Neural Comput & Applic (2019) 31 (Suppl 2):S845–S855
https://doi.org/10.1007/s00521-017-3042-4

ORIGINAL ARTICLE

Knowledge building through optimized classification rule set generation using genetic based elitist multi objective approach

Preeti Gupta[1] · Tarun Kumar Sharma[1] · Deepti Mehrotra[2] · Ajith Abraham[3]

Received: 21 June 2016 / Accepted: 26 April 2017 / Published online: 27 May 2017

Abstract Knowledge building is an important activity taken up by various organizations. The paper exemplifies the creation of a knowledge-centric environment for a non-profit sector such as a higher education. Building knowledge and thereafter using it are important aspects of knowledge-centric environment; this further helps the organization to gain competitive advantage. With the increase in popularity of genetic algorithm (GA), the technique has been used in building efficient classifiers for creating effective rule sets. The paper makes use of multi-objective genetic algorithm for building GA-based efficient classifier because classification rule mining is itself, a multi-objective problem. Knowledge expressed through classification rules help in establishing relationships between attributes that are not visible openly. The study assumes importance as curriculum planning is an important aspect of any academic institution, the knowledge derived in the form of rules residing in the knowledge base help to substantiate proper curriculum development, making a sizeable contribution toward professional growth and advancement of the students. On implementation of the findings, educational organizations will be able to institute themselves as knowledge centric.

Keywords Knowledge · Genetic algorithm · Classification · Multi-objective optimization · Higher education

1 Introduction

Implementing business intelligence for decision making brings about rapid changes in the business landscape of various organizations [1–3]. The ability of an organization to adjust to the dynamics of the business and the working environment prevailing in the organization governs its survival and success. Using suitable instruments for extraction of knowledge from the accumulated data helps the organizations to thrive and take advantage of knowledge economy. Extraction of information from data facilitates knowledge building. Information which can be termed as a subset of data stimulates action in an entity, whereas knowledge defines the action of an entity in a particular setting. Though there are numerous ways to represent knowledge [4], using production rules, written in the form of If–Then rules, is one of the most popular approaches used for knowledge representation [5]. The If–Then rules adopt a modular approach, each defining principally independent and a relatively minor piece of knowledge. A rule-based system will include universal rules and actualities about the knowledge domain covered. An optimized rule-based system will always help to make effective deductions or choices. Real-world problems primarily aim to acquire optimization on certain output criteria. Many at times, modeling of real-world problems is much more complex as many objectives have to be

✉ Preeti Gupta
pgupta@jpr.amity.edu

Tarun Kumar Sharma
tksharma@jpr.amity.edu

Deepti Mehrotra
dmehrotra@amity.edu

Ajith Abraham
ajith.abraham@ieee.org

1 Amity University Rajasthan AUR, Jaipur, India

2 Amity University Uttar Pradesh AUUP, Noida, India

3 Machine Intelligence Research Labs, Auburn, Washington, USA

minimized or maximized simultaneously tending toward multi-objectivity.

GA-based on Darwin's theory of evolution and natural selection can be used to exploit large search spaces. Through the work of eminent researchers, the usefulness of GA was proved for function optimization [6, 7]. Hence, real-world problems inherently complex and multi-objective in nature, significantly relying on knowledge base, can be modeled using GA. GA does not only deal with single objective optimization problems such as minimizing or maximizing any variable or a function [8] but later through research work taken up, GA was also used to solve multi-objective problems [9].

This paper undertakes creation of a knowledge-centric environment that depends on creation of knowledge set through optimization of classification rule set using genetic-based elitist multi-objective approach. Generation of such a classification rule set is a multi-objective optimization problem. Pareto approach has been used to cope up with multi-objectivity of the problem, giving rise to many non-dominated solutions in the form of rules. In due course, relevance index is derived for rules in the knowledge base, boosting value of the knowledge base. The knowledge derived in form of rules is interpreted and could be used to supplement processes like syllabus planning, lesson planning, structuring criteria for evaluation of the student's performance and implementation of appropriate instructive practices for improving academic performance.

The paper is organized as follows. The related work in the domain is discussed in Sect. 2. Section 3 concentrates on the proposed application of multi-objective genetic algorithm (MOGA) for classification rule mining in the working scenario of higher education. Section 4 concludes the findings, summarizing the significance of the technique for setting up of knowledge-centric higher education organizations.

2 Related work

It is often said that we are drowning in data but starving for knowledge [10]. Extraction of information from data facilitates knowledge building. Number of researchers have classified knowledge on different basis, sometimes defining the manner of codification and occurrence [11] or on the basis of know–what, know–how, know–why and know–when aspect of knowledge [12, 13]. Some have even mapped knowledge in diverse domains [14].

Very similar to other business domains, knowledge utilization facilitates formulation of a successful and result-oriented system in education sector [15–19]. Numerous crucial processes existing in any educational organization pertaining to strategic planning, admission, curriculum development, teaching–learning, examination and evaluation, research initiatives, bonding with alumni and many more thrive on implementing knowledge-centric activities [14].

Classification rule mining problems are considered as multi-objective problem rather than single objective [20].

Earlier work on classification rule mining using MOGA proposed three different approaches to cope with multi-objective optimization problems [21].

1. Weighted max min approach
2. The lexicographical approach
3. The Pareto approach

In [22], work carried out explained hybridized GA with Tabu search to do classification rule mining. Further through research work, a technique involving elitist multi-objective genetic algorithm for classification rule generation was proposed. Parameters such as prediction accuracy, comprehensibility and measure of interestingness of the rules were taken into consideration. In the work, simultaneous optimization of these objectives was achieved through a hybrid crossover operator using multi-objective genetic algorithm. On comparing with simple GA, the technique was found to have an edge over it [23]. Researchers also applied MOGA for classification in which each chromosome represented a classification rule. Different measures have been used to evaluate the classification rules. In the work [24], researcher used lexicographic Pareto based multi-objective genetic algorithm to extract useful rules from Iris Data set. The rules were evaluated on three parameters namely confidence, coverage and attractiveness. A novel technique of designing autonomous classifiers via MOGA based on three objectives was discussed by researchers [25]. Measure to quantify the understandability of the classifiers, classification accuracy and average support value was taken into consideration. Apart from designing efficient classifiers, researchers have also catered to other issues like handling missing attributes. In [26], a technique handles real-coded MOGA for mining classification rules with missing attribute values.

