Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Technology 1 (2012) 44 – 49

Procedia Technology

INSODE - 2011

Using Artificial Intelligence Techniques for Large Scale Set Partitioning Problems

Musa Peker[a] *,Baha Sen[a], Safak Bayir[b]

[a]Karabük University, Computer Engineering Department, 78050, Karabük, Turkey
[b]Karabük University, Distance Education Research and Practice Centre, 78050, Karabük, Turkey

Abstract

Set partitioning problems are among NP-Hard problems due to their complexities. It is difficult to prepare an algorithm that will give a precise solution in these types of problems which are difficult to solve. This study proposes a genetic algorithm based approach among artificial intelligence optimization algorithms in order to find simpler solutions to set partitioning problems. The proposed method was applied to the solution of problem of partitioning the 53 teams in the Turkish Third League into 5 subsets. The distribution of the teams into subsets was undertaken with the aim of minimizing the travel costs and the travel fatigue and preventing the subjective distinctions in the determination of subsets. This study was carried out for 2009-2010 football season and it includes a comparison of the plan prepared by Turkish Football Federation with the plan obtained through the study. The application software was developed by using C# programming language of .Net technologies.

Keywords: Set Partitioning Problems, Genetic Algorithm, NP-Hard Problem

1. Introduction

Set partitioning problems (SPP) deal with distributing the elements of a group to specific number of sub-groups according to a given purpose. These problems are the best known examples of combinatorial problems. SPPs can be implemented in many fields such as workforce planning, vehicle routing, identification of sub-groups in football leagues and team scheduling. SPPs are among the most difficult (NP-Hard) problems to solve due to their complexities [1]. Solving these problems utilising linear programming techniques and traditional algorithms require difficult and long processes. Therefore, various researches have been being carried out to obtain more available algorithms in order to solve these kinds of problems [2].

Linderoth, Lee and Savelsbergh proposed a parallel linear programming-based heuristic approach in their studies for large-scale SPPs [3]. Hoffman and Padberg also used linear based programming approaches [4]. Joseph suggested a synchronous operations frame in order to solve large-scale SPPs by dividing them into small subsets [5]. Klabjan proposed a practical algorithm with sub additive dual function computing [6]. Barahona and Anbil used volume algorithm and dual simplex method together [7]. It is seen that Chu and Beasley; Lin, Kao and Hsu applied genetic algorithm applications to solve set partitioning problems in different manners [8]. Güngör and Küçüksille used genetic algorithm approach to solve SPPs [2].

* Corresponding author. Tel.: +90-370 433 20 21/187; Fax: +90-370 433 322 90.
E-mail address: pekermusa@gmail.com

doi:10.1016/j.protcy.2012.02.010

An implementation on sports leagues was used in this study in order to find a solution to SPPs. Teams in sub-groups travel to the cities of the opponent teams to play the matches during the football league in the 2nd and 3rd football leagues in Turkey. The distance between cities is an important criterion in the identification of sub-groups since the less is the distance, the less the travel costs and fatigue of the players will be. However, the biggest problem lies in determining the groups for the teams that are situated in the borders and subjective assessments are inevitable in areas like football which attracts intense discussions and deception. The problem faced by many sports leagues is handled as a SPP. In the studies that search the solution for identification of sub-groups in sports leagues, 0-1 integer linear programming models are investigated [2][9]. In order to set the 0-1 integer linear programming, it is first necessary to form the set partitioning table [10]. It requires many operations to form the table mentioned above and to put forward the 0-1 integer linear programming model [10].

This study utilises a genetic algorithm approach as an alternative method to traditional algorithms in order to minimise the problems mentioned above.

