DETERMINATION OF TEAMS IN GROUPS OF TURKISH FOOTBALL FEDERATION THIRD LEAGUE CLASSIFICATION GROUPS BY GENETIC ALGORITHMS

Dr. İbrahim Zeki Akyurt[1], Dr. Timur Keskintürk[2], Barış Kiremitci[3] Serap Kiremitci[4]

Abstract - *In this paper we have emphasized the use of genetic algorithms both as a tool and applicable to the formation of Sports League Groups. Determination of team members of each group is considered as a set partitioning problem which has diverse application fields. As most of the set partitioning problem variants are NP-hard combinatorial optimization problem, classical exact solution methods are confined to small size problem instances, we proposed a genetic algorithm approach for this problem. The primary objective of this paper is the determination of teams of each group for the purpose of minimizing the total travel distance taken by the teams of Turkish Football Federation Third League Classification Groups which have 5 separate groups and 53 different teams during the season of 2009-2010. While calculating the total travel distance matrices, distances between each team were obtained from Google Maps. The comparison of the computational results of proposed genetic algorithm and the actual Turkish Football Federation results was made and reported. It is found that almost 13 % percent improvement in solution value can be achieved. All the results were displayed in tables and on the maps.*

Key Words- *Football League, Genetic Algorithms, Set Partitioning Problem, Turkey*

SET PARTITIONING PROBLEM

The Set Partitioning Problem (SPP) is known to be NP-hard and it can be used to model many important real-world decision problems [1] including those involving railroad crew scheduling, truck deliveries, airline crew scheduling, tanker routing, information retrieval, switching circuit design, stock cutting, assembly line balancing, capital equipment decisions, location of offshore drilling platforms, some other facility location problems, political districting [2]. The problem is NP-Hard therefore heuristic algorithms have been developed. Since obtaining the results of the large size problems with standard methods, is very hard and complicated.

Set Partitioning problem can be formulated as follows:

$$Min \sum_{j}^{N} c_j x_j$$

$$Ax = e$$

$$x_j = \{0,1\}$$

A=(a_{ij}) is an *m*n* matrix which consists of 0 and 1 (elements). *c* is an arbitrary *n*-vector, *e=(1,…,1)*is an *m*-vector and N ={1,...,n}. If the rows of A are associated with the elements of the set M={1,...,m} and each column a_j of A with subset of M_j of those i € M such that a_{ij}=1, then SPP is the problem of finding a minimum-weight family of subsets M_j, j € N ,which is a partition of M, each subset being weighted with c_j .[2]

As SPP has wide applicability in many real – life problems, it has both exact and heuristics methods for its solution. The most common one of them, are those based on linear programming.

There are different methods among the exact methods. The most classic one is to solve SPP with linear programming. It has noted that obtaining the solution is very easy in small size instances with this one. [3]-[4]-[2]. The other method among the exact ones is branch and bound method.

[1] Dr. İbrahim Zeki Akyurt, Istanbul University, Faculty of Business Administration, Department of Production Management, Avcılar , Istanbul, Türkiye, akyurt@istanbul.edu.tr

[2] Dr. Timur Keskintürk, Istanbul University, Faculty of Business Administration, Department of Quantitative Methods, Avcılar , Istanbul, Türkiye, tkturk@istanbul.edu.tr

[3] Barış Kiremitçi, Istanbul University, The School of Transport and Logistics, Avcılar , Istanbul, Türkiye, baris@istanbul.edu.tr

[4] Serap Kiremitçi, Istanbul University, The School of Transport and Logistics, Avcılar , Istanbul, Türkiye, serapy@istanbul.edu.tr

There are many papers concerning heuristic methods for the solution of SPP [5][6][7][8][9][10] [11][12][13]. Genetic algorithms (GA) and Tabu Search are the heuristic methods used for. In references [14][15] [16][17][18][19], genetic algorithms is used for the solution of SPP and usually the application is about airline crew scheduling. Besides Tabu search method has also used to solve SPP [20][21].

In literature, Güngör ve Küçüksille [22] has used SPP in Football Leagues. They have seperated 51 teams into three groups in Turkish Football League B Cathegory with a GA based approach. In seperation of the groups, distances between the teams are minimized..

