A Multi-objective Evolutionary Approach to Pareto Optimal Model Trees. A Preliminary Study

Marcin Czajkowski(✉) and Marek Kretowski

Faculty of Computer Science, Bialystok University of Technology,
Wiejska 45a, 15-351 Bialystok, Poland
{m.czajkowski,m.kretowski}@pb.edu.pl

Abstract. Decision tree induction is inherently a multi-objective task. However, most of the conventional learning algorithms can only deal with a single-objective that may possibly aggregate multiple objectives. This paper proposes the multi-objective evolutionary approach to Pareto optimal model trees. We developed a set of non-dominated model trees for a Global Model Tree framework using efficient sort and specialized selection. Performed study covers variants with two and three objectives that relate to the tree error and the tree comprehensibility. Pareto front generated by the GMT system allows the decision maker to select desired output model according to his preferences on the conflicting objectives. Experimental evaluation of the proposed approach is performed on three real-life datasets and is confronted with competitive model tree inducers.

Keywords: Data mining · Evolutionary algorithms · Model trees · Multi-objective optimization · Pareto optimality

1 Introduction

The most important role of data mining [10] is to reveal important and insightful information hidden in the data. Among various tools and algorithms that are able to effectively identify patterns within the data, the decision trees (DT)s [20] represent one of the most frequently applied prediction technique. Tree-based approaches are easy to understand, visualize, and interpret. Their similarity to the human reasoning process through the hierarchical tree structure, in which appropriate tests from consecutive nodes are sequentially applied, makes them a powerful tool [29] for data analysts.

Despite 50 years of research on DTs, there is still a space for the improvement [21], such as: the search for better structure, splits and leaves models; multi-objective optimization or efficient analysis of the cost-sensitive data. To help to resolve some of these issues, evolutionary algorithms (EA)s [23] are applied to DTs induction [2]. The strength of this approach lies in the global search for splits and predictions and results in simpler but still accurate trees in comparison to ones induced with greedy strategies [5].

C. Martín-Vide et al. (Eds.): TPNC 2016, LNCS 10071, pp. 85–96, 2016.
DOI: 10.1007/978-3-319-49001-4_7

The objective of this paper is to allow the decision maker to select desired output model according to his preferences on tree comprehensibility and accuracy. The main contribution is a multi-objective evolutionary approach to Pareto optimal model trees. To the best of our knowledge, such study on multi-objective optimization for regression or model trees, surprisingly, has not yet been addressed in the literature. Despite the popularity of DTs, the topic has not yet been adequately explored even for classification trees.

In this work, we focus on the Global Model Tree (GMT) framework [5] that can be used for the evolutionary induction of different kinds of regression and model trees [6] and be applied in real-life applications [4]. We have extended the actual fitness function of the GMT system that applied weight-formula or lexicographic analysis with Pareto-based multi-objective optimization methodology. The efficient non-dominated sort (ENS) [31], archive list of non-dominated solutions as well as updated crowding functions were applied for the GMT system. We have also incorporated the knowledge about tree induction into the evolutionary search.

Experimental study was performed on three publicly available real-life datasets and covered two-objective: tree error and tree comprehensibility; and three-objective optimization: tree error, number of nodes and the number of attributes in regression models located in the leaves. We have also confronted obtained results with the competitive model tree inducers.

This paper is organized as follows. Section 2 provides a brief background and Sect. 3 describes in detail proposed Pareto optimal search for the GMT framework. Section 4 presents experimental validation of our approach on real-life datasets. In the last section, the paper is concluded and possible future works are outlined.

2 Background

In this section, we want to present some background information on DTs and the multi-objective optimization.

2.1 Decision Trees

Different variants of DTs [21] may be grouped according to the type of problem they are applied to, the way they are induced, or the type of their structure. In this paper, we focus on a model tree that can be seen as an extension of the typical regression tree [27] which, in turn, is considered as a variant of DT designed to approximate real-valued functions instead of being used for classification tasks. Although, regression and model trees are not as popular as classification trees, they are highly competitive with different machine learning algorithms [13].

