National Research University
Higher School of Economics

Sergey Vladimirovich Shvydun

Superposition models in data analysis

PhD Dissertation summary
for the purpose of obtaining academic degree
Doctor of Philosophy in Applied Mathematics

Academic supervisor:
Doctor of Technology
Senior Researcher
Fuad Tagievich Aleskerov

Moscow– 2020

Statement of the problem

The problem of multi-criteria decision-making and individual and social choice has attracted a considerable interest in the context of Big Data analysis. Thereby, one of the most important characteristics of decision-making procedures is their computational complexity. Unfortunately, most existing procedures have a high computational complexity (quadratic or even higher), which leads to the application of approximate procedures that have lower computational complexity but significantly reduce the quality of the results.

One possible solution of the problem is the application of superposition principle that consists in the sequential application (composition) of different procedures where the result of the previous procedure is the input for the next procedure. The main advantage of superposition principle is that it reduces the computational complexity by applying choice procedures with a low computational complexity on first stages and more accurate choice procedures on final stages. Superposition models also have a high degree of interpretability, since they use a combination of simple methods rather than a complex one. However, as superposition operation is not commutative, i.e. the final results depend on the sequence of procedures, there is a need to analyze the properties of various superposition models.

To what extent the problem has been developed

In 1900, D. Hilbert formulated 23 most important problems, which were the most significant for mathematics. The problem of whether functions of several variables can be expressed as superposition of functions of fewer variables is one of these problems, which is also known as the Hilbert's thirteenth problem. D. Hilbert himself put forward a hypothesis about the impossibility of such expression and showed that analytic functions of three variables cannot be represented as the superposition of two variables. However, the affirmative answer to this general question for the class of continuous functions was given in 1950s by V.I. Arnold and A.N. Kolmogorov. For the class of choice functions this problem was formulated and studied by M.A. Aizerman and F.T. Aleskerov in 1990. The

authors studied choice functions closedness with respect to superposition operator, considered different types of two-stage procedures and defined under which conditions these procedures can be reduced to one-stage procedures. There are also some other works of M.A. Aizerman, A.V. Malishevsky, V.I. Vol'skiy, F.T. Aleskerov and Y. Cinar related to the analysis of some two-stage choice procedures. Nevertheless, only some general properties of two-stage procedures have been observed while the problems of their computational complexity and subsequent application to data analysis problems have not been considered yet.

Problems and goal of the research

The goal of the research is to define the properties of various superposition models and to construct models based on superposition principle which provide high performance in applied research.

Other goals of the research

1. analyze foreign and domestic studies on superposition of choice functions;
2. define a list of normative properties showing how a final choice is changed due to the changes of the set of alternatives or criterial scores.
3. study normative properties of existing choice procedures which are the basis for constructed superposition models;
4. study normative properties and computational complexity of superposition models;
5. construct an effective model based on superposition of choice functions for the search of relevant documents.
6. construct an effective model based on superposition of choice functions for tornado detection problem.
7. construct an effective model of areas allocation in the Arctic, including models based on the superposition principle.

Relevance of the study

The analysis of properties of different choice procedures and their superposition is of great importance for the following reasons. Since these properties determine how stable and constant the final set of alternatives is, it gives a more detailed perception of the main features, advantages and disadvantages of applied choice procedures.

The analysis of these properties has also a significant impact to Big Data analysis. In particular, an information about the properties allows the researcher to understand whether the choice procedure should be applied to the whole set of alternatives or to its subsets, whether it is possible to exclude irrelevant alternatives from the future analysis or whether it is necessary to recalculate the results of the procedure if criterial scores for some alternatives have been changed. Another important part of the study is the analysis of the computational complexity of superposition models which allows to determine which of them can be applied to large data sets.

Finally, there are a number of areas in which the choice problem cannot be solved efficiently in a sufficient time due to high computational complexity of the existing models. Thus, there is a need to consider new models, including those based on the idea of superposition principle.

