**Speeding Up Many-Objective Optimization of Hydro-Generator Parameters**

**Abstract.** Multi-objective evolutionary algorithms (MOEAs) are broadly accepted methods in multi-criteria decision-making. Numerous advanced genetic optimization algorithms compute the complicated sets of Pareto-optimal solution. However, with hard-to-calculate objectives, the computation becomes intractable. This paper presents the approach to accelerate the solution of a multi-objective optimization problem, by the example of hydro-generator parameters. Using clustering and regression analysis allow us to reduce the sample size and optimization time, which is especially relevant for problems where each individual requires a long computation time. This paper will describe the underlying genetic algorithm, clustering analysis, regression analysis, and some initial results.

1. **Introduction**

Nowadays optimization algorithms are commonly used in almost all industry areas and life in general. The application of optimization methods is especially relevant in those areas where a significant number of repetitive calculations. These areas include electrical engineering and, in particular, electrical machines.

The choice of candidate solutions is intractable by the complexity of the electric machine calculation algorithm and the requirement to take into account mutually exclusive parameters. Often the optimal electrical machine options are chosen based on the designer skills and field experience. This problem belongs to multi-objective optimization (MOO). A solution is Pareto frontier, which is a state of a system when each value of objective functions could not be improved without making others worse. Objectives are usually contradictory, and optimal solutions are compromises between criteria and are called Pareto solutions. The general task of multi-objective optimization is formulated as follows:

parameters vector should be found

with boundary conditions

and limitation functions

for objective function vector minimizing

.

In the optimization of electrical machine design, geometric dimensions, winding current density, material properties, etc. can be variables, while objective functions can be e.g. machine weight, steel and copper losses, subtransient resistance, etc. International technical standards, electromagnetic, thermal, mechanical and technological conditions are used to limit parameters. The general concept is to find a set of vectors, which will be Pareto frontier for entire ranges of variables while keeping the specifications within acceptable limits.

In order to solve this problem it is essential obtain Pareto frontier. Objective function values must be calculated for all solutions. Finite element method is used to obtain solutions, thereby making the weight problem, since significant computational resources is required. Determining Pareto set requires multiple calculations defined by the vectors . Thus, once the solution is complete, there is a vast set of electrical machine designs that can be used for regression analysis.

Despite the constant growth of computational resources and the appearance of new optimization algorithms, MOO still requires the participation of the decision maker to accelerate the solution process. We describe the underlying genetic algorithm, clustering analysis, regression analysis, and some initial results.

This paper uses the results of MMO of hydro-generator parameters [1-3] by the non-dominant sorting genetic algorithm II (NSGA-II) [4]. There are four objectives: stator core mass, stator steel losses, rotor current, and short circuit ratio. In this paper, the vector of design parameters includes thirteen geometric dimensions: outer and inner diameters of stator core ( and ); stator active steel length ; stator slot height ; air gap size between rotor and stator ; width, height, and radius of curvature of the rotor pole tip ( , , and ); width of the rotor pole core ; diameter, radius of location, and width of slot of damper rods (, and

1. **Methodology**

The flowchart describing actions of the decision maker is presented in Figure 1. In addition to the underlying genetic algorithm, the basic steps include:

* Сlustering to find a new set of categories. Clustering groups objects into subsets in such a way that identical objects are combined together and the rest belong to other groups.
* Correlation analysis quantifies the relationship of a linear relationship between two variables. This step is optional because some regression models allow for a linear relationship between the two variables.
* Features selection and regression analysis allow us to obtain compact models of the dependent variables.
* Testing the regression model for adequacy.