In case of the simplest regression tree [3], each leaf is associated with a constant value, usually an average value of the target attribute. In the model tree, this value is replaced by a linear (or nonlinear) regression function. To predict the target value, a new tested instance is followed down the tree from a root node to a leaf using its attribute values to make routing decisions at each internal node. Next, the predicted value for the tested instance is evaluated based on a regression model in the leaf.

In this paper we study the evolutionary induced model trees; therefore, to go further, we must briefly describe the process of learning of DT based on the training set. The two most popular concepts for the DT induction are the top-down and global approaches. The first is based on a greedy procedure known as recursive partitioning [28]. In the top-down approach, the induction algorithm starts from the root node where the locally optimal split is searched according to the given optimality measure. Next, the training instances are redirected to the newly created nodes, and this process is repeated for each node until a stopping condition is met. Additionally, post-pruning [8] is usually applied after the induction to avoid the problem of over-fitting the training data. Inducing trees with the greedy strategy is fast and generally efficient but often produces only locally optimal solutions. One of the most popular representatives of top-down induced regression trees is Classification And Regression Tree (CART) [3]. As for the model trees, the M5 system [27] is the most recognized algorithm that induces a tree with multiple linear regression models in the leaves.

The global induction for the DTs limits the negative effects of locally optimal decisions. It simultaneously searches for the tree structure, tests in the internal nodes, and models in the leaves. This approach is mainly represented by systems based on an evolutionary search [2,5] and may reveal hidden regularities that are often undetectable by greedy methods. There are relatively fewer approaches for the evolutionary induction of regression and model trees than for the classification trees. Popular representatives of EA-based regression trees are: the TARGET system [9] that evolves a CART-like tree with basic genetic operators and a strongly typed GP (Genetic Programming) approach called STGP [14]. In case of globally induced model trees there is E-Motion [1] that is a counterpart of the M5 system and a GP approach called GPMCC [26] with nonlinear regression models in the leaves.

2.2 Multi-objective Optimization in the Decision Trees

Real-world optimization problems are usually characterized by multiple objectives which often conflict with each other. In case of the DT induction, it is advisable to maximize the predictive performance and the complexity of the output tree. A single evaluation measure may degrade the other measures; therefore, multi-objective optimization may present more acceptable overall results. In the context of DTs, a direct minimization of the prediction accuracy measured in the learning set usually leads to the over-fitting problem [8].

There are three popular multi-objective optimization strategies [2]: the weight formula, lexicographic analysis, and Pareto-dominance. The weight formula transforms a multi-objective problem into a single-objective one by constructing a single formula that contains all objectives. The main drawback of this strategy is the need to find adjusted weights for the measures. The lexicographic approach analyzes the objective values for the individuals one by one based on the priorities. This approach also requires defining thresholds; however, adding up non-commensurable measures, such as tree error and size, is not performed. In contrast to Pareto-dominance approach, both aforementioned solutions were already applied for evolutionary regression and model tree induction [1,5].

Pareto-dominance [25] searches not for one best solution, but rather for a group of solutions in such a way, that selecting any one of them in place of another will always sacrifice quality for at least one objective, while improving it for at least one other. Consider m conflicting objectives that need to be minimized simultaneously. A solution $\mathbf{A} = \{a_1, a_2, \ldots, a_m\}$ is said to dominate solution $\mathbf{B} = \{b_1, b_2, \ldots, b_m\}$ (symbolically denoted by $\mathbf{A} \prec \mathbf{B}$) if and only if:

$$(\mathbf{A} \prec \mathbf{B}) \Leftrightarrow (\forall i)(a_i \leq b_i) \wedge (\exists i)(a_i < b_i). \tag{1}$$

The Pareto optimal set is constituted only with solutions that are not dominated by any other solutions:

$$\{\mathbf{A} | \neg(\exists \mathbf{B}, \mathbf{B} \prec \mathbf{A})\}. \tag{2}$$

The set of all Pareto optimal solutions is referred to as the Pareto front. Next, these multiple alternative solutions are be presented to the decision maker for a consideration.