3 Proposed work: multi-objective classification problem in higher education

Education field has progressed over the years. It has definitely steered the growth of many academic institutions having private ownership [27] or government undertaking. For creating their stand in the market, these universities and colleges compete among themselves. The management aims to take measures in bringing excellence to the professional education by provision of congenial environment for teaching and learning that eventually broadens the mental horizon of the student. Undoubtedly, designing the academic curriculum is an important internal activity in any academic institution. For measuring the effectiveness of the designed curriculum, the students' academic performance during end term examination can be treated as an indicator. Different factors affect the grades of a student in the end term examination for the course. There is a need to trace the relationship between these factors. The relationship existing between these factors can be represented in the form of rules. The collection of such rules can form a knowledge base that can eventually supplement the formulation of strategies in designing effective curriculum. In the study taken up, the engineering subject of analog electronics is considered. The dependent attributes identified for the study is the grade a student acquires in analog electronics for the end term examination. Various independent attributes exercising a positive or negative effect on the identified dependent attributes are given below in Table 1 [28].

Multi-objective genetic algorithms As already stated in a single objective problem one objective is either maximized or minimized. There will be only one solution of these types of problem where minimum or maximum value of objective is achieved. In multi-objective problem, objective is not one. There are many objectives that are to be optimized [29].

3.1 Multi-objective GA-based classification

Through implementation of GA, the large search space in classification rule problem can be explored. In order to deal with multi-objective problem like classification rule mining, three different approaches, namely: (1) the weighted sum methodology, assigning weights to the attributes, (2) the lexicographical methodology, ranking the objectives according to significance, and (3) the Pareto methodology relying on numerous non-dominated solutions, are followed. On carrying out comparison, it can be found that the weighted sum approach is a cobbled approach, created or used for a particular purpose as necessary, the lexicographic and the Pareto approach on the other hand are more evolved approaches. Pareto approach relies on direct tackling of multi-objective problem in place of reducing a multi-objectivity of a problem to single objectivity and further solving it by any evolutionary algorithm. In the approach of Pareto dominance, solutions are evaluated against each other. A solution dominates another only if it performs better in at least one criterion and has non-inferior performance in all the other criteria. A solution achieves Pareto optimality if it remains non-dominated by any other solution existing in the search space. As exhaustive search is infeasible in complex search spaces, achieving Pareto

Table 1 Attributes (independent and dependent) considered in the study

Attribute Independent attributes	Explanation	Values Grade* (Marks obtained)
Cem	Performance of the students, continuously evaluated by the faculty member with respect to class assignments, marks obtained in class test, performance in viva-voce, etc. (maximum marks −30)	Grade F (0–5) Grade BA (5.1–10) Grade Avg (10.1–15) Grade *G* (15.1–20) Grade VG (20.1–25) Grade OS (25.1–30)
		(Grade point obtained)
Prev_sgpa	Previous semester performance of the student, evaluated in terms of semester grade point average (SGPA) (maximum pointer 10)	Grade *A* (9–10) Grade *B* (8–8.9) Grade *C* (7–7.9) Grade *D* (6–6.9) Grade *E* (5–5.9) Grade *F* (4–4.9) Grade *G* (3–3.9) Grade *H* (2–2.9) Grade *I* (1–1.9) Grade *J* (0–0.9)
		Satisfaction level value (satisfaction level)
Satisfaction_student_wrt_faculty	Level of student's satisfaction for the course faculty member	Good (*A*) Fair (*B*)
		Attendance categorization (attendance)
Course_presence	Presence of the student in the class of the subject under consideration (in percentage)	Poor (75%–Below) Satisfactory (75.1–80%) Fair (80.1–85%) Good (85.1–90%) VG (90.1–95%) Excellent (95.1–100%)
		Grade* (marks obtained)
Practical_orientation_in_the_subject	Measure of the practical intellect of the student in the subject under consideration	Grade *F* (0–21) Grade BA (21.1–31) Grade Avg (31.1–41) Grade *G* (41.1–51) Grade VG (51.1–61) Grade OS (61.1–70)

Table 1 continued

		Grade*(marks obtained)
Performance_ base_sub	Performance exhibited by the student in the end term examination of Physics (base subject), studied during the earlier semester (maximum marks –70)	Grade *F* (0–21) Grade BA (21.1–31) Grade Avg (31.1–41) Grade *G* (41.1–51) Grade VG (51.1–61) Grade OS (61.1–70)
Dependent attributes		Grade*(marks obtained)
Performance_ EndSemester	Acquired marks in the end semester examination for the subject under consideration (maximum marks –70)	Grade *F* (0–21) Grade BA (21.1–31) Grade Avg (31.1–41) Grade *G* (41.1–51) Grade VG (51.1–61) Grade OS (61.1–70)

* Grade—*F* Fail, *BA* Below Average, *Avg* Average, *G* Good, *VG* Very Good, *OS* Outstanding

optimality is a real challenge. In that case instead of constructing Pareto set exhibiting optimal solutions, a set of non-dominated solutions with values as close as possible to the objective values is achieved. Such non-dominated solutions fall on the Pareto front [30].

The paper follows real-coded multi-objective GA-based classification system and optimizes the rule measures of coverage and comprehensibility simultaneously by Pareto approach.

3.2 Rule generation: mathematical formulation

The proposed method provides a general framework for elitist multi-objective rule generation and creation of the knowledge set [29].