2. Set Partition Problems (SPPs)

A big majority of discontinuous optimization problems involve integer linear programming problems [2]. A part of 0-1 integer linear programming problems consist of set covering and SPPs [11]. These problems are the most common of combinatorial problems and have many important applications in several areas such as scheduling, locations, routing etc. [1]. SPPs are among the most difficult (NP-Hard) problems to solve due to their complexities. Since traditional algorithms and techniques are insufficient to solve problems classified as NP-Hard problems, new artificial intelligence techniques have been developed. Formally, the SPP can be written as follows [12]:

$$z_{sp} = \min \sum_{j \in J} C_j X_j \quad (1)$$

$$X_j \in \{0,1\} (j \in J) \quad (3)$$

$$\text{subject to: } \sum_{j \in J} a_{ij} X_j = 1 \quad (i \in R) \quad (2)$$

Here, R is the set of constraints or rows (ground set) and J is the collection of subsets or columns. The matrix $A=\{a_{rj}\}$ is defined such that a_{rj} is equal to 1 if subset j contains row r and 0 otherwise [12]. The costs of a subset j are given by c_j. Furthermore, we define R(j) as the set of rows that are contained in subset j and J(r) as the collection of subsets that contain row r. Without loss of generality we assume that the costs vector c is integer [12].

3. Genetic Algorithms

Genetic algorithms are adaptive heuristic search algorithms based on the principle of creation of new series by chromosomes in order to solve complex regular problems [13]. At present, the application field of genetic algorithms is increasing. Some of these fields are atelier scheduling, artificial neural networks design, image control, optimization, expert systems, packaging problems, travelling salesman problem, economic modelling etc. [13].

4. Genetic Operators

Genetic operators are operations that are implemented on the existing population. The aim of this operation is to create generations with better qualities and to expand the field of the search algorithm [13]. Although different operators are used with different applications, there are three standard operators in genetic algorithms. These are selection, crossover and mutation operators.

4.1. Selection

The parents that will go through crossover are selected from among the individuals that form the community. The main problem here is how to select the individuals for crossover [14]. Many methods such as roulette wheel selection, Boltzman selection, tournament selection, rank selection and steady-state selection can be used in the selection of parents. This study uses roulette wheel selection method (see Fig 3). In this method, parents are selected according to their level of suitability of fit hence it is more probable (*Pselection*) for individuals with higher levels to be selected [14].

$$P_{selection} = u_i / \sum u \quad (4)$$

Here, u_i shows the value of suitability of the individual, whereas Σu is the total value of suitability for the individuals in the community.

4.2. Crossover

It works on the principle of creating a new solution by utilizing the structures of two solutions. Crossover operation usually occurs as the exchange of the parts of the binary series [15]. Since different coding methods are used in different applications, there are also different crossover methods: single-point, position-defined, sequence-defined, circular, linear and serial. The study uses serial crossover from crossover methods explained earlier in the chapter (see Fig 3). In this method, 2 chromosomes are randomly selected from the gene pool. 2 cut-of points are randomly determined on these chromosomes [16]. It is important to ensure having the same number of chromosomes between these two cut off points in both of the chromosomes. The chromosomes between the cut-off points are exchanged [17]. If repetitive genes form in the genes that are in the zone which is outside the cut-off point, these are replaced by genes that do not exist in the chromosome (written from left to right).

4.3. Mutation

The purpose is to create a new chromosome by changing one or more of the genes of the existing chromosome [15]. Through new and continuous gene reproduction, the chromosomes of the future generation may repeat each other after a certain period and this may cause the termination or reduction in the production of different chromosomes [15]. Because of this reason, some of the chromosomes are put through mutation in order to create diversity in the chromosomes of the generation.

5. Turkish Third Football League Application

Turkey Third Football League consists of 53 teams in 2009-2010 football season. The third league covers 53 teams including 5 sub-groups. When forming these sub-groups, it is planned to place the geographically closer teams in the same group in order to decrease the transportation costs and fatigue for footballers and the fans. Through genetic algorithm approach, this study searches a solution for the problem that is in the framework of SPPs.

5.1. Formation of initial population

A data set was formed as the first stage of forming the initial population. The data set consisted of the distances of the teams to each other. The information about distances taken from Turkish Republic Highways was transferred into a two-dimensional series [17]. The first value in these series displays the first team while the second value is for the second team. After transferring all the values for distances into the series, a matrix was formed to show the distances of the teams from each other.

After these stages, the structure of the chromosome that declares the individuals in the Genetic Algorithm (GA) approach was determined. The formation of the chromosome structure expresses the representation of the solution. This study uses the permutation series presentation. In this presentation, it is easy to see which parts are in which location and it is possible to search for solution alternatives by implementing operations such as mutation and crossover [18]. The representation using integers made use of a series with one row and fifty three columns. That was randomly created. The structure of the chromosome used in the study is given in Figure 3.