In this paper, the teams in Turkish Football Federation (TFF) Third League Classification Groups during the season of 2009-2010 are dealed with. There are 53 teams in 5 seperate groups. The first four groups are composed of 11 teams and the last group is composed of 9 teams. During the season, two matches are played between for each of the team and the other competitors in the same group. In the study comparison of the groups determined by the Turkish Football Federation (TFF) and by the proposed genetic algorithm was made according to the total distance taken by the teams during the whole season. The other constraints (if any) are not taken into consideration in the group assignment process.

GENETIC ALGORITHMS FOR SP PROBLEM AND APPLICATION

Genetic algorithms (GAs) which is a population-based meta-heuristic technique, was developed by Holland.[23][24] (Goldberg, 1989; Reeves, 1995) GA evolves a population of individuals encoded as chromosomes by creating new generations of offspring through an iterative process until some convergence criteria are met[25]. Solution values (variables) are represented in the vectors named chromosomes. This representation may be binary coding or actual values can also be used. For a given number interval, initial solutions are generated with values determined randomly. This group of solutions produced with the number of chromosomes (population size) is called Initial Population. Then the quality of each solution candidates is evaluated according to the problem-specific fitness function. Fitness function in our model is considered as minimization of the total travel distance. The creation of new generation population belonging to the next iterations, are made through GA operators like selection, crossover and mutation. The steps of GA are as follows:

Step 1: Generation of Initial Population
Step 2: Determination of Fitness Value
Step 3: Selection
Step 4: Crossover
Step 5: Mutation
Step 6: If stopping criterion is not achieved, go to the second step.
Step 7: Select the best solution as the result.[26][27]

The objective of this paper is the determination of teams of each group for the purpose of minimizing the total travel distance taken by the teams of Turkish Football Federation (TFF). The problem formulation and the notation is given as follows:

Notations:

N	:	Number of Teams
n	:	Team index n=1,..,N
K	:	Number of Groups
k	:	Group index k=1,.., K
V	:	Set of Teams
T_k	:	Number of teams in group k
S_k	:	Set of teams in group k, $S \leq V$
$c_{(x_i x_j)}$	:	Distance between team x_i and x_j, $\forall\ (i,j) \in V$
x_i	:	Problem variables, $\forall\ i \in V, x_i = [1, N]$

and integer formulation of Mathematical Model (Problem) is

$$min \sum_{k=1}^{K} \sum_{i \in S_k} \sum_{j \in S_k} c_{(x_i, x_j)}, \quad i \neq j$$

Subject to:

$$|S_k| = T_k, \quad k = 1, \ldots, K$$

x_i=[1,N] and integer

In the problem, there are integer variables for each of the teams. These integer variables can take values between 1 and the total team number (N). According to this, there are variables as many as the team numbers and they can take integer values between [1, N] interval. In GA, variables are defined in genes within chromosomes. Each of the chromosomes represents an alternative solution as shown in Figure 1.

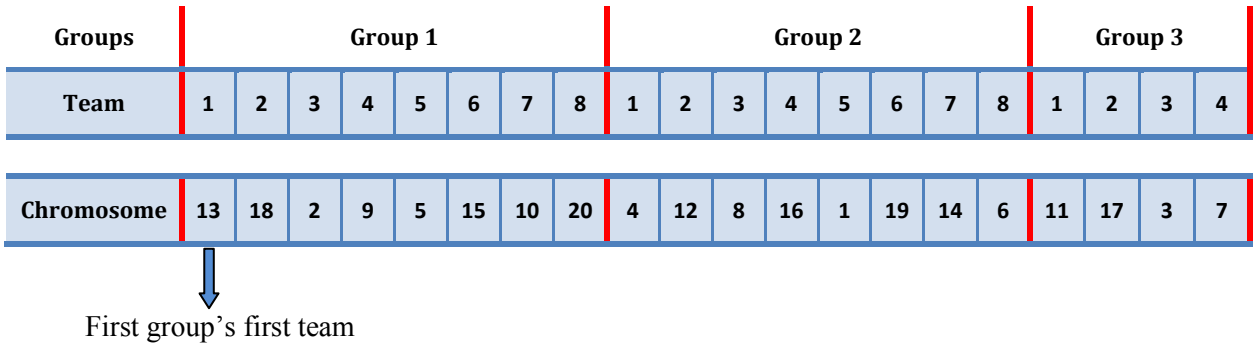

FIGURE 1
Chromosome Example

As can be seen in example chromosome, there are three groups which are composed of 8, 8 and 4 teams respectively and there are 20 teams totally. Permutation encoding is used here and each of the genes represents a team which takes place in the related group. Fitness function is the sum of total travel distance of each teams with all the others in the same group, reciprocally as indicated in the equation.

