it is often true that the latter more difficult to obtain or even if they do, may not be in complete set. Therefore, a robust model should hold high accuracy rate at the presence of uncertainty or data incompleteness.

This paper presents a soft computing technique, which comprises an evolutionary algorithm (EA), i.e., modified micro genetic algorithm (MmGA) (Tan, Lim and Cheah, 2013), coupled with a decision tree (DT)-based classifier, namely, C4.5 for the classification and optimisation of system. The MmGA works well in MOP context as shown by a series of previous successes (Tan, Lim and Cheah, 2013; Tan *et al.*, 2014; Tan, Lim, Cheah and Tan, 2013; Tan, Hanoun and Lim, 2015; Tan *et al.*, 2017). In order to evaluate the proposed soft computing model, an empirical-based case study is conducted. In particular, we aim to answer the first research question raised at the beginning of the paper. That being said, the case study attempts to uncover possibly appealing activities of students' online interaction against test score achievement in an open distance learning institution in Malaysia. Student performance modelling is noted as the second most popular areas in EDM research (Peña-Ayala, 2014). The implications of the present work can be translated into helpful information that may assist educators in designing instructions suitable for students customisable to their learning behaviours so as to improve achievement. Kremer *et al.* (2013) share the similar view that technology can be used to tailor learning to student's level of knowledge.

The remaining of the paper is organised as follows: the second section covers the background review of EDM and EA; followed by a short overview of C4.5 classifier. Third section is devoted to the description of our proposed methodology. The proposed model is first verified using a set of benchmark problems. Next, a case study for demonstrating the true efficacy of our model is described in fourth section. Finally, the paper concludes with some promising avenues for future work in fifth section.

Preliminaries

Review on EDM and learning analytics (LA)

In recent years, the potential of DM and analytics has transformed field after field. DM is the process of analysing and gleaning useful but hidden information from huge data sets (Mukhopadhyay *et al.*, 2014a, b). DM is popular due to its ability to discover data patterns, classify objects, cluster homogeneous objects, and unveil numerous kinds of new findings (Peña-Ayala, 2014). In the educational domain, a specific form of the DM is known as EDM. EDM emerges as an approach to explore educational massive data in enhancing the educational sector. It leverages the data mined through DM to improve learning, cognition, and assessment (Sachin and Vijay, 2012).

On the other hand, LA refers to the use of learner-data in reporting and analysing models for the purpose of predicting and advising learners (Ferguson, 2012; Hwang *et al.*, 2017). Though LA is commonly identified with EDM, they differ in terms of goals and scopes. Baker and Inventado (2014) contrast between EDM and LA in a recent review: from the technical perspective, EDM deals with the development of methodologies for analysis of learning data, while LA focusses on the interpretation of these data for optimising learning and its environment. EDM also emphasises the modelling of relationships among specific constructs, whereas LA relates the interplay among constructs from a holistic view of the system. Additionally, EDM research works concentrate mostly on the development of automated support for learners. LA, on the other hand, puts more effort in informing and empowering instructors about learners' performance progress.

To date, EDM has taken on and extended many other related fields including text mining, machine learning, statistics, and psychometrics. Romero *et al.* (2013) propose a prediction model of student performance using DM techniques such as clustering and class association rules. The model has been claimed to be more representative of student groups (clusters) compared to previous rule-based model. EDM is also popularly used for student modelling (Lemmerich *et al.*, 2011; Nandeshwar *et al.*, 2011), which aims at characterising student by emotion, achievements, skills, learning preferences, and fulfilling individual's learning requirements through adaptation of teaching experiences. Another area of EDM application is on student assessment and evaluation, which enables student proficiency to be distinguished at a fine-grained level (Lopez *et al.*, 2012). EDM also facilitates student feedback and support (Leong *et al.*, 2012). More generally, EDM can be applied to educational problems with regards to emotion in context, engagement, meta-cognition, and collaboration tasks (Baker, 2014).

From the lens of LA, students' engagement and learning outcomes can be improved with proper intervention by instructors. To be effective, the intervention should be provided at the right time. LA and DM techniques are helpful in this case. A commonly used approach for discovering sequential patterns among events is known as sequential pattern mining. Chen *et al.* (2017) adopt the frequent sequent mining and lag sequential analysis (LSA) in order to study how learners collaborate in knowledge-building discourse. Similarly, LSA facilitates the exploration of learners' sequential patterns in other learning settings such as online interactions behaviour (Cheng *et al.*, 2017) and problem-solving behaviour (Chiang, 2017; Hu *et al.*, 2017). It is also not uncommon to adopt LA in analysing learners' behavioural patterns as a result of interaction with strategies or technological tools. For instance, Kizilcec *et al.* (2017) investigate various self-regulated learning strategies in MOOCs environment in hope of uncovering the most effective strategies and how they manifest in online behaviour. Meanwhile, Van Leeuwen *et al.* (2014) examine how teacher supporting tools in the context of computer-supported collaborative learning affect teacher guidance behaviour.

In general, more and more educators are turning to both EDM and LA for improving the educational outcomes. As shown by Xing *et al.* (2015), synthesising LA approaches and EDM techniques supplemented by genetic programming produces an effective student performance prediction model. The model has been claimed to possess higher prediction rate and interpretability compared to traditional models. Whether EDM or LA, educators can continue to benefit from the various scientific and systematic analysis methodologies available.

Review on EAs and multi-objective optimisation
Natural evolution provides a promising collection of inspirations for computational algorithms. The group of computing methodologies, which analogises the evolutionary process of biological population in finding optimal solutions to optimisation problems, is known as EA (Golberg, 1989). Generally, EA can be divided into four major classes: genetic algorithm (Holland, 1992), evolutionary programming (Fogel, 1966), evolution strategies (Schwefel, 1993), and genetic programming (Koza, 1992).

Unlike traditional methodologies, EAs are distinguished mainly by the use of a population of search space. Each member of the population corresponds to a potential solution. The quality of the solution is determined by a fitness value associated with each member. During each iteration step (generation), better fitness members receive higher chances of survival or become the parents of the next generation. Offspring which are the new population members are generated using some variation operations, like mutation and/ or crossover. The evolutionary process ends after some termination criteria are met. These synergetic combinations of population-based, fitness-based, and variation-driven search have reported success in many complex optimisation problems (Tan, Lim and Cheah, 2013; Lim *et al.*, 2015a, b, 2016; Tan *et al.*, 2017). Meanwhile, the literature of GA runs

a long list of variance diverging from its original, yet maintaining the novelty of GA characteristics. Among the more popular ones are the micro-GA, monogamous GA (Lim *et al.*, 2015a, 2016), island model GA, and cellular GA, to name but a few[1].