Each numerical in the figure represents a team. The rule when determining the structure of the chromosome is as follows: The city for each team would be visited only once and the total distances between groups of teams would be the shortest.

5.2. Provision of the suitability of fit value

As it is known, while being reformed, chromosomes are selected according to the suitability of fit function and transferred to the new generation. The suitability of fit used here is the chromosome with the smallest value. The suitability of fit for the chromosome is the tour distance, i.e. the total of the distances between the cities [19]. The objective (z) function in the equation calculates the suitability of fit value for the chromosome.

$$z= \sum_{i=1}^{a} \sum_{j=1}^{a} x(i,j) + \sum_{i=a+1}^{b+a} \sum_{j=a+1}^{b+a} x(i,j) + \sum_{i=b+a+1}^{c+b+a} \sum_{j=b+a+1}^{c+b+a} x(i,j) + \sum_{i=c+b+a+1}^{(d+c+b+a)} \sum_{j=c+b+a+1}^{(d+b+c+a)} x(i,j) + \sum_{i=d+c+b+a+1}^{(e+d+c+b+a)} \sum_{j=d+c+b+a+1}^{(e+d+c+b+a)} x(i,j) \quad (5)$$

With the calculated objective function; for each group, total distance for travel was calculated and later values for each group is added to find the suitability of fit value for the chromosome. The number of elements in the groups is kept equal to the values range specified by the Turkey Football Federation in order to make a reliable comparison between the results obtained within the scope of the study and with those of the Federation.Accordingly, a, b, c, d and values vary between 9-12 values (number of teams) specified by the Turkey Football Federation.As a result of the experiments done within the settings of this research utilising trial and error approach, the following values which give the best results were obtained as follows respectively: a=12,b=10, c=9, d=11, e=11.

6. Application and Evaluation of Results

The application software was developed by C# programming language of .Net technologies. It was observed that crossover value of 0.7 and mutation rate of 0.003 produce effective and speedy results in this system where chromosome structures with 53 genes are used. The best solution is obtained in 3 minutes with the help of the software. Function list for the developed software interface is shown in Table 1.

Table 1: Function list for the developed software

Tools	Function
Form Team	Allows the formation of desired number of teams. Since this study determined sub-groups for 53 teams, team number is selected as 53. After clicking the form team button, the system randomly distributes the boxes with symbols on them throughout the map. Each symbol represents a team. In order to realize the optimization operation visually, the provinces or districts of the teams are identified on the map and the boxes are randomly placed on the coordinates of these points.
Start Simulation	Starts the optimization process and displays the paring of the teams as can be seen in Figure 1. Once simulation starts it keeps searching for the best solution unless stop button is clicked.
Stop Option	Stops the simulation.
Draw Graphic	Option makes it possible to represent the optimization results in the format of graphics.

The best result can be seen in Figure 1. Each sub-group in the image are shown as in different colours.

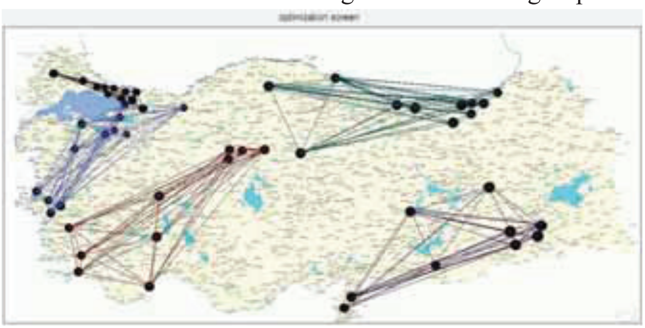

Figure1. Screenshot of the best result

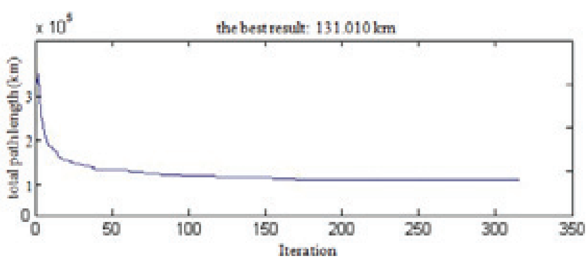

Figure2. Graphic representation of optimization realized by GA

Table2 displays the lists of the sub-groups and the distance for travel identified by this study. Table3 gives the lists of the sub-groups and the distance for travel identified by Turkey Football Federation. When the obtained values are compared, it is seen that the sub-group formation determined by this study saved 141.117-131.010=10.107 km. (around 7.2%) compared to the plan of the Football Federation. Figure 2 shows the values GA obtained for each generation until the optimum solution was reached.