Although, Pareto optimal approach is popular in machine learning [17], it has not been explored for regression or model trees yet. However, in the literature we may find some attempts performed for classification trees. In [32] the author created Pareto optimal DTs to capture the trade-off between different types of misclassification errors in a cost-sensitive classification problem. Such multi-objective strategy was also applied [19] to top-down induced trees to minimize two objectives: classification error rate and the tree size (measured by the number of the tree nodes). Finally, the Pareto-optimality for greedy induced oblique DTs was investigated in [24]. The authors show that a inducer, that generates the most accurate trees does not necessarilly generate the smallest trees or the ones that are included in Pareto-optimal set.

3 Pareto-Optimal Search in GMT

In this section, we present a new multi-objective approach for evolutionary induced regression and model trees. At first, we briefly describe a system called Global Model Tree (GMT). Next, we illustrate how to efficiently adapt Pareto-based approach in the fitness function for GMT.

3.1 Global Model Tree

The general structure of the GMT system follows a typical EA framework [23] with an unstructured population and a generational selection. The GMT framework allows evolving all kinds of tree representations [6] e.g.: univariate, oblique, regression, model and mixed. In our description we focused on univariate model trees [5], however; our study can be easily adapted to the different types of trees.

Model trees are represented in their actual form as traditional univariate trees, so each split in the internal node is based on a single attribute. Each tree leaf contains a multiple linear regression model that is constructed with learning instances associated with that leaf. Tree-based representation requires developing specialized genetic operators corresponding to classical mutation and crossover. The GMT framework [5] offers several specialized variants that can modify the tree structure, tests in internal nodes, and models in the leaves.

Fitness function is one of the most important and sensitive element in the design of EA. It drives the evolutionary search process by measuring how good a single individual is in terms of meeting the problem objective. Currently there are two multi-objective optimization strategies implemented in the GMT: weight formula and lexicographic analysis. Among various weight formulas tested within the GMT system, the Bayesian information criterion (*BIC*) [30] has the highest performance with regression and model trees. The BMC is given by:

$$Fit_{BIC}(T) = -2 * ln(L(T)) + ln(n) * k(T), \tag{3}$$

where $L(T)$ is the maximum of the likelihood function of the tree T, n is the number of observations in the data, and $k(T)$ is the number of model parameters in the tree. The log(likelihood) function $L(T)$ is typical for regression models and can be expressed as:

$$ln(L(T)) = -0.5n * [ln(2\pi) + ln(SS_e(T)/n) + 1], \tag{4}$$

where $SS_e(T)$ is the sum of squared residuals of the tree T. In this measure of goodness of fit the term $k(T)$ can be viewed as a penalty for over-parametrization. It reflects the tree complexity which for regression trees equals to the number of nodes (denoted as $Q(T)$), where as for model trees it also includes the number of attributes in the linear models in the leaves (denoted as $W(T)$).

When the lexicographic analysis is applied in fitness evaluation, each pair of individuals is analyzed, in order of priorities, one of the measures: $SS_e(T)$, $Q(T)$ and $W(T)$. The first priority is set for the tree accuracy measure, next the number of terminal nodes to prevent overfitting and overgrown trees. The last measure $W(T)$ keeps the models in the leaves as simple as possible and also penalizes for over-parametrization.

The selection mechanism is based on the ranking linear selection [23] with the *elitist strategy*, which copies the best individual founded so far to the next population. Evolution terminates when the fitness of the best individual in the population does not improve during the fixed number of generations (default: 1000). In case of a slow convergence, maximum number of generations is also specified (default value: 10000), which limits the computation time.

3.2 Pareto-Based Approach for GMT

The main goal of the multi-objective optimization is to find a diverse set of Pareto-optimal solutions, which may provide insights into the trade-offs between