1. *Gen = 1; KnowledgeSet(Gen) =* ϕ*;*
2. *Population initialization P(Gen);*
3. *Evaluation of objective function, EvalObj(P(Gen))*
4. *Pareto dominance-based rank assignment of fitness value, AssignFV(P(Gen))*
5. *Insert (KnowledgeSet(Gen),Chromosomes ranked as 1)*
6. *While (Gen* $\leq$ *Specified Generations) do*
   (a) *Applying Selection, Select(p(Gen))* $\rightarrow$ *P1(Gen)*
   (b) *Applying CrossOver and Mutation CroMut(P1(Gen))* $\rightarrow$ *P2(Gen)*
   (c) *Replace(P(Gen), P2(Gen))* $\rightarrow$ *P(Gen + 1)*
   (d) *Evaluation of Objective Function, EvalObj(P(Gen + 1))*
   (e) *Pareto Dominance-based rank assignment of Fitness Value, AssignFV(P(Gen + 1))*
   (f) *Insert (KnowledgeSet(Gen), Chromosomes ranked as 1)* $\rightarrow$ *KnowledgeSet(Gen + 1)*
   (g) *Gen = Gen + 1;*
7. *End While*
8. *Decode the chromosome as If–Then rule, Decode (KnowledgeSet(Gen))*

The steps in the pseudocode are followed and implemented.

3.2.1 Building population P(Gen)

Initial population (IP) is built from the training data set. IP size is fixed to 10% of training data set size. As the searching capability of GA is guided by population size, its size should increase with the size of the data set making searching of complex search space possible in a large data set.

For the data set used in the experiment, any two members of initial population are listed below:

$$\mathrm{MemberOfPopulation}_1 = \mathrm{Avg}|B|E|\mathrm{Good}|\mathrm{Good}|\mathrm{Avg}$$
$$\mathrm{MemberOfPopulation}_2 = \mathrm{Avg}|B|C|\mathrm{Fair}|\mathrm{Fair}|\mathrm{Avg}$$

where the entities denote continuous evaluation marks | level of student's satisfaction for the course faculty member | previous semester's SGPA of the student | presence of the student in the course under scrutiny | base subject (Physics) marks | practical orientation of students toward the subject under scrutiny.

3.2.2 Encoding and interpretation of chromosomes

Encoding of chromosome The chromosome can be encoded as given below:

Taking into consideration all members of initial population having valid values, for any run of GA, number of genes in chromosome (NGC) and valid gene number (VGN) have to be fixed.

For the study undertaken, number of genes in chromosome (NGC) = 4 and valid gene number (VGN) = $a_{n1}, a_{n2},\ldots, a_{n\mathrm{NGC}} = 1, 2, 4, 5$. Then, chromosomes from above-mentioned population set, in accordance with the member of the initial population, are:

$$\mathrm{Ch}_1 = \mathrm{Avg}|B|\mathrm{Good}|\mathrm{Good}$$
$$\mathrm{Ch}_2 = \mathrm{Avg}|B|\mathrm{Fair}|F$$

Interpretation of chromosome The chromosome can be interpreted as given below:

Chromosome 1 $(\mathrm{Ch}_1) = \mathrm{Avg}|B|\mathrm{Good}|\mathrm{Good}$

Continuous evaluation marks = Avg (Average, 10.1–15)

Level of student's satisfaction for the course faculty member = B (Fair).

Presence of the student in the subject under consideration = Good (85.1–90%).

Base subject marks (Physics) = Good (41.1–51).

Chromosome 2 (Ch_2) = Avg | *B* | Fair | *F*.

Continuous evaluation marks = Avg (Average, 10.1–15).

Level of student's satisfaction for the course faculty member = *B* (Fair).

Presence of the student in the subject under consideration = Fair (80.1–85%).

Base subject marks (Physics) = *F* (Fail, 0–21).

3.2.3 Evaluation of objective function

The proposed rule generation problem is stated as follows, listed in Eq. (1):

$$\text{Maximize}(f1(R), f2(R)) \tag{1}$$

Such that the search space is defined by $0 < (f1(R), f2(R)) \leq 1$.

Where

$$f1(R) = \text{Comprehensibility of a rule}\,(R). \tag{2}$$

If C_c defines the maximum number of conditions in a rule *R*, the comprehensibility of the rule *R* can be defined as in Eq. (3):

$$\text{Comp}\,(X \rightarrow Y) = 1 - \frac{N_c(R)}{C_c} \tag{3}$$

where $N_c(R)$ is the number of conditions in the rule *R*.

$$f2(R) = \text{Coverage of a rule}\,(R) \tag{4}$$

Coverage is sometimes called as antecedent support or LHS support, given in Eq. (5). The value of the metrics lies between 0 and 1.

$$\text{Cover}\,(X \rightarrow Y) = \text{Supp}\,(X) \tag{5}$$

Thus, the encoding the chromosomes by following NGC and VGN can be seen in Table 2. Moreover, the antecedent part of the rule is measured by comprehensibility and coverage.

3.2.4 Assigning fitness to P(Gen) by assigning rank based on Pareto dominance

All objective functions values (fitness values) are to be calculated for all the chromosomes. Ranks are assigned to each individual of the population (chromosome) based on non-dominated sorting that is done front wise. The implemented method is shown in Table 3 below. Chromosomes ranked as 1 are chosen for *KnowledgeSet*(*Gen*).

Therefore, the chromosomes that are added to the knowledge set are the chromosomes having rank 1, i.e., BA || *B* || Poor || BA.

Till the number of generations are not equal to the specified number of generations, the below procedure is to be followed.

3.2.5 Selection and reproduction

During the stage of selection in GA, individuals from a population are chosen for later breeding. The study randomly chooses few individuals from the population and runs a 'tournament' among them. The survivor is finally selected (the one with the best fitness) to carry out crossover. In tournament selection, individuals are chosen randomly from the population. Crowding operator then decides the designation of individuals as best, for their selection as parents for cross over. Crowding distance mentioned in Eqs. (6) and (7) below is an estimate of density of solutions, surrounding a particular solution in population. It helps in obtaining uniform distribution. The calculation of crowding distance is done according to the Fig. 1.

$$1_{CD} = l_{CD} = \infty \tag{6}$$

$$i_{CD} = \sum_m \left(\frac{f[i+1]_m - f[i-1]_m}{f_m^{max} - f_m^{min}} \right) \quad \text{where } i = 2, 3, \ldots (l-1) \tag{7}$$

$f[i]_m$ represents *m*th objective value of *i*th solution and f_m^{max} is the maximum value of the function f_m in the Pareto front. The formation of *P*1(Gen) is shown in Table 4 below.