Many real-world problems are made up of performance measures (objectives) that are often conflicting in interest. They ought to be optimised simultaneously in order to achieve a trade-off. In this light, a special domain of the EA that deals with MOPs is known as the multi-objective evolutionary algorithm (MOEA). In any MOP, it is not surprising that a set of optimal solutions (as opposed to single optimum) is obtained. The optimal solution set usually consists of a number of solutions that are close in fitness according to Pareto dominance concept. As a result, comparing among the different optimal solution sets is a challenging task (Jiang *et al.*, 2014). Various quantitative performance metrics exists in the literature of MOP for defining the optimality of different solution sets. These included the generational distance (GD) metric (Durillo and Nebro, 2011), generalised spread metric (Zhou *et al.*, 2006), and hypervolume metric (Zitzler and Thiele, 1999).

In the meantime, the MOEA can be broadly classified into aggregation-based, indicator-based, and Pareto-based approaches. The aggregation-based approaches treat MOP as single-objective optimisation problem that can then be solved using conventional EAs after combining all its objective functions into a single weighted scalar value. However, the major shortcoming of this approach is that the scalar function and weights are critical in determining the efficiency of the algorithm. However, finding suitable weights is an optimisation problem in itself. On the other hand, the indicator-based MOEAs typically adopt selection mechanism with specific performance metric (Zitzler and Künzli, 2004; Beume *et al.*, 2007). They have the advantage of being scalable to the number of objectives, usually four or more. However, they are generally more computationally expensive, especially when using hypervolume metric.

Finally, a representative Pareto-based MOEA approach is the MmGA (Coello and Pulido, 2005). MmGA is also an extension of the micro-GA. It has been used with great success in handling various multi-objective benchmark problems (Tan, Lim and Cheah, 2013), job-shop scheduling problems (Tan, Hanoun and Lim, 2015; Tan *et al.*, 2017) as well as classification problems (Pourpanah *et al.*, 2017). Even though the MmGA uses only a small size population relative to the other GA variants, it is able to achieve good convergence rate (see third section for more details). As such, this work employs the MmGA as an optimisation means. The MmGA uses GD as its performance metric.

From the DM perspective, MOEAs are popular underlying optimisation solutions for a variety of DM tasks, enumerated as clustering (Kirkland *et al.*, 2011; Ripon and Siddique, 2009), association rule mining (Matthews *et al.*, 2011; Martín *et al.*, 2011), classification (Tan, Lim, Cheah and Tan, 2013; Tan *et al.*, 2014; Pangilinan and Janssens, 2011; Pourpanah *et al.*, 2017), and feature selection (Tan *et al.*, 2014; Venkatadri and Rao, 2010; Brester *et al.*, 2014). For a complete review of the various MOEAs for DM, interested reader is referred to Mukhopadhyay *et al.* (2014a, b).

Review on C4.5 classifier
This section provides a quick overview of the C4.5 classifier, which is commonly used for generating a DT. First and foremost, a DT is a tree-like structure composed of decision rules. These rules regulate the grouping of independent variables into homogeneous zones in recursion (Cho and Kurup, 2011). DT is commonly used in acquiring information for decision making. This is in conjunction with the observation that by constructing a DT, the outcome of a set of input variables can be predicted simply by finding the set of decision rules (Pradhan, 2013). In fact, DT has been ranked as the second most popular classification methods in EDM in a recent survey conducted in Peña-Ayala (2014).

Even though there exist a plethora of DT model constructing algorithms, for instance, the chi-square automatic interaction detector DT (Michael and Gordon, 1997) and classification and regression tree (Breiman *et al.*, 1984), this paper focusses on the use of C4.5 classifier (Quinlan, 1986) for reason of simplicity and wide applications.

C4.5 is an extended algorithm to the ID3 (Quinlan, 1986), which is based upon the Hunt's algorithm (Hunt and Kübler, 1984). It addresses many problems that were not accounted for by its predecessor, including continuous and categorical attributes, pruning, and rule derivation. In C4.5 algorithm, a DT is built from a set of training data, $S = s_1, s_2, \ldots$. Each sample s_i is made up of n-dimensional vector $(x_{1, i}, x_{2, i}, \ldots, x_{p, i})$, where $x_{k,i}$ refers to the sample features or attribute values of class s_i. When encountering continuous attributes, the algorithm simply divides the attribute values into two partitions as specified by a given threshold. In order to remove any bias of information gain, especially when an attribute has many outcome values, the C4.5 algorithm relies on gain ratio as its selection measure. Starting from the highest information gain attribute, the algorithm recurs to smaller sub-lists. In this way, the root node has the maximum gain ratio. The higher information gain attribute will be chosen for decision making (Quinlan, 1993).

Proposed model
In this work, a soft computing model to classify and optimise students' online behaviours in a distance learning environment is presented. Students' online behaviours as characterised by a set of web data, forms the input to our proposed model. The web data represents the frequency of students' interactions with courses within the distance learning environment. Our aim is to classify students' frequency of access to the learning repository against their examinations achievement at the end of a semester. Followings are elaboration of the proposed model.

Initially, a standard C4.5 classifier (Quinlan, 1993, 1996) is applied. It uses a divide-and-conquer approach to growing DTs from a set C of cases. Suppose that C fulfils a stopping criterion of decision making. The tree of C is a leaf associated with the most frequent target class in C, which contains only cases of the similar target class. Meanwhile, the proportion of cases in X of jth class is identified. The uncertainty about the class for a case of X, and its corresponding information gained by a test T with k outputs are computed.

Next, a specific MOEA, namely, the MmGA (Tan, Lim and Cheah, 2013) is deployed. The MmGA performs optimisation on two objective functions, i.e., maximising the classification accuracy rate (α) and minimising the number of features (β) of classification process. Note that α describes the systematic errors and measures the statistical bias in handling predictors and outcomes of C4.5 classifier processes. As articulated earlier (recall section "Review on EAs and Multi-objective Optimisation"), the MmGA is able to achieve good convergence rate as indicated by the GD metric. MmGA's search process terminates when objective functions has reached the maximum round of evaluation or achieved convergence as measured by true Pareto. Details on C4.5 classifier as well as objective functions α and β with relation on MmGA are presented in the Appendix.