Crossover and mutation are two important operators that make changes in existing chromosomes in the search for better solutions. The aim of the crossover is to exchange of information between the chromosomes, so it enables creating new and better individuals (chromosomes). In binary and real-value coding, crossover step is taken by changing one side of the predetermined crossover point between two strings reciprocally. Thus, two new different individuals are obtained. In permutaion encoding in order to prevent unfeasible solutions, one-point crossover is used like in Figure 2. Mutation consists of randomly modifiying some gene(s) of a single individual at a time to further explore the solution space and ensure, or preserve, genetic diversity. The occurance of mutation is generally associated with a low probability [27].

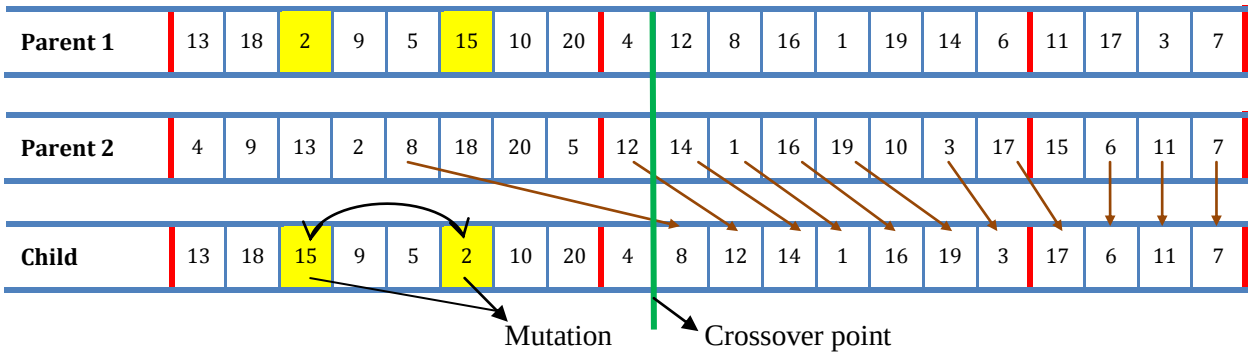

FIGURE 2
Creation of New Generation

Modified population by crossover and mutation steps, are eliticized with Selection step. ile değiştirilen population Selection ile daha elit hale getirilir. Chormosomes with a better fitness value, have higher probability of being chosen for the next generation. Ones with worse fitness are eliminated and ones with better fitness are reproduced and the population size remains constant. This loop continues till getting the predetermined iteration number.

In this study, the object is the determination of teams of each group for the purpose of minimizing the total travel distance taken by the teams of Turkish Football Federation Third League Classification Groups which have 5 separate groups and 53 different teams during the season of 2009-2010. The first four groups are composed of 11 teams and the last group is composed of 9 teams. Distances between the teams are calculated by a developed program with using coordinates as road distance in terms of kilometers which have obtained from the website of Googlemaps.com.

In GA crossover rate was determined 0.9 mutation rate was determined 0.001 and the population size was determined 20. Algorithm was coded with MATLAB R2009 and implemented in an Intel(R) Pentium(R) 4 CPU 3.20 GHz, 480 MB RAM configured PC for 1000 iterations, 100 times. Comparative results are presented in Table 1.