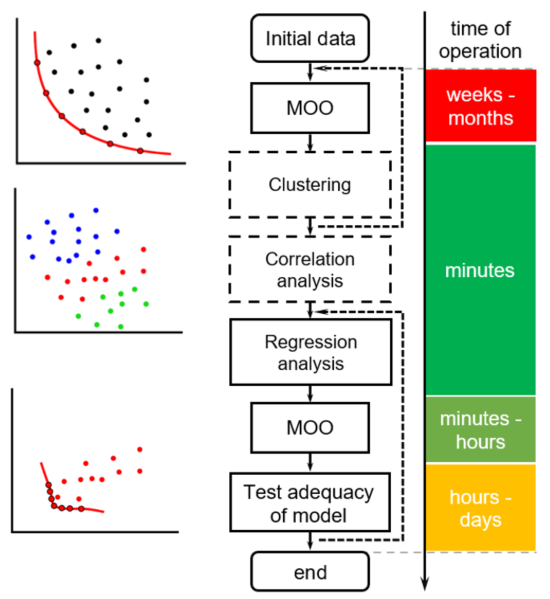


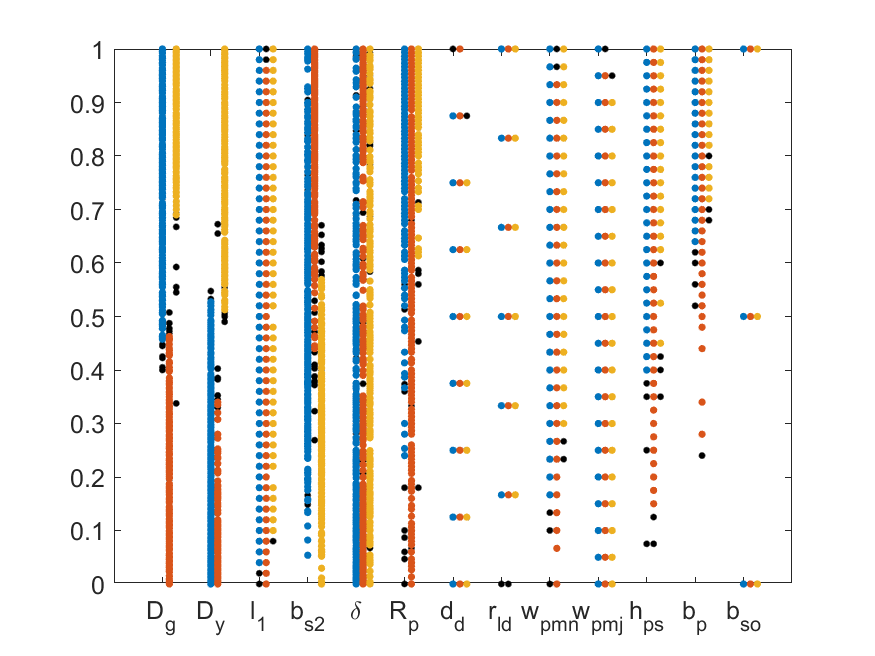
Figure 1 — Schematic diagram of methodology with operation time

* 1. **Clustering and correlation analysis**

The large variety of clustering methods make it difficult to pick out a single one, but the key point is to decrease the number of samples to minimize the calculation time. In the authors' experience, the choice of the most common k-means or k-medoids with variance ratio criterion is sufficient to define cluster groups and select the most appropriate among them [5].

The results of numerical calculation of hydro-generator used at the customer’s facility and designed according to the method [6], were chosen as base values {1; 1; 1}. The base generator is marked in black. Figure 2 shows the clustering of the parameter population obtained by NSGA-II. The figure shows all machines analyzed in the construction of Pareto frontier. Calculation of this set required 330 hours on the SPbPU cluster (14-core Intel Xeon E5 2695 v3 processors and 64 GB) using ANSYS Electronics Desktop to solve problems. Each cluster is highlighted in color. The NSGA-II method alter the vector according to the algorithm and the constraints of the problem. As can be seen, parameters number were changed discretely and continuously (, are continuous, all others are discrete.)

Clustering makes it easy to trace that cluster #1 and #3 belong to the so-called extreme variants in the population, excluding them from further analysis.



Δ, p.u.

parameter

— cluster #1

— cluster #2

— cluster #3



Figure 2 — Parameters vector distribution by clusters

(black dots are remoted outliers)

While clustering reduced the size of the population correlation analysis decreases the number of parameters. Initially, it is recommended to select the parameters so that they do not have a linear relationship. If this condition cannot be met, then correlation analysis and variance inflation factor allow us to exclude an extra parameter from consideration [7].