Benchmark tests
The proposed model aims to yield a solution set, which fulfils the objective functions f_1 and f_2 such that the classification accuracy rate is maximised, while minimising the number of features during the classification stage. Prior to application on a real-world case study, we first examined the proposed model's performance on a set of benchmark data obtained from the University of California at Irvine (UCI) machine learning repository (University of California, 2017). The benchmark data set comprises students' achievements in mathematics and Portuguese in two Portugal secondary schools. The data attributes include student

grades, demographic, social, and school-related features. They were collected by using both school reports and questionnaires as published in Cortez and Silva (2008).

Note only mathematical achievements, which were modelled as binary classification, but five-level classification and regression tasks were adopted in this study. We adhere to the original performance evaluation of Portugal education. That is, students are evaluated in three periods during the school year based on a 20-point grading scale (with values between $0 =$ lowest score to $20 =$ perfect score). Hence, the data set is split into three classes according to period grade, i.e., first period grade (G1), second period grade (G2), and final grade (G3). As a result, each newly created data set has originally 30 features, which correspond to variable x in the Equation (A5) for each target class G1, G2, or G3, separately. To begin, the collected grades for each class were binarised prior to classification processing: student grades were re-categorised into two groups, namely, well performed (those above or equal to score 8) and not well performed (those below score 8).

For comparison purposes, the proposed model first uses only a standard C4.5 classifier (note: in the remaining of this paper, we merely refer this model by C4.5 classifier). Subsequently, an enhanced model which incorporates the MmGA coupled with standard C4.5 classifier is deployed. It is coined as the MmGA-based classifier. The MmGA analogises the evolutionary process of biological population in finding optimal solutions for MOP. In this context, by maximising α (Equation (A3)) and minimising β (Equation (A4)). Each member of the population corresponds to a potential solution, which is created with MmGA extended population formation. We also employ a ten-fold cross-validation method in producing the experimental results. All experiments involving both methods are repeated over 30 runs with randomised seed.
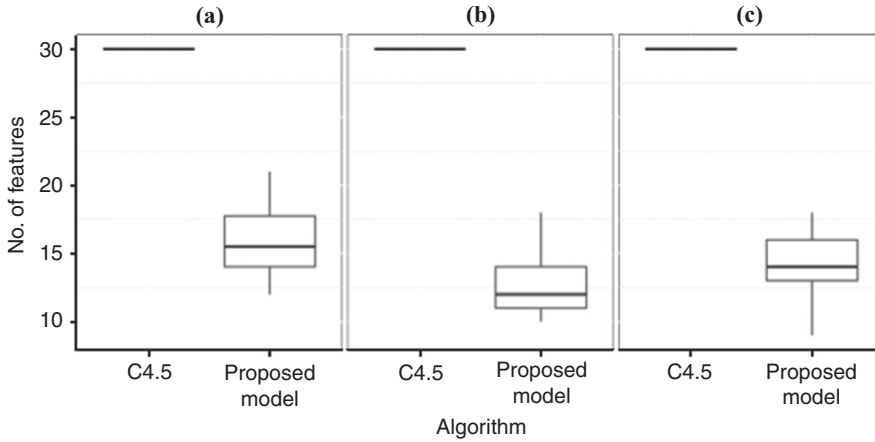
Results and discussion
Figures 1 and 2 depict the performance of the proposed model as compared to the standard C4.5 when simultaneously optimising the objective functions $f_1(x)$ and $f_2(x)$. Apart from a lower β achievement, our proposed model reported a higher α relative to the standard C4.5 classifier. For completeness, the mean and standard deviation values obtained for each experiment are tabulated in Table I. Mean values marked in italics indicate best statistical significance results at 95% confidence interval under the pairwise t-test (Hall and Holmes, 2003; Götz et al., 2008) comparison.
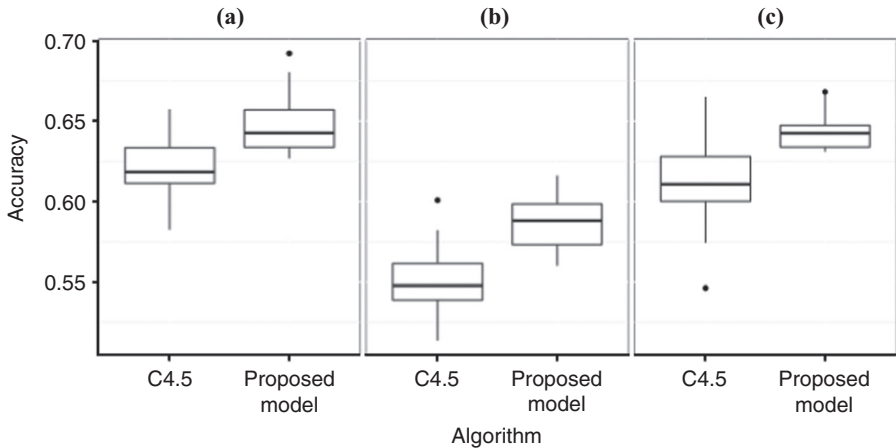
Figure 1.
A comparison of the accuracy rate between the standard C4.5 classifier and the proposed model in the data set

Notes: (a) G1; (b) G2; (c) G3

Notes: (a) G1; (b) G2; (c) G3

	Model		
Indicator	Standard C4.5 classifier	Proposed model	*p*-value
(a) Benchmark data with G1			
α	0.6193 ± 0.0184	*0.6472 ± 0.0170*	9.69E−008
β	30.00 ± 0.00	*15.73 ± 2.36*	< 2.2E−16
(b) Benchmark data with G2			
α	0.5499 ± 0.0189	*0.5863 ± 0.0153*	3.01E−011
β	30.00 ± 0.00	*12.97 ± 2.46*	< 2.2E−16
(c) Benchmark data with G3			
α	0.6131 ± 0.0259	*0.6433 ± 0.0108*	2.12E−007
β	30.00 ± 0.00	*14.03 ± 2.27*	< 2.2E−16

Notes: Mean ± standard deviation for 30 runs experimental results and computed *p*-values. Mean values marked in italics indicate best statistical significance results at 95% confidence interval under the pairwise *t*-test comparisons

The obvious yet encouraging results obtained inform us that our proposed model is superb in optimising the given data set using lesser number features but at the same time yielding much higher accuracy rate of classification. We attribute this to the superiority of MmGA in performing multi-objective optimisation. Consider a population of probable solutions (aka members) in our proposed model. Each member is represented as a variable x following Equation (A5) and is further associated with multi-objective-based fitness values, in this case α and β. The quality of the member is determined by its fitness values. Like all EAs, MmGA biases members with better fitness: At each iteration step, better fitness members receive higher chances of survival or become the parents of the next generation under an elitism strategy. Offspring, or new population members, are generated using mutation, crossover, and selection operators. The evolutionary process ends with both objectives converging in MmGA nominal and outlier evolution cycles; yielding p (Equation (A5)) in response to α and β.