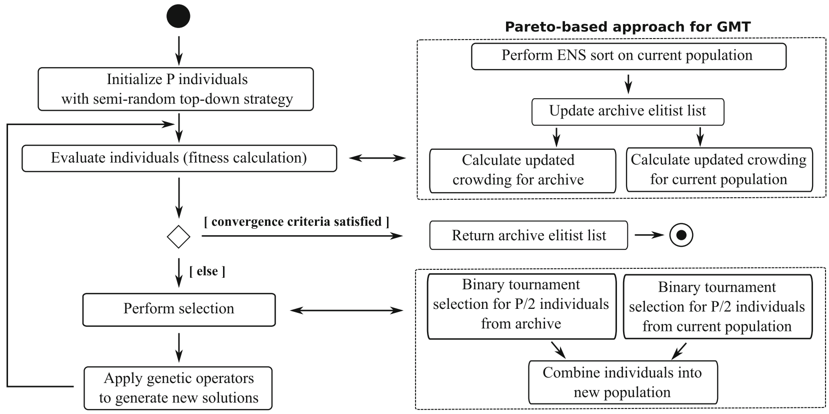

Fig. 1. General GMT schema together with proposed Pareto-based extension.

the objectives. Current GMT fitness functions: weight formula and lexicographic analysis yield only a limited subset of the solutions that may not even belong to the Pareto front.

Multiple EAs were developed to tackle the multi-objective optimization problems, in particular the search for a set of Pareto-optimal solutions [15]. Among various dominance comparison mechanisms, non-dominated sorting genetic algorithm NSGA-II [7] has been shown to be very effective. It showed fast convergence to the Pareto-optimal set, good spread of solutions and became a framework for many future algorithms.

In the GMT system, we have applied the basics of the NSGA-II workflow. Most of the elements like sorting strategy itself, crowding and elitism differ from the NSGA-II algorithm as they were specialized to fit more accurately to the problem of evolutionary model tree induction. Figure 1 shows the general GMT schema together with proposed Pareto-based extension.

In the first step, more recent search strategy called efficient non-dominated sorting strategy (ENS) [31] is applied. The reason why ENS was selected is due to its efficiency. Experimental evaluation of ENS showed that it outperforms other popular non-dominated sorting approaches especially for optimization problems having a small number of objectives which is here the case. The ENS algorithm is conceptually different from most existing non-dominated sorting methods. ENS determines the front each solution belongs to one by one, where typical non-dominated sorting approaches determine the front of all solutions on the same front as a whole. This way ENS avoids duplicate comparisons, since a solution to be assigned only needs to be compared with solutions that have already been assigned to the front.

In the second step of proposed extension (see Fig. 1) the archive fronts are updated. The NGSA-II approach maintains a population size set of non-dominated solutions that is later combined with the next population. However, in case of GMT, where population size is small (50 individuals), many interesting from the decision-maker point of view non-dominated solutions may be lost. Therefore, we have applied different strategy [33] that allows storing all non-dominated solutions investigated so far during the search. Solutions from Pareto front are stored in elitist list, which is updated each time a new solution from the current population dominates one in the list. Although, this operation is more computationally expensive, it is still acceptable as the Pareto front in case of GMT is not very large.

In proposed extension we also have adapted the updated crowding distance procedure [11]. The crowded comparison operator ($\prec_n$) helps ordering (ranking) the solutions. In NSGA-II the crowding-distance is used for diversity preservation and to maintain a well-spread Pareto front. The main improvement of the crowding distance calculation focuses on using unique fitnesses when two or more individuals share identical value. Such case in NSGA-II algorithm causes the crowding distance of the individual to either become null, or to depend on the individuals position within the Pareto front sequence.

Finally, proposed approach differs from NSGA-II in a way of creating a new population. In NSGA-II the archive and current population are merged into new population using the binary tournament as a selection method. Each solution is assigned a rank equals to its non-domination level (1 is the best level, 2 is the next-best level, and so on) and in case of a draw, crowding distance is considered. Due to storing the full list of non-dominated solutions in the archive, we have applied strategy proposed in [16]. We reserve a room for p elitist solutions in the next population (default: half of the population size P). In this strategy, $P - p$ solutions are selected from parents and newly created offspring and p solutions are selected from the stored elitist list. Both sets use the binary tournament as a selection method. The elitist solutions are scored with the crowding distance (as they all belong to a non-dominated set) and the current population is scored like in NSGA-II algorithm.