* 1. **Optimal feature selection and regression analysis**

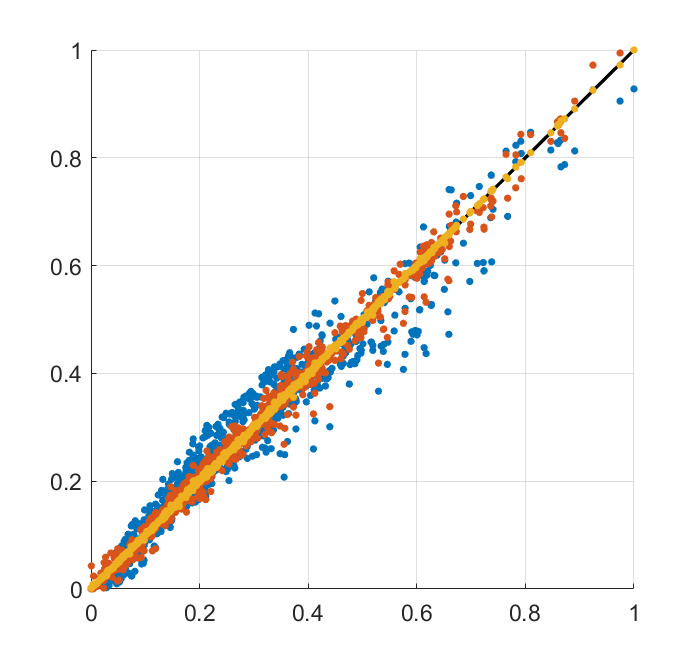
Optimal feature selection is a kind of logical continuation of correlation analysis and diagnosis of model collinearity. One of the common challenges of data analysis is optimal predictors selection for the model. Reducing number of independent variables is intended to reduce the model dimensionality by removing all unimportant and redundant predictors, thus simplifying the model. In addition, redundant predictors add noise to the estimation of the influence of other factors of interest.

There is no universal and well-established approach to select the optimal set of predictors. Instead of the traditional linear and polynomial regressions, let us consider a method based on committee of decision trees (random forest — RF) [8].

The algorithm essence can be summarized by making a random sample of variables at each iteration, and then running a decision tree on this new sample [9]. In this case, a random sample of two-thirds of the observations is selected for training, and the remaining third is used to evaluate the result. This operation is performed a specified number of times.

A comparative analysis was performed for three regression models: linear, polynomial and RF. Automatic selection of informative predictors, taking into account possible interactions between them, the absence of the need to define relationship between an objective and predictors, and the high coefficient of determination show the efficiency and ease of use of the RF. The analysis of Table 1 and Figure 3 shows that the RF is the best regression model of the three presented and the most robust to reduce the independent number of variables. The determinism of the results and the absence of a significant contribution from part of the predictors indicates the strict dependence of the predictors and factors on the set of Pareto-optimal solutions. In checking the adequacy of the regression model, a standard deviation and a coefficient of determination were used.

Predicted data, p.u.



Initial data, p.u.

— LR

— PR

— Random forest

Figure 3 — Relative importance of random forest predictors

Table 1 — Standard deviation and coefficient of determination of different regressions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Set of predictors | LR | | PR | | RF | |
|  |  |  |  |  |  |
| Mass of stator core | | | | | | |
|  | 0.146 | 0.979 | 0.060 | 0.996 | 0.018 | 0.999 |
|  | 0.137 | 0.981 | 0.040 | 0.998 | 0.003 | 1.000 |
|  | 0.131 | 0.983 | 0.039 | 0.999 | 0.003 | 1.000 |
| whole set | 0.126 | 0.984 | 0.030 | 0.999 | 0.001 | 1.000 |
| Iron losses | | | | | | |
|  | 0.235 | 0.945 | 0.067 | 0.996 | 0.014 | 0.999 |
|  | 0.233 | 0.946 | 0.063 | 0.996 | 0.009 | 0.999 |
| whole set | 0.215 | 0.955 | 0.035 | 0.999 | 0.006 | 1.000 |
| Rotor current | | | | | | |
|  | 0.273 | 0.926 | 0.192 | 0.941 | 0.060 | 0.996 |
|  | 0.239 | 0.943 | 0.128 | 0.984 | 0.010 | 0.999 |
| whole set | 0.229 | 0.949 | 0.093 | 0.991 | 0.007 | 1.000 |