Personal contribution of the author to the research

Theorems 1-2 which determine the properties of various choice procedures have been formulated and proved by the author personally. The author also evaluated the computational complexity of superposition models under consideration.

The research related to the tornado prediction problem has been performed in collaboration with N.N. Baiborodov, S.S. Demin, Professors V.I. Yakuba and F.T. Aleskerov (HSE University) as well as Professors M. Richman and T. Trafalis (the University of Oklahoma). The author proposed a model based on superposition of super-threshold procedures and evaluated its efficiency.

The research related to the search problem has been performed in collaboration with E.O. Mitichkin, Professor V.I. Yakuba and Professor F.T.

Aleskerov. The author constructed feature selection procedure, superposition models, evaluated their efficiency and implemented them as a software.

The research related to areas allocation problem in the Arctic has been performed in collaboration with Professor F.T. Aleskerov. The author worked on data collection and preprocessing procedures, constructed several models of areas allocation, including the superposition model, and performed their software implementation.

The results of the study were presented at the following scientific conferences, workshops and seminars:

1. 3rd International Workshop “Models of Influence and Network Theory”, Paris (France). Topic: "Over-lath superposition algorithm and its use to the ranking of options in the problem of search", 14.05.2012-15.05.2012;
2. All-Russian seminar „Expert evaluation and data analysis”, ICS RAS, Moscow (Russia). Title: "Ranking Algorithms Based on Superposition Principle and Their Application to the Information Retrieval", 28.11.2012;
3. ITQM 2013 conference, Suzhou (China). Title: "Super-threshold Procedures and Their Application to the Search Problem", 16.05.2013-18.05.2013;
4. Seminar of DeCAn lab (HSE University), Moscow (Russia). Title: "Choice Procedures and their Properties", 10.06.2013;
5. 12th All-Russia Meeting on Control Problems, Moscow (Russia). Title: "The Analysis of Normative Properties of Two-Stage Choice Procedures", 19.06.2014;
6. IFORS 2014 conference, Barcelona (Spain). Title: "Two-Stage Superposition Choice Procedures and their Properties", 13.07.2014-18.07.2014;
7. All-Russian seminar „Expert evaluation and data analysis”, ICS RAS, Moscow (Russia). Title: "Two-Stage Choice Procedures and Their Properties ", 28.01.2015;
8. MCO 2015 conference, Metz (France). Title: "Properties and Complexity of Some Superposition Choice Procedures", 11.05.2015-13.05.2015;

9. EURO2015 conference, Glasgow (United Kingdom). Title: "Normative properties of the superposition of multi-criteria choice procedures", 12.07.2015-15.07.2015;
10. CYBCONF 2017 conference, Exeter (United Kingdom). Title: "A Mathematical Approach to Conflict Resolution in the Arctic Region", 21.06.2017-23.06.2017.
11. IFORS 2017 conference, Quebec (Canada). Title: “Conflict resolution models in the Arctic region”, 17.07.2017-21.07.2017;
12. MLSD'2017 conference, Moscow (Russia). Title: “Superposition Models of Conflict Resolution in the Arctic Region”, 02.10.2017-04.10.2017;
13. UTFORSK Norwegian-Russian Workshop on Arctic Logistics, Molde (Norway). Title: “Allocation of Disputable Zones in the Arctic”, 10.10.2017-12.10.2017;
14. Seminar "Mathematical Economics", Central Economic Mathematical Institute (CEMI), Moscow (Russia). Title: “Allocation of Disputable Zones in the Arctic Region”, 06.02.2018;
15. EURO2018 conference, Valencia (Spain). Title: “On various solutions of areas allocation problem”, 8.07.2018-11.07.2018;
16. All-Moscow seminar "Mathematical methods of decision analysis in economics, finance and politics", HSE University, Moscow (Russia). Title: "Superposition Models in Data Analysis", 17.10.2018;
17. Autumn school „Current trends in decision-making analysis“,HSE University, Moscow (Russia). Title: "Superposition Models in Data Analysis", 07.11.2018;
18. SVF-8063 School of Society and Advanced Technology in the Arctic, Longyearbyen (Norway). Title: “Superposition (Composition) Models in Data Analysis”, 13.10.2019-19.10.2019.