4 Experimental Validation

In this section, we perform a preliminary validation of proposed Pareto optimal extension. While evolving model trees one can distinguish three objectives that could to be minimized: prediction error measured with Root Mean Squared Error (RMSE), number of nodes and the number of attributes in regression models located in the leaves. The last two objectives are partially depended and may fall under one single-objective denoted as a tree comprehensibility. Thus, in performed experiments we present the results for the fitness function with 2 objectives where number of nodes and the attributes in models is summed; and with 3 objectives where all measures are analyzed separately.

To assess the performance of the proposed approach in solving real-life problems three real-life publicly available datasets from Louis Torgo repository [22] were analyzed: Abalone (4177 instances, 7 real-valued, 1 nominal attribute), Kinematics (8192, 8, 0) and Stock (950, 9, 0). Each dataset was divided into the training (66.6 %) and testing (33.4 %) set. In our experimental validation of

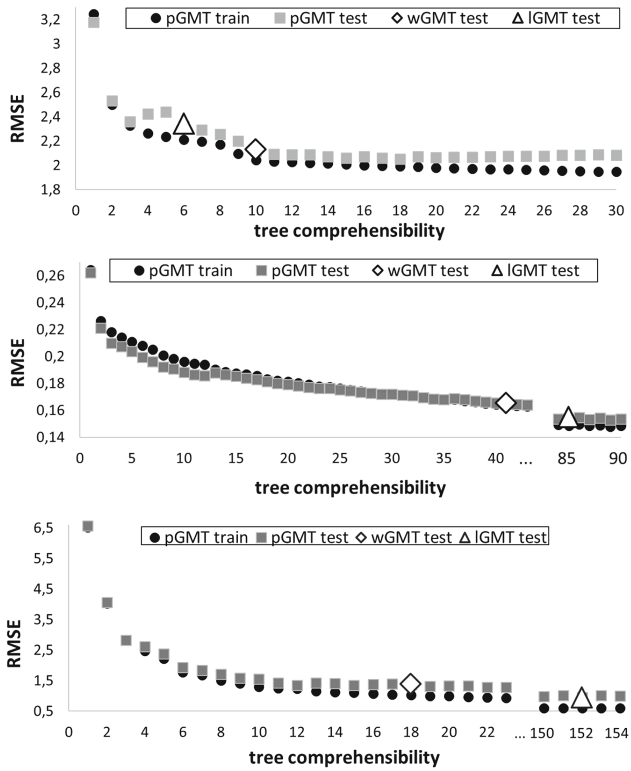

Fig. 2. Pareto front for GMT (pGMT) for 2 objectives on training and testing set of Abalone (a), Kinematics (b) and Stock (c). Results on testing set for GMT with weight (wGMT) and lexicographic (lGMT) fitness functions are also enclosed.

the proposed Pareto GMT (denoted in experiments as pGMT) we also enclosed the results for competitive systems. We have tested GMT with the weight fitness function (wGMT), GMT with the lexicographic fitness function (lGMT) and two popular top-down inducers: REP Tree (RT) that builds a regression tree and state-of-the-art model tree called M5 [27] which is the most adequate greedy counterpart of GMT.

Figure 2 shows the results achieved for the GMT system with different fitness functions. The Pareto front was achieved for bi-objective optimization problem that minimized the RMSE and the tree comprehensibility. One can observe, that for all tested dataset the GMT system with weight or lexicographic fitness functions managed to find non-dominated solutions, as they belong to the Pareto front. However, open question is if the induced trees by wGMT or lGMT will satisfy the decision maker. In case of the results for Abalone dataset (Fig. 2a) both wGMT and lGMT managed to find simple model trees with decent prediction performance. However, if the analyst wants to have slightly more accurate model, he might select trees with higher number of nodes/attributes. Opposite situation is for the Kinematics (wGMT and lGMT) and Stock (lGMT) datasets where they find accurate but complex prediction models which could be difficult to analyze and interpret. Although, the trade-off between prediction performance and tree comprehensibility can be partially managed by ad-hoc settings of the complexity term (wGMT) and thresholds (lGMT), there is no guarantee that founded solutions will belong to the Pareto front. With the proposed Pareto-based approach for GMT the decision maker can easily balance between the tree prediction performance and its comprehensibility, depending on the purpose of the analysis goals.