7. Conclusion

The proposed method was applied to the solution of the problem of partitioning the 53 teams in the Turkey Third Football League into 5 subsets. It was observed that effective and speedy results were produced in this system where chromosome structures with 53 genes are used. The plan of the football federation prepared by using the suggested method for 2009-2010 football season creates cost savings of approximately 7.2% in terms of distance that needs to be travelled. The study draws attention by the speed, functional interface and simulation support.

With the study, it was seen that use of GA yielded good results in a short time in the solution of the problem of arrangement of league teams which was selected as an example of SPPs. The study can be developed and applied to other SPPs in order to increase performance in those areas. The selected example of organizing league teams can be developed with the same method and implemented to the leagues in Turkey and other countries to organize more economical leagues.

References

1.Çetin, E.I., Kuruüzüm, A. and Irmak, S.,"The Solution of Crew Scheduling Problem with Set Partitioning Model", Aviation and Space Technology Journal, 3 (4), 47-54, 2008

2.Güngör, I. and Küçüksille, E.,"Optimization of SPPs by Genetic Algorithms: Second Football League Application", Review of Social, Economic and Business Studies,5/6, Gazimağusa/KKTC, 2005

3.Linderoth, J.T., Lee, E.K. and Savelsbergh, M.W.P., "A Parallel, Linear Programming Based Heuristic for Large Scale SPPs", Informs Journal on Computing, 13/3, 191-209, 2001

4.Hoffman K.L. and Padberg M., "Set Covering, Packing and Partitioning Problems", Encyclopedia of Optimization, 3482-3486, 2009

5.Joseph, A., "A Concurrent Processing Framework For The SPP", Computers & Operations Research, 29/10, 1375-1391, 2002

6.Klabjan, D., "A Practical Algorithm for Computing a Subadditive Dual Function for Set Partitioning", Computational Optimization and Applications, 29/3, 347-368, 2004

7.Barahona, F. and Anbil, R., "On Some Difficult Linear Programs Coming From Set Partitioning", Discrete Applied Mathematics, 118/1, 3-11.,2002

8.Chu, P.C. and Beasley, J.E.,"Constraint Handling in Genetic Algorithms: The SPP", Journal of Heuristics, 4/4,323-357, 1998

9.Mitchell, J. E., "Realignment in the National Football League: Did They Do It Right?", Naval Research Logistics, 50/7, 683-701

10. Güngör, I. "Küme Bölme Problemlerinin Optimizasyonu ve Türkiye Futbol Liglerine Uygulanması", Doktora Tezi, Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, 1993

11.Van Krieken, M.G.C., Fleuren, H., Peeters, M.J.P., " Problem Reduction in SPP",Discussion Paper, Tilburg University,2003

12.Van Krieken, M.G.C., Fleuren, Hein and Peeters, Rene, "A Lagrangean Relaxation Based Algorithm for Solving Set Partitioning Problems", Center Discussion Paper No: 44, 2004

13.Mitchell, M. and Forest S., "Genetic Algorithms and Artificial Life", Vol. 1, No. 3, pp. 267-289. Reprinted in C. G. Langton (Ed.) Artificial Life: an Overview, MIT Press, Cambridge, MA, 1995

14.Kalenderli, Ö. and Güclü, A., "Yüksek Gerilim Tekniğinde Tek Değişkenli Optimizasyon Problemlerinin Genetik Algoritma ile Çözümü", Eleco'2004 Symposium on Electrical-Electronics and Computer Engineering, Bursa, pp. 248-252, December 8-12, 2004.