A case study

Satisfied with the preliminary results, let us now consider applying the proposed model to a real-world case study encompassing a Malaysian private institution of higher education with more than a decade of history in open distance education. The institution offers tertiary education to working adults via open distance learning mode. The learning environment has been catered for adult learners seeking to purse tertiary qualifications for professional development and self-enrichment in a flexible manner. Furthermore, students and tutors come from different regions across Malaysia and worldwide.

Rather unique in its kind, the open distance learning institution provides five face-to-face tutorial classes that are spread over a period of five months to its students every semester. It also offers learning-support services via an open source learning management system (LMS). The LMS is an important platform for collaborative learning involving massive teaching-learning activities among course instructors, tutors, and students. For example, apart from the face-to-face classes, students and tutors continue to interact via video conferencing tools supported by the learning platform. Students are also free to engage in online activities such as downloading course materials, posting discussion in forums, participating in online quizzes, submitting assignments, and many more at any time anywhere convenient to them. On the other hand, instructors and tutors often play the role of system administrator in the online platform by uploading course materials, initiating discussion groups, setting up quizzes, marking assignments, answering posts, and others. It should be noted that, throughout the semester, students are generally assessed using three instruments on three periods: assignment 1 ($T1$), assignment 2 ($T2$), and final examination on the second, fourth, and fifth month, respectively.

Experimental setup

Moving on, the proposed MmGA-based soft computing model is depicted in Figure 3. Initially, data extracted from the LMS go through a pre-processing stage. It involves gathering various students' interaction data from courses and converting their frequency into required raw data in a tabular format. Noise from the raw data are removed and transformed into a structured data format, i.e., an Extensible Markup Language file format, so that the C4.5 classifier may perform further processing. The processing stage involves employing MmGA-based soft computing model. Lastly, the output of the processing stage is made available for interpretation. In most complex systems, the interpretation may involve end-users and incorporation of other tacit knowledge to uncover the existence of any
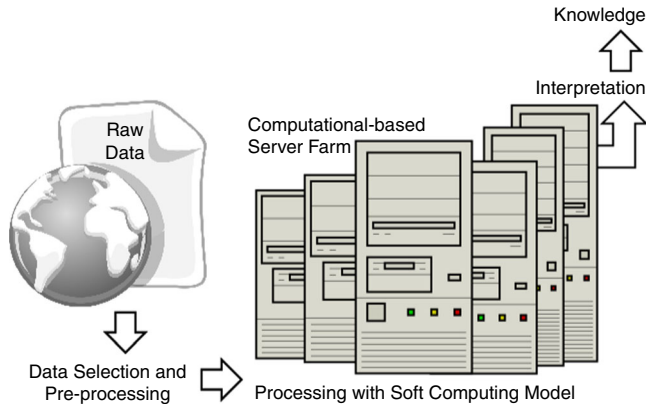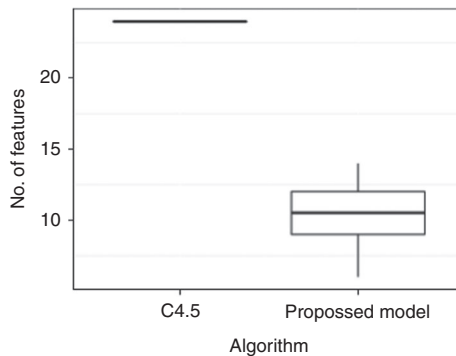
Figure 3.
Proposed model for the EDM in an open distance learning institution in Malaysia

possible relationship between the trends of students' online interaction activities with the e-learning platform and their examination performance, for instance.

Students' daily online interaction activities for every course are captured in LMS. In this study, a total of 46 courses offered in the said institution are examined. The data are collected throughout the entire semester for 24 consecutive weeks, including two weeks prior to the start of semester and two weeks after the end of semester. This contributed to 24 features, which are further grouped into two targeted classes: well-performed and not well-performed student classes corresponding to examination scores above or equal to 60 marks and below 60 marks, respectively. Note that the number of features is determined by a fixed interval of seven days. They form the input features for the classification and optimisation processes in the subsequent experimental studies carried out within the institution's computational-based server farm.

Results and discussion
As depicted by Figure 4, there were initially many chosen features (i.e. weeks) resulted from the application of C4.5 classifier. After applying MmGA, a significant reduction in the number of features is observed. Worth remarking that the effect is achieved without reduction in the accuracy rates as shown in Figure 5. On average of 30 runs, there is approximately 6 per cent improvement in the accuracy rate when employing our proposed model compared to the standard C4.5 classifier. In addition, Figures 4 and 5 depict that the classification accuracy rate of the proposed model has high level of agreement with each other. This performance indicator comes from the observation of a lower number of features,

Figure 4.
A comparison of the features number between the standard C4.5 classifier and the proposed model
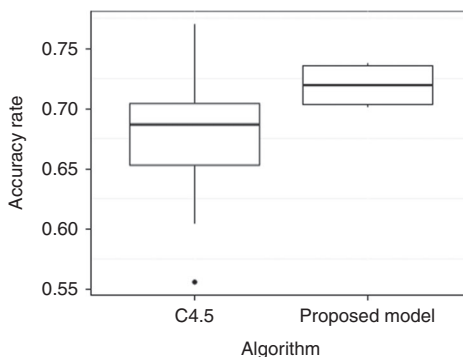
Figure 5.
A comparison of the accuracy rate between the standard C4.5 classifier and the proposed model

i.e., a reduction of up to 57 per cent relative to the standard C4.5 classifier. On closer inspection, the box plot distribution also reflects that the proposed model is more consistent and stable as compared to the standard C4.5 classifier since the former has narrower box and shorter tails. Reader is referred to Table II for the numerical results comprising mean, standard deviation, and p-value of the pairwise t-test comparisons. The best mean values marked in italics are statistically significant at 95% confidence interval.

To take a step further, let us examine the major determinants for the proposed model more closely. As illustrated by Figure 6, ten most prominent features (shaded in black) have been identified by our proposed model after optimisation. They represent the most significant weeks with student interaction activities that are adopted by the proposed model in classifying students achievement (recall Figure 4 and Table II). The captured interaction activities are not limited to students' post-reply inquiry on tutorials, technical hands-on, examination-related discussions, online quizzes, manipulation of teaching-learning materials, academia-related consultation, and clarification.