The Pareto front for three-objective optimization problem is illustrated in Fig. 3. For all datasets, one can see a trend that either induced trees are small but with large number of attributes, either large but with smaller number of the attributes. In all cases, more compact trees have higher prediction performance (smaller RMSE) than larger ones but with simpler models in the leaves. We can also observe, that the lGMT for Kinematics and Stock dataset finds solutions that do not belong to the Pareto front. The three-objective optimization enables obtaining more variants of the output trees, but it can also cause choosing right prediction more difficult.

Table 1 illustrates achieved results for evolutionary induced model trees with different fitness functions as well as popular greedy counterparts of GMT. For all datasets, three metrics are shown: RMSE on the testing set, number of nodes and number of attributes in regression models located in the leaves. One can observe that GMT variants induce much more comprehensible predictors with smaller number of nodes and less complex regression models in the leaves, which was also noticed in [5]. The RT algorithm induces only regression trees, thus the attribute metric does not apply.

In the real-life applications the predictive accuracy is usually considered more important than the comprehensibility of DT. Consider, for instance, two trees: T1 and T2 where T1 has 20 % smaller prediction error but also 20 % larger size. Most of the researches would clearly prefer the T1 over T2; however, the

Table 1. Performance results for evolutionary induced model trees with different fitness functions as well as popular greedy counterparts of GMT. Results for three solutions from the Pareto front (denoted as pGMT *) are also included.

Algorithm	Abalone			Kinematics			Stock		
	RMSE	nodes	attributes	RMSE	nodes	attributes	RMSE	nodes	attributes
wGMT	2.127	2.0	7.92	0.163	7.12	34.32	1.386	3.94	14.17
lGMT	2,341	1.0	5	0.154	9.7	74.5	0.935	41.4	111
M5	2.122	12.0	96	0.162	106	848	0.937	47	423
RT	2.223	291	-	0.194	819	-	1.469	137	-
pGMT 1*	2.359	1	2	0.184	2	13	1.531	3	8
pGMT 2*	2.102	2	9	0.174	4	25	0.928	31	38
pGMT 3*	2.079	3	12	0.149	17	116	0.782	61	121

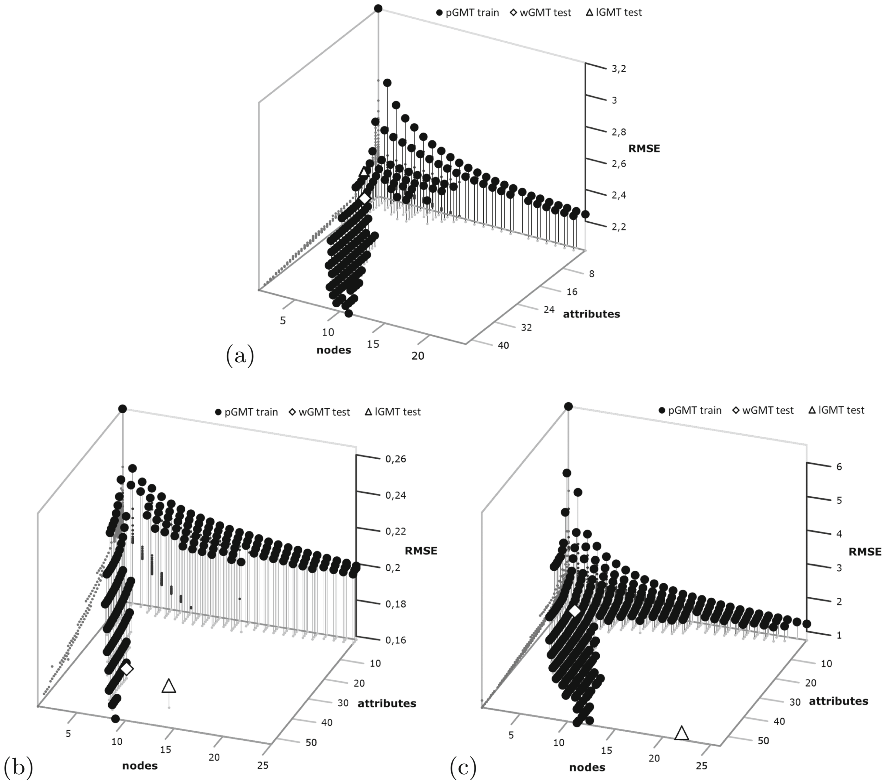

Fig. 3. Pareto front for GMT (pGMT) for 3 objectives on testing set of Abalone (a), Kinematics (b) and Stock (c). Results on testing set for GMT with weight (wGMT) and lexicographic (lGMT) fitness functions are also enclosed.