15. Işçi Ö. and Korukoğlu S., "Genetik Algoritma Yaklaşımı ve Yöneylem Araştırmasında Bir Uygulama", CBÜ IIBF Yönetim ve Ekonomi Dergisi(ISSN-1302-0064),Cilt 10 sayı:2,191-208,2003

16.Engin, O. and Fığlalı A., "Determining the Optimum Crossover Operator in the Solution of Flowshop Scheduling Problems By Genetic Algorithm", Dogus University Journal, 6, 27 - 35, 2002

17.http://www.kgm.gov.tr (Access Time: 15.03.2011)

18.Gülsün B.,Tuzkaya G. and Duman C., "Facility Layout Design with Genetic Algorithms and an Application", Dogus University Journal, Vol: 10 Issue: 1 Pages/record No: 73-87

19.Farajollahi, M., & Moenikia, M. (2011). The effect of computer-based learning on distance learners self regulated learning strategies. *World Journal on Educational Technology, 3*(1), 28-38.

20.Doyranlı Ş., "Solve of TSP with Genetic Algorithms",http://www.doruk~.net.tr/Ãrifprogramlar/delphiT, 1999 (Access time: 16.03.2011)

Appendix 1

Table 2.Plan obtained through the developed program

Groups	Sub-groups (Team number)
A	Balıkesirspor(1), Bandırmaspor(2), BursaNilüferSpor(4), İnegölSpor(6), OyakRenaultspor(8), Torbalıspor(12), Menemenspor(15), Altınordu(17),İzmirspor(20), Darıca Gençlerbirliği(24), Orhangazispor(28), Düzcespor(31)

B	Gaziosmanpaşa(3) , Lüleburgazspor(5), Alibeyköy(7), Çerkezköybelediyespor (10), Bayrampaşa(9), Küçükköyspor(11), Anadolu Üsküdar 1908(23), Beylerbeyi A.Ş(26), Maltepespor(30), Gölcükspor(29), Kartal Belediyepor(32)
C	Keçiören Sportif(13),Afyonkarahisarspor(16), Ispartaspor(14), Tekirovabelediyespor (19), Nazillibelediyespor (18), Muğlaspor(21), MarmarisGençlik(22), Keçiörengücü(25), MKE Kırıkkalespor(27), Ankara Demirspor(33)
D	SiirtSpor(45), Kırıkhansspor(46), Malatya belediyespor (47), Hatayspor(49), Bingölspor(48), Batmanspor(50), Batman Petrolspor(51), Diyarbakır Kayapınarspor(52), ŞanlıurfaBelediyespor(53)
E	Araklıspor(35), Pazarspor(34), YimpaşYozgatpor(36), Ünyespor(38), Yalıspor(37), Gümüşhanespor(41), Sürmenespor(40), Bafrabelediyespor (39), Arsinspor(43), Kastamonuspor(42), Bulancakspor(44)

Table 3.Plan determined by the Federation

Groups	Sub-groups (Team number)
A	Balıkesirspor(1),Bandırmaspor(2), Gaziosmanpaşa(3), BursaNilüferSpor(4), Lüleburgazspor(5), İnegölspor(6), Alibeyköy(7), OyakRenaultspor(8), Bayrampaşa(9), Çerkezköybelediyespor(10), Küçükköyspor(11)
B	Anadolu Üsküdar 1908 spor(23), Darıca Gençlerbirliği (24), Keçiörengücü(25), Beylerbeyi A.Ş (26), MKE Kırıkkalespor (27), Orhangazispor(28), Gölcükspor(29), Maltepespor(30),Düzcespor(31), Kartal Belediyespor (32), Ankara Demirspor (33)
C	Torbalıspor(12), Keçiören Sportif (13),Ispartaspor(14), Menemenspor(15), Afyonkarahisarspor(16), Altınordu(17), Nazillibelediyespor(18), Tekirovabelediyespor(19),İzmirspor(20),Muğlaspor(21), MarmarisGençlik (22)
D	Siirtspor(45),Kırıkhanspor(46),Malatya belediyespor(47), Bingölspor (48),Hatayspor(49),Batmanspor(50),Batman Petrolspor(51), Diyarbakır Kayapınarspor (52), ŞanlıurfaBelediyespor (53)
E	Pazarspor(34),Araklıspor(35),YimpaşYozgatspor(36),Yalıspor(37),Ünyespor(38),Bafrabelediyespor(39),Sürmenespor(40),Gümüşhanespor(41), Kastamonuspor(42),Arsinspor(43),Bulancakspor(44)