TABLE 1
Results of The Problem

	TFF	**GAs (Mean)**
Total Distance	292164,4	253800,8
Average Deviation of Groups Distance	653,0317576	1371,493333
Average Deviation of Teams Distance	1239,386686	1561,851691
Standard Deviation of Teams Distance	1700,976937	2185,013234
Standard Deviation of Groups Distance	724,0845	1736,731471
CPU	-	15,0198732

When the results are considered, proposed GA result is a good alternative to TFF's result in terms of total distance. The difference between the results in total distance is 38363.6 km and approximately % 13.1 . It is considered that this improvement may enable serious benefits in economic, health and environment issues. Group based distributions of teams acoording to the TFF are shown in Turkish Map in Figure 3. Group based distributions of teams acoording to GA are shown in Turkish Map in Figure 4.

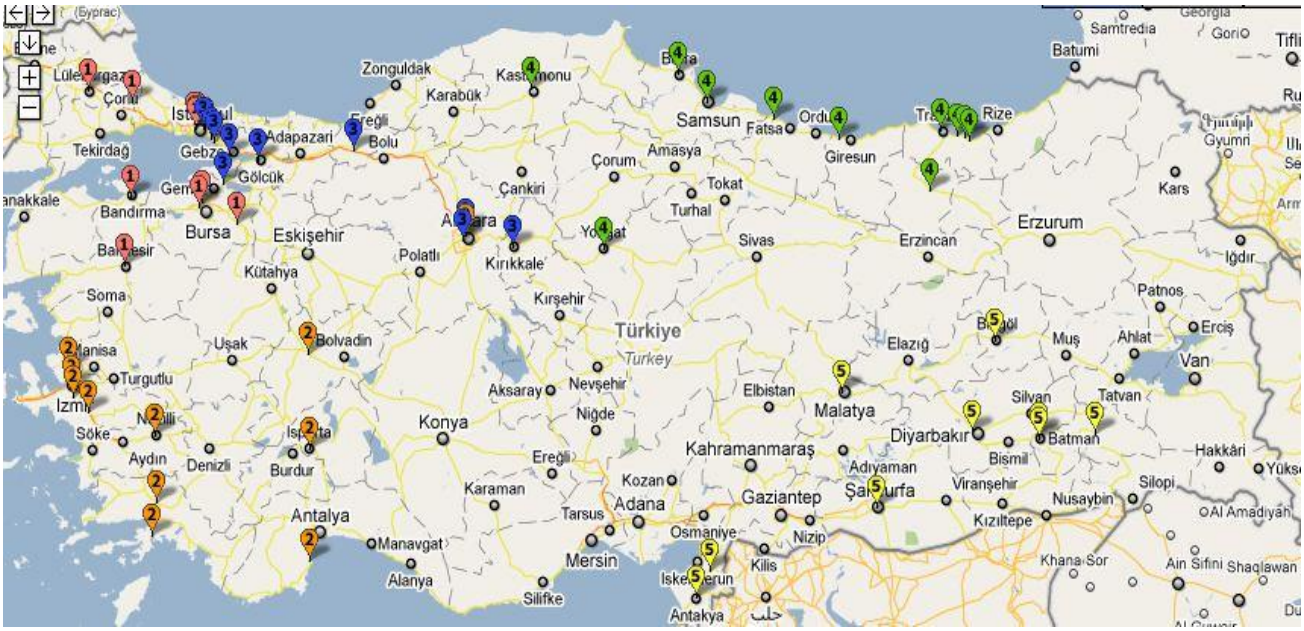

FIGURE 3
TFF's Current Group Distribution on Turkish Map

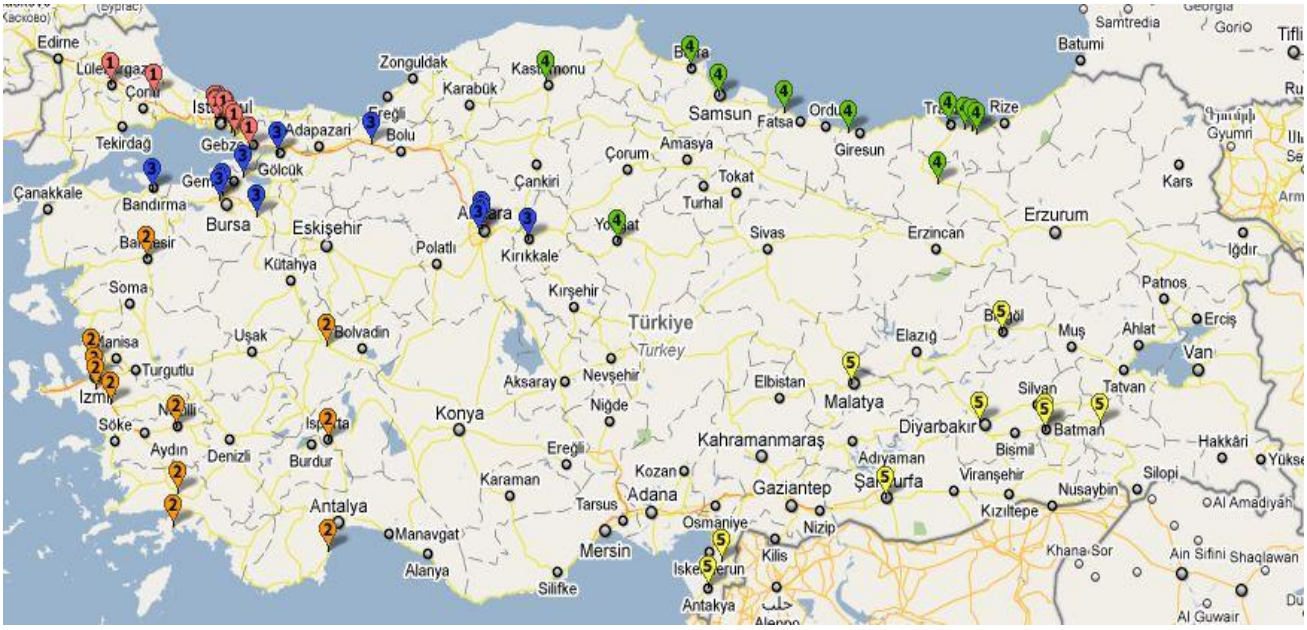

FIGURE 4
Proposed GAs Group Distribution on Turkish Map

Examining the figures, it can be seen that Proposed GA forms better groups than TFF's and this enables an improvement in total distance. But there might be imbalances between groups and teams as the objective function is the minimization of the total distance. This imbalance is clearly seen when you look at group's and team's means and deviations in Table 1.

CONCLUSIONS

In this paper, determination of Turkish Football Federation (TFF) 3. League classification groups with Genetic algorithms is presented as a set partitioning problem. The aim of the study is to provide an alternative solution with a scientific approach to TFF's current solution. Objective function is the minimization of the total distance taken by teams between the others in the same group. According to the results, an improvement which is approximately % 13.1 is obtained. This means 38363.6 km in terms of distance. So it can be noted that GA may be a good alternative. But when team's and group's means and deviations are taken into consideration, an increase in the imbalance between the groups and teams is seen ; although there is an improvement in total distance, assignments of teams to groups have more inequity. The competition is not only in the field in football and economic power is also a competition issue so such an assignment may prevent the balance of competition between teams and groups. According to this, it is not enough to determine the objective function that covers only minimization of total distance, it must include also balance of total distance between teams and groups.

As Genetic algorithm presents a good performance about total distance, it is taught to be the same algorithm can be used to balance the distances between teams and groups as well. With further reseach on this subject, transportation costs which are important cost items can be balanced in competitive intense sector like football.

REFERENCES

[1] Joseph, A., 2002, "A Concurrent Processing Framework For The Set Partitioning Problem", Computers&Operations Research, 29, 1375-1391

[2] Balas, E. and Padberg, M., 1976, "Set Partitioning: A Survey", SIAM Review, 18, 710–760

[3] Gershkoff, 1989, "Optimizing Flight Crew Schedules", Interfaces, 19, 29–43

[4] Marsten, R.E., 1974, "An Algorithm For Large Set Partitioning Problems", Management Science, 20, 774–787

[5] Baker, E.K., Bodin, L.D., Finnegan, W.F. and Ponder, R.J., 1979, "Efficient Heuristic Solutions To An Airline Crew Scheduling Problem", AIIE Transactions, 11, 79–85

[6] Balas, E. and Ho, A., 1980, "Set Covering Algorithms Using Cutting Planes, Heuristics, And Subgradient Optimization: A Computational Study", Mathematical Programming Study, 12, 37–60