Table 2 Evaluation of objective functions

Student	Chromosome	Comprehensibility	Coverage
Chromosome 1	Good | *B* | Excellent | *F*	0.333333	0.012195122
Chromosome 2	Good | *A* | Excellent | *F*	0.333333	0.004878049
Chromosome 3	VG | *B* | *B* | Sat | *F*	0.166667	0.002439024
Chromosome 4	BA | *B* | Fair | *F*	0.333333	0.007317073
Chromosome 5	VG | *B* | Fair | Good	0.333333	0.002439024
Chromosome 6	BA | *A* | *D* | VG | Good	0.166667	0.002439024
Chromosome 7	VG | *A* | Excellent | Avg	0.333333	0.009756098
Chromosome 8	Good | *B* | *D* | Fair | *F*	0.166667	0.002439024
Chromosome 9	Avg | *B* | Fair | *F*	0.333333	0.017073171
Chromosome 10	BA | *B* | Poor | BA	0.333333	0.031707317

Table 3 Assignment of fitness value

Student	Chromosome	Comprehensibility	Coverage	Fitness value (rank based)
Chromosome 1	Good ǀ *B* ǀ Excellent ǀ *F*	0.333333	0.012195122	2
Chromosome 2	Good ǀ *A* ǀ Excellent ǀ *F*	0.333333	0.004878049	5
Chromosome 3	VG ǀ *B* ǀ *B* ǀ Sat ǀ *F*	0.166667	0.002439024	7
Chromosome 4	BA ǀ *B* ǀ Fair ǀ *F*	0.333333	0.007317073	4
Chromosome 5	VG ǀ *B* ǀ Fair ǀ Good	0.333333	0.002439024	6
Chromosome 6	BA ǀ *A* ǀ *D* ǀ VG ǀ Good	0.166667	0.002439024	7
Chromosome 7	VG ǀ *A* ǀ Excellent ǀ Avg	0.333333	0.009756098	3
Chromosome 8	Good ǀ *B* ǀ *D* ǀ Fair ǀ F	0.166667	0.002439024	7
Chromosome 9	Avg ǀ *B* ǀ *A* ǀ Fair ǀ *F*	0.166667	0.007317073	5
Chromosome 10	BA ǀ *B* ǀ Poor ǀ BA	0.333333	0.031707317	1

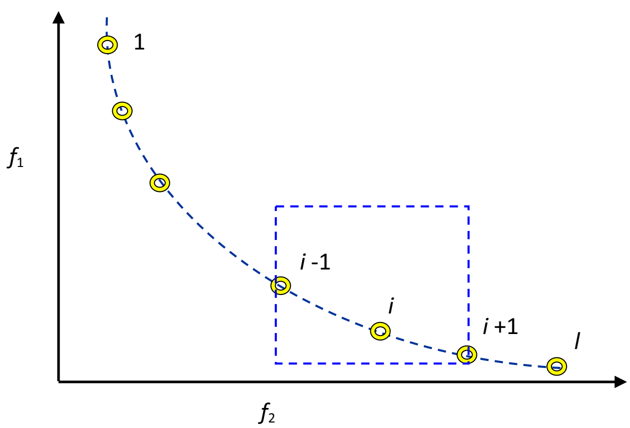

Fig. 1 Crowding distance calculation

Table 4 Selection procedure

Tournament selection		*P*1(Gen)	
First rule	Second rule	Selected rule	Selected chromosome
Rule 4	Rule 7	Rule 4	BA ǀ *B* ǀ Fair ǀ *F*
Rule 1	Rule 4	Rule 1	BA ǀ *B* ǀ Poor ǀ BA
Rule 9	Rule 10	Rule 10	Good ǀ *B* ǀ *D* ǀ Fair ǀ *F*
Rule 8	Rule 10	Rule 10	Good ǀ *B* ǀ *D* ǀ Fair ǀ *F*
Rule 3	Rule 6	Rule 3	VG ǀ *A* ǀ Excellent ǀ Avg
Rule 4	Rule 8	Rule 4	BA ǀ *B* ǀ Fair ǀ *F*
Rule 2	Rule 3	Rule 2	Good ǀ *B* ǀ Excellent ǀ *F*
Rule 5	Rule 7	Rule 5	Good ǀ *A* ǀ Excellent ǀ *F*
Rule 1	Rule 4	Rule 1	BA ǀ *B* ǀ Poor ǀ BA
Rule 3	Rule 8	Rule 7	VG ǀ *A* ǀ Excellent ǀ Avg

3.2.6 Crossover

The phase of crossover follows the selection step that enriches the population with fit individuals. The notion behind implementing crossover is to improve the child population by taking the best characteristics from each of the parents. This will enhance the child chromosome, as compared to the parents. The aim of crossover is to produce chromosomes better than both parents. Parent chromosomes are again quality sequences selected during reproduction. Single point crossover is implemented in the study. It includes random selection of crossover site over the stretch of the mated sequence and swapping genes subsequent to the crossover position.

3.2.7 Mutation

The process of mutation follows crossover. Study implements, mutation involving flipping of a gene, with a small mutation probability. The mutation operator preserves the diversity of population which is primarily important. In the study, the probability is taken as 0.20–0.25. On applying crossover and mutation, Table 5 is obtained.

3.2.8 Population for the next generation

The population for the next generation is created by the replacement of the origination chromosomes present in the population by the newly obtained one after mutation. The newly created population is evaluated by the objective function by assigning fitness, subjected to ranking based on Pareto dominance. Chromosomes sequences that acquired rank 1 are added to the original *KnowledgeSet*(*Gen*) and the number of generation is incremented by 1.

3.2.9 Convergence

Once a set number of generations are over, the convergence is achieved. The algorithmic procedure should then be halted and the *KnowledgeSet*(*Gen*) extracted. The chromosomes stored in *KnowledgeSet*(*Gen*) are then decoded to obtain the If–Then rules, i.e., class labels are assigned to the chromosome.