Appendix 2

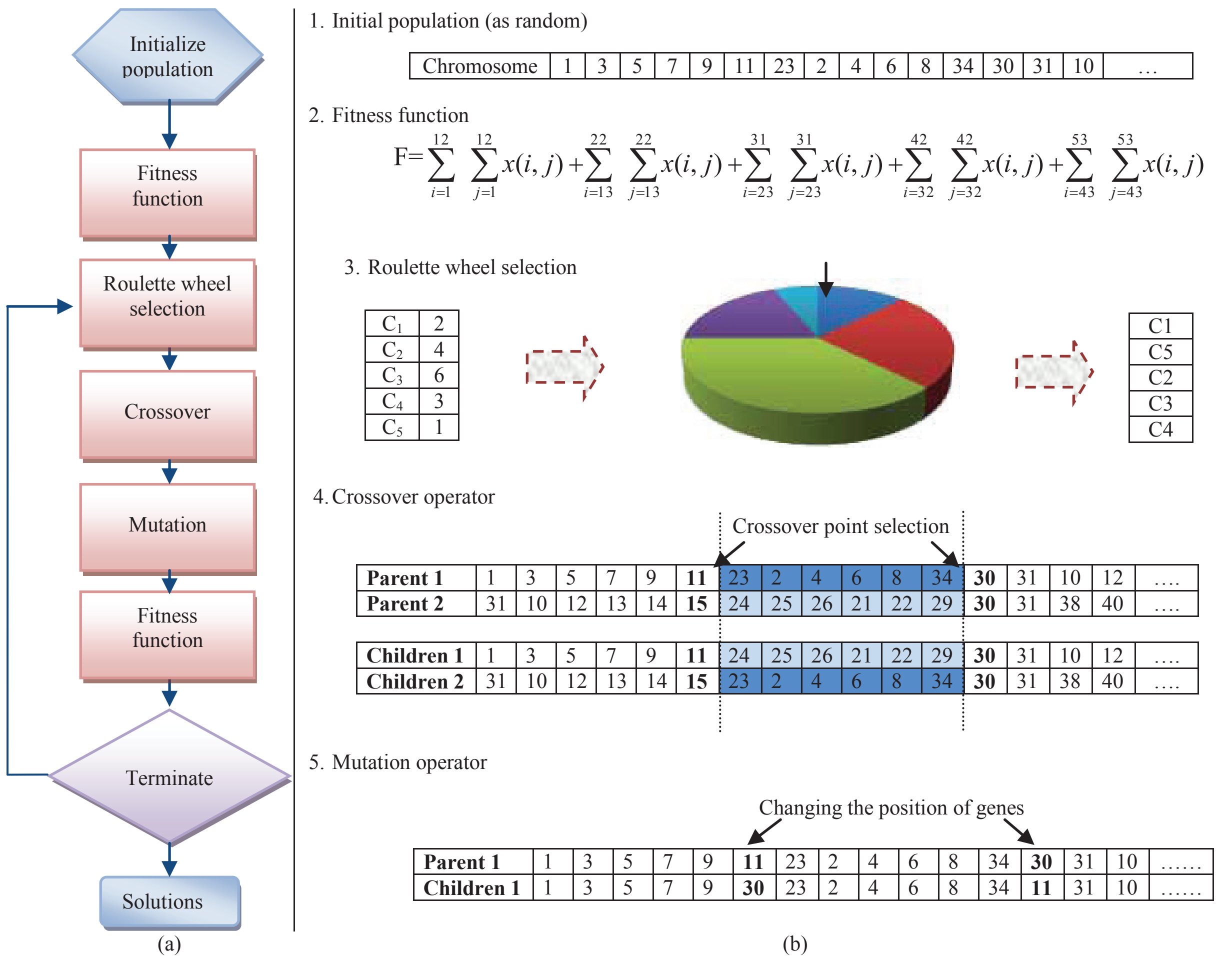

Figure 3.(a) Step of the genetic algorithms (b) Representation of genetic operators