Pareto approach would consider that none of these two trees dominates the other. Therefore, in context of DT and the Pareto front, the weights preferences could be introduced to the multi-objective optimization [12].

5 Conclusion and Future Works

In the paper we propose a new fitness function to evolutionary induced regression and model trees. Preliminary experiments showed that our approach is capable of finding Pareto front for the GMT framework. Its a first step towards searching for efficient and easy to interpret Pareto optimal DTs. There are, however; still many open issues that need more research and clarification.

The impact of this new multi-objective optimization on the GMT performance need to be analyzed more deeply. The proposed approach increases the calculation time of each evolutionary loop and may affect the convergence of the EA. Additional efficiency improvements, especially in context of storing and preprocessing full list of non-dominated solutions need to be considered. Performance issue may also be partially mitigated with parallelization of the GPGPU approach for GMT [18].

Next issue that need to be resolved is the comprehensibility of the generated Pareto front. Currently, due to the size of the front and the number of possible trees, the decision maker may have problem to decide which predictor he should choose. Thus, more research need to be performed in context of limiting the output elitist front as well as improving the crowding function with aforementioned weights.

Acknowledgments. This project was funded by the Polish National Science Center and allocated on the basis of decision 2013/09/N/ST6/04083. The second author was supported by the grant S/WI/2/13 from Bialystok University of Technology founded by Ministry of Science and Higher Education.

References

1. Barros, R.C., Ruiz, D.D., Basgalupp, M.P.: Evolutionary model trees for handling continuous classes in machine learning. Inf. Sci. **181**(5), 954–971 (2011)
2. Barros, R.C., Basgalupp, M.P., de Carvalho, A.C.P.L.F., Freitas, A.A.: A survey of evolutionary algorithms for decision-tree induction. IEEE Trans. Syst. Man Cybern. Part C **42**(3), 291–312 (2012)
3. Breiman, L., Friedman, J., Olshen, R., Stone, C.: Classification and Regression Trees. Wadsworth and Brooks, Monterey (1984)
4. Czajkowski, M., Czerwonka, M., Kretowski, M.: Cost-sensitive global model trees applied to loan charge-off forecasting. Decis. Support Syst. **74**, 57–66 (2015)
5. Czajkowski, M., Kretowski, M.: Evolutionary induction of global model trees with specialized operators and memetic extensions. Inf. Sci. **288**, 153–173 (2014)
6. Czajkowski, M., Kretowski, M.: The role of decision tree representation in regression problems - an evolutionary perspective. Appl. Soft Comput. **48**, 458–475 (2016)
7. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comp. **6**(2), 182–197 (2002)
8. Esposito, F., Malerba, D., Semeraro, G.: A comparative analysis of methods for pruning decision trees. IEEE Trans. Pattern Anal. Mach. Intell. **19**(5), 476–491 (1997)
9. Fan, G., Gray, B.J.: Regression tree analysis using TARGET. J. Comput. Graph. Stat. **14**(1), 206–218 (2005)
10. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.): Advances in Knowledge Discovery and Data Mining. American Association for Artificial Intelligence, Menlo Park (1996)
11. Fortin, F.A., Parizeau, M.: Revisiting the nsga-ii crowding-distance computation. In: Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation, pp. 623–630. GECCO 2013. ACM, New York, NY, USA (2013)
12. Friedrich, T., Kroeger, T., Neumann, F.: Weighted preferences in evolutionary multi-objective optimization. Int. J. Mach. Learn. Cybern. **4**(2), 139–148 (2013)
13. Guzman, F.M.O., Valenzuela, O., Prieto, B., Saéz-Lara, M.J., Torres, C., Pomares, H., et al.: Comparing different machine learning and mathematical regression models to evaluate multiple sequence alignments. Neurocomputing **164**, 123–136 (2015)