Table 5 Crossover and mutation

$P1$(Gen)	$P2$(Gen)	
Selected chromosome	Crossover	Mutation
BA ∣ B ∣ Fair ∣ F	BA ∣ B ∣ Poor ∣ BA	BA ∣ B ∣ Sat ∣ BA
BA ∣ B ∣ Poor ∣ BA	BA ∣ B ∣ Fair ∣ F	Avg ∣ B ∣ Fair ∣ F
Good ∣ B ∣ D ∣ Fair ∣ F	Good ∣ B ∣ D ∣ Fair ∣ F	Good ∣ B ∣ D ∣ Poor ∣ F
Good ∣ B ∣ D ∣ Fair ∣ F	Good ∣ B ∣ D ∣ Fair ∣ F	Good ∣ B ∣ E ∣ Fair ∣ F
VG ∣ A ∣ Excellent ∣ Avg	VG ∣ A ∣ Fair ∣ F	VG ∣ A ∣ Good ∣ F
BA ∣ B ∣ Fail ∣ F	BA ∣ B ∣ Excellent ∣ Avg	BA ∣ B ∣ Excellent ∣ BA
Good ∣ B ∣ Excellent ∣ F	Good ∣ B ∣ Excellent ∣ F	Good ∣ A ∣ Excellent ∣ F
Good ∣ A ∣ Excellent ∣ F	Good ∣ A ∣ Excellent ∣ F	Good ∣ A ∣ Excellent ∣ BA
BA ∣ B ∣ Poor ∣ BA	BA ∣ B ∣ Excellent ∣ Avg	BA ∣ B ∣ Excellent ∣ BA
VG ∣ A ∣ Excellent ∣ Avg	VG ∣ A ∣ Poor ∣ BA	VG ∣ B ∣ Poor ∣ BA

Table 6 Knowledge set preparation

Generation number	Knowledge set
0	BA ∣ B ∣ Poor ∣ BA
1	BA ∣ B ∣ Poor ∣ BA, Avg ∣ B ∣ Fair ∣ F
2	BA ∣ B ∣ Poor ∣ BA, Avg ∣ B ∣ Fair ∣ F, VG ∣ A ∣ Excellent ∣ BA
3	BA ∣ B ∣ Poor ∣ BA, Avg ∣ B ∣ Fair ∣ F, VG ∣ A ∣ Excellent ∣ BA, BA ∣ B ∣ Excellent ∣ BA, Good ∣ A ∣ VG ∣ F
4	BA ∣ B ∣ Poor ∣ BA, Avg ∣ B ∣ Fair ∣ F, VG ∣ A ∣ Excellent ∣ BA, BA ∣ B ∣ Excellent ∣ BA, Good ∣ A ∣ VG ∣ F, VG ∣ A ∣ Excellent ∣ Good, Good ∣ A ∣ Poor ∣ Avg, VG ∣ A ∣ Good ∣ BA
5	BA ∣ B ∣ Poor ∣ BA, Avg ∣ B ∣ Fair ∣ F, VG ∣ A ∣ Excellent ∣ BA, BA ∣ B ∣ Excellent ∣ BA, Good ∣ A ∣ VG ∣ F, VG ∣ A ∣ Excellent ∣ Good, Good ∣ A ∣ Poor ∣ Avg, VG ∣ A ∣ Good ∣ BA, Avg ∣ A ∣ VG ∣ F
⋮	⋮
100	BA ∣ B ∣ Poor ∣ BA, Avg ∣ B ∣ Fair ∣ F, VG ∣ A ∣ Excellent ∣ BA, BA ∣ B ∣ Excellent ∣ BA, Good ∣ A ∣ VG ∣ F, VG ∣ A ∣ Excellent ∣ Good, Good ∣ A ∣ Poor ∣ Avg, VG ∣ A ∣ Good ∣ BA, Avg ∣ A ∣ VG ∣ F,…

3.2.10 Assignment of class labels

The chromosomes stored in *KnowledgeSet*(*Gen*) form the antecedent(Ant) part of the If–Then rules. The class label can be assigned by following the below-mentioned computations in Eq. (8).

$$\frac{\mathrm{Ant}_{ij}}{\mathrm{Cons}_j} \tag{8}$$

where Ant_{ij}—number of instances where antecedent is true for ith chromosome and jth class, Cons_j—number of instances for jth class.

Particular Cons_j that is assigned as class label to the chromosome is the one for which $\mathrm{Ant}_{ij}/\mathrm{Cons}_j$ is highest.

3.2.11 Result depiction and analysis

Chromosomes added to the *KnowledgeSet*(*Gen*) generations after generations are listed in Table 6.

Rule Generation On computing the class label for the chromosomes in the knowledge set, the final rules that are inferred are given below