[7] Vasko, F.J. and Wilson, G.R., 1984, "An Efficient Heuristic For Large Set Covering Problems", Naval Research Logistics Quarterly, 31, 163–171

[8] Ball, M. and Roberts, A., 1985, "A graph partitioning approach to airline crew scheduling", Transportation Science, 17, 4–31

[9] Ryan, D.M. and Falkner, J.C., 1988 "On The Integer Properties Of Scheduling Set Partitioning Models", European Journal of Operational Research, 35, 422–456

[10] Beasley, J.E., 1990, "A Lagrangian Heuristic For Set-Covering Problems", Naval Research Logistics, 37, 151–164

[11] Jacobs, L.W. and Brusco, M.J., 1993, "A Simulated Annealing-Based Heuristic For The Set-Covering Problem Working Paper", Operations Management and Information Systems Department, Northern Illinois University, Dekalb, IL.

[12] Sen, S., 1993, "Minimal Cost Set Covering Using Probabilistic Methods", In: Proc. 1993 ACM/SIGAPP Symposium on Applied Computing, 157–164

[13] Atamtürk, A., Nemhauser, G.L. and Savelsbergh, M.W.P., 1995, "A Combined Lagrangian, Linear Programming And Implication Heuristic For Large-Scale Set Partitioning Problems", Journal of Heuristics, 1, 247–259

[14] Liepins, G. E., Hilliard, M. R., Palmer, M. and Monow, M., 1987, "Greedy genetics", in Grefenstette J. J., editor, Genetic Algorithm and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms.

[15] Liepins, G. E., Hilliard, M. R., Richardson, J. and Palmer, M., 1990, "Genetic algorithms applications to set covering and traveling salesman problems", in Brown (ed), Operations Research and Artificial Intelligence: The Integration of Problem-Solving Strategies, Kluwer Academic Publishers.

[16] Al Sultan, K.S., Hussain, M.F. and Nizami, J.S., 1996, "A Genetic Algorithm for The Set Covering Problem", Journal of The Operational Research Society, 47, 702-709

[17] Levine, D., 1996, "A Parallel Genetic Algorithm For The Set Partitioning Problem" In: I.H. Osman and J.P. Kelly, Editors, Meta-heuristics: theory and applications. Kluwer Academic Publishers, Dordrecht, 23–35

[18] Chu, P.C. and Beasley, J.E., 1998, "Constraint Handling In Genetic Algorithms: The Set Partitioning Problem", Journal of Heuristics, 4, 323–357

[19] Beasley, J. E., Chu, P. C., 1996, "A Genetic Algorithm For The Set Covering Problem", European Journal of Operational Research, 94 (2), 392-404

[20] Kelly, J. and Xu, J., 1999, "A Set Partitioning Based Heuristic For The Vehicle Routing Problem", INFORMS Journal on Computing, 11, 161–172

[21] Mark, L., Kochenberger, G., and Alidae, B., 2008, "A New Modeling And Solution Approach For The Set-Partitioning Problem", Computers & Operations Research, 35 (3), 807-813

[22] Güngör, İ. and Küçüksille, E.U., 2005, "Küme Bölme Problemlerinin Genetik Algoritmayla Optimizasyonu: Türkiye İkinci Futbol Ligi Uygulması", Review of Social Economics & Business Studies, 5 (6), 381-394

[23] Goldberg, D.E., 1989, "Genetic Algorithms in Search Optimization and Machine Learning", Addison Wesley Publishing Company, USA.

[24] Reeves, C.R., 1995, "Modern Heuristic Techniques for Combinatorial Problems", Mcgraw-Hill Book Company Inc., Europe.
[25] Braysy, O., Gendreau, M., 2005, "Vehicle Routing Problem with Time Windows, Part II: Metaheuristics", Transportation Science, 39 (1), 119-139
[26] Keskintürk, T., 2007, "Portföy Seçiminde Markowitz Modeli Için Yeni Bir Genetik Algoritma Yaklaşımı", Yönetim, 18 (56), 81.
[27] Akyurt, I.Z. Keskintürk, T. and Arıkan, M.V., 2009, "Determining The Parameters Of Periodic Optional Replenishment Model With Genetic Algorithm", International Logistics And Supply Chain Congress', Istanbul, Turkey.