At second glance, the emergent patterns unfold several interesting events. First, students are actively involved in pre-semester activities before the start of a course (weeks 1 through 2). An obvious example of such activity includes exploration of course materials by students. The trend extends towards the second week just before the commencement of a new semester, which in turn corresponds to the closing date of course registration. This has come with little surprise as students are naturally more curious and eager to know about a new course being enroled in. But the obvious may have yet gotten the attention of educators. As evidential here, educators eager to improve students' first perception about a course, and subsequently leading to higher motivation in continuing the course (student retention strategy) should at least invest more time in the preparation work. Early content availability and accessibility,

Table II.
Numerical results comparing the accuracy rate and number of features between the standard C4.5 classifier and proposed model on case study

| Indicator | Model | | p-value |
	Standard C4.5 classifier	Proposed model	
α	0.6762 ± 0.0511	*0.7201 ± 0.0142*	7.19E−05
β	24.00 ± 0.00	*10.30 ± 2.23*	< 2.2E−16

Notes: Mean ± standard deviation for 30 runs experimental results and computed p-values. Mean values marked in italics indicate statistical significance results at 95% confidence interval under the pairwise t-test comparison

Figure 6.
Adopted rate for study weeks (features) in case study
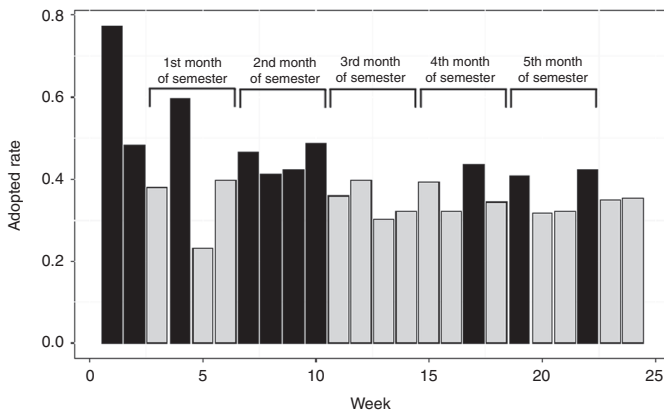
for example, would promote first positive impression and invite future interaction. A wide range of research works in cognition and social psychology attest to how initial impressions influence human interpretation of later information (Dougherty *et al.*, 1994).

Second, weeks 7 through 10 have been recognised as other significant indicators of student performance. Inherently, plenty of practical labs preparation and $T1$ discussions take place throughout the second month of the semester. To educators, this is likely the best time to engage more with students in ensuring that they are well with the course. The notion arises from the assumption that increasing the two-way interactions between tutor and student will enhance both student motivation and achievement. To this end, Allen *et al.* (2011) find strong correlation between high-quality tutor-student interactions and improved student achievement gains.

Whereas, the third week of the fourth month and the first week of the fifth month are the remaining two indicators identified for student performance. The former reflects the presence of discussion activities following $T2$, while the latter arises as a result of revision activities in conjunction with final examination. The preceding makes it clear that revision is important in boosting student performance. Educators may plan revision activities on gradual basis. Revision may be treated as a form of reinforcing previously learned knowledge through practises. Intuitively, educators should pay more attention to retrieval practice in consolidating learning (Karpicke and Roediger, 2008; Karpicke and Blunt, 2011) as part of the revision process. Finally, the common activities involved in the remaining two weeks (weeks 24 through 25) may include closing of posts and reporting of web statistics. They reflect the closing of the semester.

Conclusions

Many of the developing and potential EDM research on student achievement are contended to statistical methods (with a few exceptions) and lack the knowledge in handling multi-objective optimisation problems. To this end, this paper fills the gap by proposing a soft computing model with MmGA coupled with a DT-based classifier, namely, C4.5 classification and optimisation of system. Our model has shown confident results in student achievement classification under the UCI benchmark problem and real-world distance learning case study in Malaysia by simultaneous optimising multiple objectives (performance factors), i.e., maximising accuracy rate, α and minimising the number of features, β. Whether benchmark or real-world case study, the proposed model successfully reduced β by at least 48 per cent while achieving higher α.

We believe that this work can expand access to knowledge and insight into understanding student interaction activities and their achievement based on our empirical results. It may serve as a potential platform to inform educators seeking to reform educational policy by enhancing its provision of learning-support services and create a better learning experience for the students. To this end, the results presented could be easily translated into useful information such as when and what should be done in order achieve the target research goal. In the case study, early educator preparation work, improving tutor-student interactions, and investing in retrieval practices may well improve student motivation and achievement.

This is only the beginning of a study that can lead to more elaborative outcomes for the educational arena. The results thus far could very well be true for the case study e-learning environment, but the proposed model is transferable to any optimisation and classification problems. We also plan to deploy other types of MOEAs and classifiers for other experiments in the near future. Finally, another promising work is investigation into the behaviour of the proposed model in response to the nebulous data covering different domains of interest to educators.

We need more rigorous research in the educational arena and soft computing models have opened up a new route. We believe that the future is bright and the vacuum of empirical evidence shall continue to be filled by the enthusiast research works of EDM and the alike.

Note

1. Due to space limitation, interested reader is referred to Lim (2014) for more comprehensive review of structured population GAs.

References

Allen, J.P., Pianta, R.C., Gregory, A., Mikami, A.Y. and Lun, J. (2011), "An interaction-based approach to enhancing secondary school instruction and student achievement", *Science*, Vol. 333 No. 6045, pp. 1034-1037.

Atkinson, R.C. and Geiser, S. (2009), "Addressing the graduation gap", *Science*, Vol. 325 No. 5946, pp. 1343-1344.

Baker, R. (2014), "Educational data mining: an advance for intelligent systems in education", *IEEE Intelligent Systems*, Vol. 29 No. 3, pp. 78-82.

Baker, R.S. and Inventado, P.S. (2014), "Educational data mining and learning analytics", in Larusson, J.A. and White, B. (Eds), *Learning Analytics: From Research to Practice*, Springer, New York, NY, pp. 61-75.

Beume, N., Naujoks, B. and Emmerich, M. (2007), "SMS-EMOA: multiobjective selection based on dominated hypervolume", *European Journal of Operational Research*, Vol. 181 No. 3, pp. 1653-1669.

Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A. (1984), *Classification and Regression Trees*, Wadsworth, Belmont, CA.

Brester, C., Sidorov, M. and Semenkin, E. (2014), "Acoustic emotion recognition two ways of features selection based on self-adaptive multi-objective genetic algorithm", *11th International Conference on Informatics in Control, Automation and Robotics*, Vol. 2, IEEE, pp. 851-855.

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M.P. and Núñez, J.C. (2016), "Students' LMS interaction patterns and their relationship with achievement: a case study in higher education", *Computers & Education*, Vol. 96, pp. 42-54.

Chen, B., Resendes, M., Chai, C.S. and Hong, H.-Y. (2017), "Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse", *Interactive Learning Environments*, Vol. 25 No. 2, pp. 162-175.

Cheng, H.N., Liu, Z., Sun, J., Liu, S. and Yang, Z. (2017), "Unfolding online learning behavioral patterns and their temporal changes of college students in SPOCs", *Interactive Learning Environments*, Vol. 25 No. 2, pp. 176-188.

Chiang, T.H.-C. (2017), "Analysis of learning behavior in a flipped programing classroom adopting problem-solving strategies", *Interactive Learning Environments*, Vol. 25 No. 2, pp. 189-202.

Cho, J.H. and Kurup, P.U. (2011), "Decision tree approach for classification and dimensionality reduction of electronic nose data", *Sensors and Actuators B: Chemical*, Vol. 160 No. 1, pp. 542-548.

Coello, C.A.C. and Pulido, G. (2005), "Multiobjective structural optimization using a microgenetic algorithm", *Structural and Multidisciplinary Optimization*, Vol. 30 No. 5, pp. 388-403.

Cortez, P. and Silva, A.M.G. (2008), "Using data mining to predict secondary school student performance", *Proceedings of 5th Future Business Technology Conference*, (FUBUTEC 2008), EUROSIS, University of Porto, Porto, pp. 5-12.

Dougherty, T.W., Turban, D.B. and Callender, J.C. (1994), "Confirming first impressions in the employment interview: a field study of interviewer behavior", *Journal of Applied Psychology*, Vol. 79 No. 5, pp. 659-665.

Durán, O., Rodriguez, N. and Consalter, L.A. (2010), "Collaborative particle swarm optimization with a data mining technique for manufacturing cell design", *Expert Systems with Applications*, Vol. 37 No. 2, pp. 1563-1567.

Durillo, J.J. and Nebro, A.J. (2011), "jMetal: a java framework for multi-objective optimization", *Advances in Engineering Software*, Vol. 42 No. 10, pp. 760-771.

Ferguson, R. (2012), "Learning analytics: drivers, developments and challenges", *International Journal of Technology Enhanced Learning*, Vol. 4 Nos 5-6, pp. 304-317.

Fiol-Roig, G. and Miro-Julia, M. (2011), "Stock market analysis using data mining techniques: a practical application", *International Journal of Artificial Intelligence*, Vol. 6 No. S11, pp. 129-143.

Fogel, L.J. (1966), *Artificial Intelligence Through Simulated Evolution*, John Wiley & Sons, Chichester.

Golberg, D.E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison Wesley, Reading, MA.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talaón, M., Dopazo, J. and Conesa, A. (2008), "High-throughput functional annotation and data mining with the Blast2GO suite", *Nucleic Acids Research*, Vol. 36 No. 10, pp. 3420-3435.

Hall, M.A. and Holmes, G. (2003), "Benchmarking attribute selection techniques for discrete class data mining", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 15 No. 6, pp. 1437-1447.

Holland, J. (1992), *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press, Cambridge, MA.

Hu, Y., Wu, B. and Gu, X. (2017), "Learning analysis of k-12 students online problem solving: a three-stage assessment approach", *Interactive Learning Environments*, Vol. 25 No. 2, pp. 262-279.

Hunt, B. and Kübler, O. (1984), "Karhunen-Loeve multispectral image restoration, part I: theory", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 32 No. 3, pp. 592-600.

Hwang, G.-J., Chu, H.-C. and Yin, C. (2017), "Objectives, methodologies and research issues of learning analytics", *Interactive Learning Environments*, Vol. 25 No. 2, pp. 143-146.

Jiang, S., Ong, Y.-S., Zhang, J. and Feng, L. (2014), "Consistencies and contradictions of performance metrics in multiobjective optimization", *IEEE Transactions on Cybernetics*, Vol. 44 No. 12, pp. 2391-2404.

Karpicke, J.D. and Blunt, J.R. (2011), "Retrieval practice produces more learning than elaborative studying with concept mapping", *Science*, Vol. 331 No. 6018, pp. 772-775.

Karpicke, J.D. and Roediger, H.L. (2008), "The critical importance of retrieval for learning", *Science*, Vol. 319 No. 5865, pp. 966-968.

Kirkland, O., Rayward-Smith, V.J. and de la Iglesia, B. (2011), "A novel multi-objective genetic algorithm for clustering", in Yin, H., Wang, W. and Rayward-Smith, V. (Eds), *Intelligent Data Engineering and Automated Learning – IDEAL 2011: 12th International Conference, Norwich*, Springer, Berlin and Heidelberg, 7-9 September, pp. 317-326.

Kizilcec, R.F., Paérez-Sanagustaín, M. and Maldonado, J.J. (2017), "Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses", *Computers & Education*, Vol. 104, pp. 18-33.

Koza, J.R. (1992), *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, MA.

Kraft, M.A. (2015), "Teacher layoffs, teacher quality, and student achievement: evidence from a discretionary layoff policy", *Education Finance and Policy*, Vol. 11 No. 4, pp. 1-41.

Kremer, M., Brannen, C. and Glennerster, R. (2013), "The challenge of education and learning in the developing world", *Science*, Vol. 340 No. 6130, pp. 297-300.

Langstrand, J., Cronemyr, P. and Poksinska, B. (2015), "Practise what you preach: quality of education in education on quality", *Total Quality Management & Business Excellence*, Vol. 26 Nos 11-12, pp. 1202-1212.

Lemmerich, F., Ifland, M. and Puppe, F. (2011), "Identifying influence factors on students success by subgroup discovery", *Proceedings of the 3rd International Conference on Educational Data Mining*, pp. 345-346.

Leong, C.K., Lee, Y.H. and Mak, W.K. (2012), "Mining sentiments in SMS texts for teaching evaluation", *Expert Systems with Applications*, Vol. 39 No. 3, pp. 2584-2589.