$\mathrm{BA}|B|\mathrm{Poor}|\mathrm{BA} \to F$

$\mathrm{BA}|B|\mathrm{Poor}|\mathrm{BA} \to \mathrm{BA}$

$\mathrm{BA}|B|\mathrm{Poor}|\mathrm{BA} \to \mathrm{VG}$

$\mathrm{Avg}|B|\mathrm{Fair}|F \to F$

$\mathrm{VG}|A|\mathrm{Excellent}|\mathrm{BA} \to \mathrm{Avg}$

$\mathrm{VG}|A|\mathrm{Excellent}|\mathrm{BA} \to \mathrm{BA}$

$\mathrm{BA}|B|\mathrm{Excellent}|\mathrm{BA} \to \mathrm{Avg}$

$\mathrm{Good}|A|\mathrm{VG}|F \to \mathrm{Avg}$

$\mathrm{VG}|A|\mathrm{Excellent}|\mathrm{Good} \to \mathrm{Avg}$

$\mathrm{VG}|A|\mathrm{Excellent}|\mathrm{Good} \to \mathrm{VG}$

⋮

$\mathrm{Good}|A|\mathrm{Poor}|\mathrm{Avg} \to \mathrm{VG}$

$\mathrm{VG}|A|\mathrm{Good}|\mathrm{BA} \to \mathrm{BA}$

$\mathrm{Avg}|A|\mathrm{VG}|F \to F$

$\mathrm{Avg}|A|\mathrm{VG}|F \to \mathrm{BA}$

$\mathrm{Avg}|A|\mathrm{VG}|F \to \mathrm{Good}$

Table 7 Rule interpretation

Rules	Rule interpretation
BA ∣ *B* ∣ Poor ∣ BA → *F*	If the continuous evaluation marks is *Below Average*, Level of student's satisfaction for the course faculty member is *Fair*, Student's attendance in the subject under consideration is *Poor* and the performance in Physics (base subject/course) which studied during the earlier semester is *Below Average*, then the student is likely to *Fail* in analog electronics' end semester examination
BA ∣ *B* ∣ Poor ∣ BA → BA	If the continuous evaluation marks is *Below Average*, Level of student's satisfaction for the course faculty member is *Fair*, Student's attendance in the subject under consideration is *Poor* and the performance in Physics (base subject/course) which studied during the earlier semester is *Below Average*, then the student's performance is likely to be *Below Average* in analog electronics' end semester examination
BA ∣ *B* ∣ Poor ∣ BA → VG	If the continuous evaluation marks is *Below Average*, Level of student's satisfaction for the course faculty member is *Fair*, Student's attendance in the subject under consideration is *Poor* and the performance in Physics (base subject/course) which studied during the earlier semester is *Below Average*, then the student's performance is likely to be *Very Good* in analog electronics' end semester examination
Avg ∣ *B* ∣ Fair ∣ *F* → *F*	If the continuous evaluation marks is *Average*, Level of student's satisfaction for the course faculty member is *Fair*, Student's attendance in the subject under consideration is *Fair* and the performance in Physics (base subject/course) which studied during the earlier semester is *Fail*, then the student is likely to *Fail* in analog electronics' end semester examination
VG ∣ *A* ∣ Excellent ∣ BA → Avg	If the continuous evaluation marks is *Very Good*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Excellent* and the performance in Physics (base subject/course) which studied during the earlier semester is *Below Average*, then the student's performance is likely to be *Average* in analog electronics' end semester examination
VG ∣ *A* ∣ Excellent ∣ BA → BA	If the continuous evaluation marks is *Very Good*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Excellent* and the performance in Physics (base subject/course) which studied during the earlier semester is *Below Average*, then the student's performance is likely to be *Below Average* in analog electronics' end semester examination
BA ∣ *B* ∣ Excellent ∣ BA → Avg	If the continuous evaluation marks is *Below Average*, Level of student's satisfaction for the course faculty member is *Fair*, Student's attendance in the subject under consideration is *Excellent* and the performance in Physics (base subject/course) which studied during the earlier semester is *Below Average*, then the student's performance is likely to be *Average* in analog electronics' end semester examination
Good ∣ *A* ∣ VG ∣ *F* → Avg	If the continuous evaluation marks is *Good*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Very Good* and the performance in Physics (base subject/course) which studied during the earlier semester is *Fail*, then the student's performance is likely to be *Average* in analog electronics' end semester examination
VG ∣ *A* ∣ Excellent ∣ Good → Avg	If the continuous evaluation marks is *Very Good*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Excellent* and the performance in Physics (base subject/course) which studied during the earlier semester is *Good*, then the student's performance is likely to be *Average* in analog electronics' end semester examination
VG ∣ *A* ∣ Excellent ∣ Good → VG	If the continuous evaluation marks is *Very Good*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Excellent* and the performance in Physics (base subject/course) which studied during the earlier semester is *Good*, then the student's performance is likely to be *Very Good* in analog electronics' end semester examination
⋮	⋮
Good ∣ *A* ∣ Poor ∣ Avg → VG	If the continuous evaluation marks is *Good*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Poor* and the performance in Physics (base subject/course) which studied during the earlier semester is *Average*, then the student's performance is likely to be *Very Good* in analog electronics' end semester examination
VG ∣ *A* ∣ Good ∣ BA → BA	If the continuous evaluation marks is *Very Good*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Good* and the performance in Physics (base subject/course) which studied during the earlier semester is *Below Average* the student's performance is likely to be *Below Average* in analog electronics' end semester examination
Avg ∣ *A* ∣ VG ∣ *F* → *F*	If the continuous evaluation marks is *Average*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Very Good* and the performance in Physics (base subject/course) which studied during the earlier semester is *Fail*, then the student is likely to *Fail* in analog electronics' end semester examination
Avg ∣ *A* ∣ VG ∣ *F* → BA	If the continuous evaluation marks is *Average*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Very Good* and the performance in Physics (base subject/course) which studied during the earlier semester is *Fail*, then the student's performance is likely to be *Below Average* in analog electronics' end semester examination
Avg ∣ *A* ∣ VG ∣ *F* → Good	If the continuous evaluation marks is *Average*, Level of student's satisfaction for the course faculty member is *Good*, Student's attendance in the subject under consideration is *Very Good* and the performance in Physics (base subject/course) which studied during the earlier semester is *Fail*, then the student's performance is likely to be *Good* in analog electronics' end semester examination

Rule interpretation Considering the valid gene number (VGN) used in the study, i.e., VGN = 1, 2, 4, 5. The above cited rules can be interpreted as mentioned below in Table 7.

Result validation In knowledge-centric environment the validity, significance and relevance of knowledge have the force to guide decisions. It should be understood that the relevance of the knowledge is of prime importance as the its deployment is only effective if it fits into the scenario, i.e., the holistic positive impact of deployment of knowledge can only be achieved if the inferred knowledge is appropriate for the situation. Knowledge could be represented in form of If–Then rules and could be evaluated by using various metrics.

For each of the rules acquired by implementing the technique, the value of classification (true positive, TP) and misclassification (false positive, FP), true negative (TN) and false negative (FN) was recorded.

True positive (TP)	Record count from the data set satisfying *A* and *C*
False positives (FP)	Record count from the data set satisfying *A* but not *C*
True negative (TN)	Record count from the data set not satisfying *A* but satisfying *C*
False negative (FN)	Record count from the data set not satisfying *A* nor *C*

where *A*—antecedent of the rule, *C*—consequent of the rule.