14. Hazan, A., Ramirez, R., Maestre, E., Perez, A., Pertusa, A.: Modelling expressive performance: a regression tree approach based on strongly typed genetic programming. In: Rothlauf, F., et al. (eds.) EvoWorkshops 2006. LNCS, vol. 3907, pp. 676–687. Springer, Heidelberg (2006). doi:10.1007/11732242_64
15. Hiwa, S., Nishioka, M., Hiroyasu, T., Miki, M.: Novel search scheme for multi-objective evolutionary algorithms to obtain well-approximated and widely spread pareto solutions. Swarm Evol. Comput. **22**, 30–46 (2015). (Complete)
16. Ishibuchi, H., Murata, T.: A multi-objective genetic local search algorithm and its application to flowshop scheduling. IEEE Trans. Syst. Man Cybern. Part C **28**(3), 392–403 (1998). (Applications and Reviews)
17. Jin, Y., Sendhoff, B.: Pareto-based multiobjective machine learning: An overview and case studies. IEEE Trans. Syst. Man Cybern. Part C **38**(3), 397–415 (2008)
18. Jurczuk, K., Czajkowski, M., Kretowski, M.: Evolutionary induction of a decision tree for large scale data. a GOU-based approach. Soft Comput. doi:10.1007/s00500-016-2280-1 (in press, 2016)
19. Kim, D.E.: Structural risk minimization on decision trees using an evolutionary multiobjective optimization. In: Keijzer, M., O'Reilly, U.-M., Lucas, S., Costa, E., Soule, T. (eds.) EuroGP 2004. LNCS, vol. 3003, pp. 338–348. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24650-3_32
20. Kotsiantis, S.B.: Decision trees: a recent overview. Artif. Intell. Rev. **39**(4), 261–283 (2013)
21. Loh, W.Y.: Fifty years of classification and regression trees. Int. Stat. Rev. **82**(3), 329–348 (2014)
22. Louis, T.: Regression data sets (2016). http://www.dcc.fc.up.pt/ltorgo/Regression/DataSets.html
23. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs, 2nd edn. Springer-Verlag, New York (1994)
24. Pangilinan, J., Janssens, G.: Pareto-optimality of oblique decision trees from evolutionary algorithms. J. Global Optim. **51**(2), 301–311 (2011)
25. Pappalardo, M.: Multiobjective optimization: a brief overview. In: Chinchuluun, A., Pardalos, P.M., Migdalas, A., Pitsoulis, L. (eds.) Pareto Optimality, Game Theory And Equilibria. Springer Optimization and Its Applications, vol. 17, pp. 517–528. Springer, Heidelberg (2008)
26. Potgieter, G., Engelbrecht, A.P.: Evolving model trees for mining data sets with continuous-valued classes. Expert Syst. Appl. **35**(4), 1513–1532 (2008)
27. Quinlan, J.R.: Learning with continuous classes. Proc. Aust. Joint Conf. Artif. Intell. **92**, 343–348 (1992). World Scientific, Singapore
28. Rokach, L., Maimon, O.: Top-down induction of decision trees classifiers - a survey. Trans. Sys. Man Cyber Part C **35**(4), 476–487 (2005)
29. Rokach, L., Maimon, O.: Data Mining with Decision Trees: Theroy and Applications. World Scientific Publishing Co. Inc., River Edge (2008)
30. Schwarz, G.: Estimating the dimension of a model. Ann. Statist. **6**(2), 461–464 (1978)
31. Zhang, X., Tian, Y., Cheng, R., Jin, Y.: An efficient approach to nondominated sorting for evolutionary multiobjective optimization. IEEE Trans. Evol. Comput. **19**(2), 201–213 (2015)
32. Zhao, H.: A multi-objective genetic programming approach to developing pareto optimal decision trees. Decis. Support Syst. **43**(3), 809–826 (2007)
33. Zitzler, E., Thiele, L.: Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. IEEE Trans. Evol. Comp. **3**(4), 257–271 (1999)