Lim, T.Y. (2014), "Structured population genetic algorithms: a literature survey", *Artificial Intelligence Review*, Vol. 41 No. 3, pp. 385-399.

Lim, T.Y., Al-Betar, M.A. and Khader, A.T. (2015a), "Adaptive pair bonds in genetic algorithm: an application to real-parameter optimization", *Applied Mathematics and Computation*, Vol. 252, pp. 503-519.

Lim, T.Y., Al-Betar, M.A. and Khader, A.T. (2015b), "Monogamous pair bonding in genetic algorithm", *2015 IEEE Congress on Evolutionary Computation*, (CEC), Sendai, pp. 15-22.

Lim, T.Y., Al-Betar, M.A. and Khader, A.T. (2016), "Taming the 0/1 knapsack problem with monogamous pairs genetic algorithm", *Expert Systems with Applications*, Vol. 54, pp. 241-250.

Lopez, M.I., Luna, J., Romero, C. and Ventura, S. (2012), "Classification via clustering for predicting final marks based on student participation in forums", *Proceedings of the International Conference on Educational Data Mining*, International Educational Data Mining Society, Chania, pp. 148-151.

Márquez-Vera, C., Cano, A., Romero, C., Noaman, A.Y.M., Mousa Fardoun, H. and Ventura, S. (2016), "Early dropout prediction using data mining: a case study with high school students", *Expert Systems*, Vol. 33 No. 1, pp. 107-124.

Martín, D., Rosete, A., Alcalaa-Fdez, J. and Herrera, F. (2011), "A multi-objective evolutionary algorithm for mining quantitative association rules", *11th International Conference on Intelligent Systems Design and Applications, IEEE, Cordoba*, pp. 1397-1402.

Matthews, S.G., Gongora, M.A. and Hopgood, A.A. (2011), "Evolving temporal fuzzy association rules from quantitative data with a multi-objective evolutionary algorithm", in Corchado, E., Kurzynski, M. and Wozniak, M. (Eds), *Hybrid artificial intelligent systems: 6th International Conference, HAIS, Wroclaw, Poland, May 23-25, Proceedings, Part I*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 198-205.

Michael, J. and Gordon, S.L. (1997), *Data Mining Technique: For Marketing, Sales and Customer Support*, John Wiley & Sons Inc., New York, NY, p. 445.

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S. and Coello Coello, C. (2014a), "Survey of multiobjective evolutionary algorithms for data mining: part II", *IEEE Transactions on Evolutionary Computation*, Vol. 18 No. 1, pp. 20-35.

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S. and Coello Coello, C. (2014b), "A survey of multiobjective evolutionary algorithms for data mining: part I", *IEEE Transactions on Evolutionary Computation*, Vol. 18 No. 1, pp. 4-19.

Nandeshwar, A., Menzies, T. and Nelson, A. (2011), "Learning patterns of university student retention", *Expert Systems with Applications*, Vol. 38 No. 12, pp. 14984-14996.

Pangilinan, J.M. and Janssens, G.K. (2011), "Pareto-optimality of oblique decision trees from evolutionary algorithms", *Journal of Global Optimization*, Vol. 51 No. 2, pp. 301-311.

Peña-Ayala, A. (2014), "Educational data mining: a survey and a data mining-based analysis of recent works", *Expert Systems with Applications*, Vol. 41 No. 4, pp. 1432-1462.

Pourpanah, F., Tan, C.J., Lim, C.P. and Mohamad-Saleh, J. (2017), "A Q-learning-based multi-agent system for data classification", *Applied Soft Computing*, Vol. 52, pp. 519-531.

Pradhan, B. (2013), "A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS", *Computers & Geosciences*, Vol. 51, pp. 350-365.

Quinlan, J.R. (1986), "Induction of decision trees", *Machine learning*, Vol. 1 No. 1, pp. 81-106.

Quinlan, J.R. (1993), *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, CA.

Quinlan, J.R. (1996), "Improved use of continuous attributes in C4.5", *Journal of Artificial Intelligence Research*, Vol. 4, pp. 77-90.

Ripon, K.S.N. and Siddique, M.N.H. (2009), "Evolutionary multi-objective clustering for overlapping clusters detection", *9th IEEE Congress on Evolutionary Computation*, IEEE, pp. 976-982.

Romero, C., Laópez, M.-I., Luna, J.-M. and Ventura, S. (2013), "Predicting students' final performance from participation in on-line discussion forums", *Computers & Education*, Vol. 68, pp. 458-472.

Romero, C., Ventura, S., Pechenizkiy, M. and Baker, R.S. (2010), *Handbook of Educational Data Mining*, CRC Press, New York, NY.

Sachin, R.B. and Vijay, M.S. (2012), "A survey and future vision of data mining in educational field", *2nd International Conference on Advanced Computing & Communication Technologies, IEEE*, Rohtak, Haryana, pp. 96-100.

Schwefel, H.-P.P. (1993), *Evolution and Optimum Seeking: The Sixth Generation*, John Wiley & Sons Inc., New York, NY.

Smith, M.S. (2009), "Opening education", *Science*, Vol. 323 No. 5910, pp. 89-93.

Tan, C.J., Bong, C.W. and Natarajan, C. (2015), "Virtual classroom for technology-enhanced teaching and learning", *29th AAOU Annual Conference of Asian Association of Open Universities, Kuala Lumpur Conventional Centre (KLCC), Kuala Lumpur, 30 November-2 December*.

Tan, C.J., Hanoun, S. and Lim, C.P. (2015), "A multi-objective evolutionary algorithm-based decision support system: a case study on job-shop scheduling in manufacturing", *9th Annual IEEE International Systems Conference*, pp. 170-174.

Tan, C.J., Lim, C.P. and Cheah, Y.-N. (2013), "A modified micro genetic algorithm for undertaking multi-objective optimization problems", *Journal of Intelligent and Fuzzy Systems*, Vol. 24 No. 3, pp. 483-495.

Tan, C.J., Lim, C.P. and Cheah, Y.-N. (2014), "A multi-objective evolutionary algorithm-based ensemble optimizer for feature selection and classification with neural network models", *Neurocomputing*, Vol. 125, pp. 217-228.

Tan, C.J., Lim, C.P., Cheah, Y.-N. and Tan, S.C. (2013), "Classification and optimization of product review information using soft computing models", *International Symposium on Affective Engineering, Japan Society of Kansei Engineering*, pp. 115-120.