The rules were further evaluated and validated on the basis of Kappa coefficient (*K*) that measures the degree of concordance, i.e., degree of homogeneity in the ratings, for categorical data. Value of Kappa coefficient is either 1 or <1. In case the value of *K* is 1, it signifies complete agreement. Equation (9) exhibits the calculation of Kappa.

$$K = \frac{(\mathrm{TP}+\mathrm{TN}) - \frac{(\mathrm{TP}+\mathrm{FN})(\mathrm{TP}+\mathrm{FP})+(\mathrm{FP}+\mathrm{FN})(\mathrm{FN}+\mathrm{TN})}{N}}{N - \frac{(\mathrm{TP}+\mathrm{FN})(\mathrm{TP}+\mathrm{FP})+(\mathrm{FP}+\mathrm{FN})(\mathrm{FN}+\mathrm{TN})}{N}} \tag{9}$$

The Table 8 exhibits the Kappa coefficient values for the rules obtained.

It can be seen that some of the rules have got considerably high Kappa value.

Result analysis Though the rules, the relationship existing between the various attributes can be witnessed, i.e., the influence of the independent variables on the dependent variable can be observed.

The Kappa statistic (or coefficient) is a metric that compares an observed accuracy with an expected accuracy (accuracy by random chance). The Kappa statistic not only evaluates a single classifier, but also evaluates classifiers among themselves. It is more meaningful than simply using accuracy as a metric as it takes random chance into account. Kappa statistics depicts the measure of reliability between two rating bodies: one reflecting the actual values of each instance to be classified, and the other reflecting classifier used to perform the classification in machine learning.

Table 8 Kappa coefficient

Rules	Kappa statistics
BA ∣ *B* ∣ Poor ∣ BA → *F*	0.458678332
BA ∣ *B* ∣ Poor ∣ BA → BA	0.779564572
BA ∣ *B* ∣ Poor ∣ BA → VG	0.589563974
Avg ∣ *B* ∣ Fair ∣ *F* → *F*	0.836607961
VG ∣ *A* ∣ Excellent ∣ BA → BA	0.711304681
VG ∣ *A* ∣ Excellent ∣ BA → Avg	0.651407393
BA ∣ *B* ∣ Excellent ∣ BA → Avg	0.738782319
Good ∣ *A* ∣ VG ∣ *F* → Avg	0.641653561
VG ∣ *A* ∣ Excellent ∣ Good → Avg	0.688679245
VG ∣ *A* ∣ Excellent ∣ Good → VG	0.796340908
Good ∣ *A* ∣ Poor ∣ Avg → VG	0.701058201
VG ∣ *A* ∣ Good ∣ BA → BA	0.841543931
Avg ∣ *A* ∣ VG ∣ *F* → *F*	0.865213523
Avg ∣ *A* ∣ VG ∣ *F* → BA	0.61659436
Avg ∣ *A* ∣ VG ∣ *F* → Good	0.811830461

The study depicts that out of the 15 rules extracted; only 6 rules bear a Kappa value of more than 0.75 (which makes them significant).

On observing and analyzing the significant rules listed in Table 9, it can be seen the performance of the student in the end semester examination is significantly dependent on the students' performance in the base subject.

Similar results were established through [28] where the authors established the fact that out of all the independent attributes taken up in the study, the performance of the student in the base subject significantly affects the performance of the student thereafter. The below Table 9 exhibits the rules having Kappa coefficient values >0.75 along with the rule interpretation.

It is observed that the rules were validated and analyzed on the basis of their Kappa coefficients. The inferred knowledge through these rules that elaborated the major effect of the base subject on the performance of the student in the end term examination of the subject under scrutiny was recorded. It was then passed to the curriculum design committee to initiate suitable steps (e.g.) starting of a bridge course to explain the fundamental principles of physics related to the subject of analog electronics or assigning of an experience faculty to teach physics at second semester level. Hence application of the knowledge in real life scenario further validated the process.

Table 9 Rules with Kappa coefficient >0.75

Rules	Kappa statistics	Rule interpretation
BA ∣ *B* ∣ Poor ∣ BA → BA	0.779564572	If the continuous evaluation marks is Below Average, Level of student's satisfaction for the course faculty member is Fair, Student's attendance in the subject under consideration is Poor and the performance in Physics (base subject/course) which studied during the earlier semester is Below Average, then the student's performance is likely to be Below Average in analog electronics' end semester examination
Avg ∣ *B* ∣ Fair ∣ *F* → *F*	0.836607961	If the continuous evaluation marks is Average, Level of student's satisfaction for the course faculty member is Fair, Student's attendance in the subject under consideration is Fair and the performance in Physics (base subject/course) which studied during the earlier semester is Fail, then the student is likely to Fail in analog electronics' end semester examination
VG ∣ *A* ∣ Excellent ∣ Good → VG	0.796340908	If the continuous evaluation marks is Very Good, Level of student's satisfaction for the course faculty member is Good, Student's attendance in the subject under consideration is Excellent and the performance in Physics (base subject/course) which studied during the earlier semester is Good, then the student's performance is likely to be Very Good in analog electronics' end semester examination
VG ∣ *A* ∣ Good ∣ BA → BA	0.841543931	If the continuous evaluation marks is Very Good, Level of student's satisfaction for the course faculty member is Good, Student's attendance in the subject under consideration is Good and the performance in Physics (base subject/course) which studied during the earlier semester is Below Average the student's performance is likely to be Below Average in analog electronics' end semester examination
Avg ∣ *A* ∣ VG ∣ *F* → *F*	0.865213523	If the continuous evaluation marks is Average, Level of student's satisfaction for the course faculty member is Good, Student's attendance in the subject under consideration is Very Good and the performance in Physics (base subject/course) which studied during the earlier semester is Fail then the student is likely to Fail in analog electronics' end semester examination
Avg ∣ *A* ∣ VG ∣ *F* → Good	0.811830461	If the continuous evaluation marks is Average, Level of student's satisfaction for the course faculty member is Good, Student's attendance in the subject under consideration is Very Good and the performance in Physics (base subject/course) which studied during the earlier semester is Fail, then the student's performance is likely to be Good in analog electronics' end semester examination

4 Conclusion

Using GA for optimization helps for excavation of patterns and establishing relationships between attributes, non-evident, openly. The study takes up multi-objective classification problem in higher education domain for creating knowledge-centric education system by coining unique and non-trivial rules. The knowledge derived in the form of rules bears relevance in the context of the domain and hence can be added to the knowledge set that can supplement the rule-based system for appropriate syllabus planning, designing structured lesson plans, structuring criteria for evaluation of the student's performance and implementation of appropriate instructive practices for improving the academic performance in a larger perspective.