For checking regression model adequacy, a population exceeding the base generator by all objectives was formed, Figure 4 (left). Should be noted, number of generators after NSGA-II satisfying this condition is but three. Figure 4 (right) shows the result of checking model adequacy with an average error about 1%. The comparison shows that design improvements are obtained. Stator core mass is 2.8% (1.5% — previous NSGA-II value), rotor current is 1.8% (1.5%), steel losses are 3.5% (1%), — 3.5% (3%).

Reducing the sample by clustering and forming populations by regression model allow us not only reducing the time to calculate multi-criteria problems, but also allows to generate populations in the desired areas without studying the dependence of predictors. The average overall acceleration is about two-three times without taking into account clustering with population reduction. Regression models allow us to conveniently form populations, with no continuous ranges of parameters.

|  |  |
| --- | --- |
| Mass of stator core, p.u.  Rotor current, p.u  Iron  losses, p.u. | Iron  losses, p.u.  Rotor current, p.u  Mass of stator core, p.u. |

Figure 4 — Pareto-optimal front of the regression model with generators exceeding the base generator by all objectives (left); Pareto-optimal front after checking model adequacy (right).

1. **Conclusion**

A speeding up MOO with hard-to-calculate objectives has been investigated empirically by the example of hydro-generator parameters. The results show that the performance of the method provides the decision maker with a tool to reduce labor costs when evaluating different options. The average overall acceleration is about two-three times without taking into account clustering with population reduction. Using clustering and regression analysis allow us to reduce the sample size and optimization time, which is especially relevant for problems where each individual requires a long computation time.

1. **References**
2. Gulay, S. On optimization hydro-generator parameters by NSGA-II / S. Gulay, D. Verkhovtsev // IOP Conference Series: Materials Science and Engineering. – 2019. – Т.643, №1. – C.1, 2009. DOI: 10.1088/1757-899X/643/1/012009
3. Korovkin N.V., Potienko A.A. The use of a genetic algorithm for solving electric engineering problems. Elektrichestvo, #11, 2002, P. 2 – 15. ISSN 00135380.
4. Korovkin, N.V. Optimization of hydro generator parameters / N.V. Korovkin, S.L. Gulay, D.A. Verkhovtsev // Proceedings of the Russian Academy of Sciences. Power Engineering. - 2019. - №4. - C.42-50. DOI: 10.1134/S0002331019040071
5. Deb K, Agrawal S, Pratap A and Meyarivan T 2000 A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II. In: Schoenauer M. et al. (eds) Parallel Problem Solving from Nature PPSN VI. PPSN 2000. Lecture Notes in Computer Science, vol. 1917. Springer, Berlin, Heidelberg.
6. Milligan, G.W. An examination of procedures for determining the number of clusters in a data set / G.W. Milligan, M.C. Cooper // Psychometrika. – 1985. – Т.50, №2. – C.159–179.
7. Glebov I A and Dombrobsky V V 1982 Hydro-generators — Energoizdat
8. James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert (2017). An Introduction to Statistical Learning (8th ed.). Springer Science+Business Media New York.
9. Breiman, L. Random Forests / L. Breiman // Machine Learning. – 2001. – Т.45, №1. – C.5–32.
10. Classification and Regression Trees / L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone: Routledge, 2017.

Llorà, X., Sastry, K., Alías, F., Goldberg, D. E., & Welge, M. (2006). Analyzing active interactive genetic algorithms using visual analytics. Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation - GECCO ’06.

Ishibuchi, H. Evolutionary many-objective optimization: A short review / H. Ishibuchi, N. Tsukamoto, Y. Nojima // 2008 IEEE Congress on Evolutionary Computation (CEC): 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence): IEEE, 01.06.2008 - 06.06.2008. – C.2419–2426.