Tan, C.J., Neoh, S.C., Lim, C.P., Hanoun, S., Wong, W.P., Loo, C.K., Zhang, L. and Nahavandi, S. (2017), "Application of an evolutionary algorithm-based ensemble model to job-shop scheduling", *Journal of Intelligent Manufacturing*, pp. 1-12, available at: https://link.springer.com/article/10.1007/s10845-016-1291-1

University of California (2017), "University of California at Irvine (UCI) Machine Learning Database", School of Information and Computer Sciences, University of California, Irvine, available at: www.ics.uci.edu/mlearn/MLRepository.html (accessed 12 December 2016).

Van Leeuwen, A., Janssen, J., Erkens, G. and Brekelmans, M. (2014), "Supporting teachers in guiding collaborating students: effects of learning analytics in CSCL", *Computers & Education*, Vol. 79, pp. 28-39.

Venkatadri, M. and Rao, K.S. (2010), "A multiobjective genetic algorithm for feature selection in data mining", *International Journal of Computer Science and Information Technologies*, Vol. 1 No. 5, pp. 443-448.

Wang, F., Lee, N., Hu, J., Sun, J., Ebadollahi, S. and Laine, A.F. (2013), "A framework for mining signatures from event sequences and its applications in healthcare data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35 No. 2, pp. 272-285.

Willman, S., Lindaen, R., Kaila, E., Rajala, T., Laakso, M.-J. and Salakoski, T. (2015), "On study habits on an introductory course on programming", *Computer Science Education*, Vol. 25 No. 3, pp. 276-291.

Xing, W., Guo, R., Petakovic, E. and Goggins, S. (2015), "Participation-based student final performance prediction model through interpretable genetic programming: integrating learning analytics, educational data mining and theory", *Computers in Human Behavior*, Vol. 47, pp. 168-181.

Zhou, A., Jin, Y., Zhang, Q., Sendhoff, B. and Tsang, E. (2006), "Combining model-based and genetics-based offspring generation for multi-objective optimization using a convergence criterion", *IEEE International Conference on Evolutionary Computation, IEEE, Vancouver, BC*, pp. 892-899.

Zitzler, E. and Künzli, S. (2004), "Indicator-based selection in multiobjective search", in Yao, X., Burke, E.K., Lozano, J.A., Smith, J., Merelo-Guervós, J.J., Bullinaria, J.A., Rowe, J.E., Tiňo, P., Kabán, A. and Schwefel, H.-P. (Eds), *Parallel Problem Solving from Nature – PPSN VIII: 8th International Conference, Birmingham*, Springer, Berlin and Heidelberg, 18-22 September, pp. 832-842.

Zitzler, E. and Thiele, L. (1999), "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach", *IEEE Transactions on Evolutionary Computation*, Vol. 3 No. 4, pp. 257-271.

Appendix

C4.5 classifier

Let C denotes the number of classes and $p(X, j)$ as the proportion of cases in X that belongs to the jth class. The uncertainty about the class for a case of X, and its corresponding information gained by a test T with k outputs are derived from the following equation (Quinlan, 1993, 1996):

$$g(X, T) = f(X) - \sum_{i=1}^{k} \frac{|X_i|}{|X|} \times f(X_i) \tag{A1}$$

subject to:

$$f(X) = - \sum_{j=1}^{C} p(X, j) \times \log 2(p(X, j))$$

$$s(X, T) = - \sum_{i=1}^{k} \frac{|X_i|}{|X|} \times \log 2 \left(\frac{|X_i|}{|X|} \right) \tag{A2}$$

Note that function $g(X, T)$ is influenced by k and is maximal for only one case in each subset X_i. Furthermore, by knowing which case the subset X_i falls, it is possible to obtained the potential information of partitioning a set of cases simply by using Equation (A2). The desirability of test is computed using the ratio of Equations (A1) and (A2).

Objective functions and MmGA

The two objective functions, i.e., α and β, of MmGA are derived as follows:

$$\text{maximize } f_1(x) = \alpha \tag{A3}$$

$$\text{minimize } f_2(x) = \beta \tag{A4}$$

subject to:

$$\mathbb{F} \equiv \left\{ x \in \mathbb{R}^n \| 0 \leqslant \alpha \leqslant 1, 1 \leqslant \beta \leqslant L \right\}, \mathbb{F} \subseteq \mathbb{S}$$

L is the total of β, and $x = (x_1, x_2, ..., x_L)^T$ is a vector of decision variables. $\mathbb{F}$ and $\mathbb{S}$ represent the feasible region and whole search space, respectively. Equation (A5) is derived to yield optimum values of $f_i(x)$ based on the given x. It is also known as the solution of the Pareto optimality, which denotes a Pareto optimal set with respect to the two functions f_1 and f_2:

$$p \equiv \left\{ [f_1(x), f_2(x)]^T, \forall x \in \mathbb{F} \right\} \tag{A5}$$

Figure A1 further depicts the pseudo-code of MmGA, which yields the Pareto (pf_{MmGA}) in correspondence to true Pareto (pf_{true}). pf has a vector of x^* that is Pareto optimal: either $\forall_{i \in k}(f_i(x) = f_i(x^*))$ or there is at least one $i \in k$ such that $f_i(x) > f_i(x^*)$ where k represents the number of objective functions and every $x \in \mathbb{F}$. As such, k is 2 in this study for optimising objectives α and β.

```
Require: Generate non-dominated solutions for MOP
 1: Procedure MmGA
 2: m = Sort(m_init, ⪯)
 3: while i < Maximum outlier evaluation and pf_MmGA ≠ pf_true do is met
 4:    p_init = InitialisePopulation(n, ratio, irm, rm, m)
 5:    p = Sort(p_init, ⪯)
 6:    repeat
 7:       u = Binary tournament θ_s on p
 8:       v = Two-Point θ_r on u
 9:       w = Uniform θ_m on v
10:       pf_MmGA_i = MmGAElitism(w, p, 1)
11:    until Maximum nominal convergence is reached or pf_MmGA = pf_true is met
12:    em_MmGA = MmGAElitism(pf_MmGA_i, em, eliteSize)
13:    if em is full when trying to insert em_elite then
14:       em = adaptiveGrid(em_elite)
15:    end if
16:    m = MmGAElitism(pf_MmGA_i, m, eliteSize)
17:    if i modulus replacementCycle then
18:       rm = MmGAElitism(em, rm, eliteSize)
19:    end if
20: end whilereturn pf_MmGA
```

Corresponding author
Choo Jun Tan can be contacted at: cjtan@wou.edu.my