As a future work, other aspects of academics like acquiring placement after program completion can be considered.

Acknowledgements Authors would like thank the anonymous reviewers for their valuable suggestions and comments to improvise the manuscript.

Compliance with ethical standards

Conflict of interest Authors declare that they do not have any conflict of interest.

References

1. Rostami NA (2014) Integration of business intelligence and knowledge management—a literature review. J Intell Stud Bus 4:30–40
2. Surbakti H (2015) Integrating knowledge management and business intelligence processes for empowering government business organizations. Int J Comput Appl 114:36–45
3. Ashrafi N, Kelleher L, Kuilboer JP (2014) The impact of business intelligence on healthcare delivery in the USA. Interdiscip J Inf Knowl Manag 9:117–130
4. Sowa JF (2000) Knowledge representation: logical, philosophical and computational foundation. Brook/Cole, Pacific Grove
5. Rich E, Knight K, Nair SB (2010) Artificial intelligence. TMH, New Delhi
6. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley Longman, Boston
7. Jong KAD, Spears WM, Gordon DF (1993) Using genetic algorithms for concept learning. Mach Learn. doi:10.1007/BF00993042

8. Aickelin U (2002) An indirect genetic algorithm for set covering problems. J Oper Res Soc 53(10):1118–1126
9. Schaffer JD (1985) Multiple objective optimization with vector evaluated genetic algorithms. In: Grefenstette JJ (ed) Proceedings of first international conference on genetic algorithms, pp 93–100
10. Han J, Kamber M (2000) Data mining: concepts and techniques. Morgan Kaufmann Publishers, Burlington
11. Polanyi M (1966) The tacit dimension. Routledge and Kegan Paul, London
12. Nickols FW (2000) The knowledge in knowledge management. In: Cortada JW, Woods JA (eds) The knowledge management yearbook 2000–2001. Butterworth-Heinemann, Boston, pp 12–21
13. Jong TD, Ferguson Hesler MGM (1996) Types and quality of knowledge. Educ Psychol 31:105–113
14. Gupta P, Mehrotra D, Singh R (2012) Achieving excellence through knowledge mapping in higher education institution. In: Proceedings of international conference on recent advances and future trends in information technology (iRAFIT 2012), Patiala, iRAFIT(8), pp 5–10
15. Rusli A, Selamat MH, Jaafar A, Abdullah S, Sura S (2008) An empirical study of knowledge management system implementation in public higher learning institution. Int J Comput Sci Netw Secur 8:281–290
16. Sahay A, Mehta K (2010) Assisting higher education in assessing, predicting, and managing issues related to student success: a web-based software using data mining and quality function deployment. In: Proceedings of Academic and Business Research Institute Conference, Las Vegas, pp 1–12
17. Sagsan M (2009) Knowledge management discipline: test for an undergraduate program in Turkey. Electron J Knowl Manag
18. Wright H (2008) Tacit knowledge and pedagogy at UK universities; challenges for effective management. Electron J Knowl Manag 6(1):49–62
19. Cranfield DJ, Taylor J (2008) Knowledge management and higher education: a UK case study. Electron J Knowl Manag 6(2):85–100
20. Dutta D, Dutta P (2010) Optimized classification rule set generation using pareto based binary coded elitist multi objective genetic algorithm. In: Conference proceeding of 4th international conference on intelligent systems and networks Haryana, India
21. Dutta D, Sil J (2013) Evolution of genetic algorithms in classification rule mining. In: Bhattacharyya S, Dutta P (ed) Handbook of research on computational intelligence for engineering, science, and business, vol 1. IGI Global, Hershey, PA, USA, pp 328–363
22. Zhong-Yang X, Lei Z, Yu-Fang Z (2004) A classification rule mining method using hybrid genetic algorithm. In: Proceedings of IEEE region 10 conference (TENCON 2004) Chiang Mai, Thailand: IEEE, pp 207–210
23. Dehuri S, Patnaik S, Ghosh A et al (2008) Application of elitist multi-objective genetic algorithm for classification rule generation. Appl Soft Comput 8(1):477–487. doi:10.1016/j.asoc.2007.02.009
24. Dutta D (2009) Classification rules generation for iris data using lexicographic pareto based multi objective genetic algorithm. In: Conference proceedings of 3rd international conference on intelligent systems and networks (IISN 2009), At Klawad, Haryana, India, pp 173–177
25. Kaya M (2010) Autonomous classifiers with understandable rule using multi-objective genetic algorithms. Expert Syst Appl 37(4):3489–3494. doi:10.1016/j.eswa.2009.10.039
26. Dutta D, Dutta P (2011) A real coded MOGA for mining classification rules with missing attribute values. In: Proceedings of international conference on communication, computing and security, India, pp 355–358
27. Gupta P, Mehrotra D, Sharma TK (2015) Genetic based weighted aggregation model for optimization of student's performance in higher education. In: Proceedings of fifth international conference on soft computing for problem solving volume 436 of the series, advances in intelligent systems and computing, pp 877–887
28. Gupta P, Mehrotra D, Sharma TK (2016) Role of decision tree in supplementing tacit knowledge for hypothetico-deduction in higher education. Int J Syst Assur Eng Manag. doi:10.1007/s13198-016-0483-6
29. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multi objective genetic algorithm: NSGA—II. IEEE Trans Evol Comput 6(2):182–197
30. Zitzler E, Thiele L (1998) An evolutionary algorithm for multiobjective optimization: the strength pareto approach computer engineering and networks laboratory (TIK). Zürich, Swiss Federal Institute of Technology Zürich (ETH). doi:10.3929/ethz-a-004288833