Description of the research methods

Methods of optimization theory, choice theory, theory of computation, game-theoretic methods and methods of modern applied algebra are used for the theoretical analysis of properties. Experimental part of the research involves computer modeling, including simulation modelling.

Theoretical significance of the research consists in

1. identifying the list of normative conditions which are satisfied for existing multi-criteria choice procedures;
2. identifying the list of normative conditions which are satisfied for two-stage superposition models based on existing multi-criteria choice procedures;
3. the analysis of computational complexity of superposition models under consideration;
4. the development of new mathematical models for the search of relevant documents, for tornado detection problem and for disputable areas allocation in the Arctic.

Practical significance of the research consists in the high efficiency of the proposed models in the search problem, tornado prediction problem and the problem of areas allocation in the Arctic. Moreover, the proposed models can be applied in relevant prediction problems as well as in the study of interaction between various agents in the economic and political fields.

Basic results to be defended

1. normative properties and computational complexity of 28 existing choice procedures and 617 two-stage procedures have been analyzed;
2. superposition models that provide high performance in the search problem, tornado prediction problem and the problem of areas allocation in the Arctic have been constructed.

Scientific novelty of the work. The following new scientific results have been obtained in the PhD thesis:

1. an exhaustive study of normative properties for 28 existing choice procedures and their superposition has been performed;
2. there has been analyzed a computational complexity of various superposition models and obtained a list of models which can be applied to large data sets;
3. there have been constructed superposition models that provide high performance in the search problem, tornado prediction problem and the problem of areas allocation in the Arctic.

Results and conclusions of the study

1. Within the framework of the study, rationality conditions of existing choice procedures and their superposition have been analyzed. Two theorems are proved showing which normative conditions are satisfied for choice procedures under consideration.

Among 28 existing choice procedures, only super-threshold choice procedure with a constant threshold value satisfies all rationality conditions (except for the non-compensatory condition). As for the remaining procedures, most of them do not satisfy any condition. The Heredity condition is satisfied for the simple majority rule and procedures that choose undominated (by majority relation) alternatives from the initial set. The Concordance condition is satisfied only for some procedures based on the pairwise comparison of alternatives. Only the minimal dominant set and the simple majority rule satisfy the Outcast condition. More details about the properties of existing rules are provided in Appendix B of the study.

As for the study of 617 two-stage procedures, it was found that most procedures do not satisfy any normative conditions. This feature can be explained by the fact that most existing choice procedures also do not satisfy these rationality conditions. Only some two-stage procedures satisfy the monotonicity condition 1 and other rationality conditions. For instance, two-stage procedure, which consequently applies the Fishburn and simple majority rules, satisfies the Heredity (**H**), the Concordance (**C**) and the 1st monotonicity conditions. More details about

two-stage procedures, which satisfy normative conditions, are provided in Appendix C of the research work.

Information about the properties of existing choice procedures and their superposition is of a great significance for data analysis. For instance, if the procedure satisfies the monotonicity condition 1 and the criterial scores have been changed for the chosen (in the initial set) alternative, the results of the procedures should not be recalculated. Another example, if the procedure satisfies condition **C**, the calculation of the results can be performed on subsets of the initial dataset, which allows to reduce the runtime of the procedure.

Although most procedures do not satisfy any rationality conditions, there was obtained a list of procedures that satisfy some interesting properties, which opens the way for their use in various data analysis problems.

2. A computational complexity of two-stage choice procedures and their runtime on test data are evaluated. It was shown that two-stage choice procedures, which use choice procedures with a high computational complexity on the first stage, require more time than other procedures and, consequently, cannot be applied to Big Data. There has been obtained a list of procedures with a computational complexity which highly depends on the total number of alternatives remained after the first stage. There were also identified superposition models which can be applied large datasets. Finally, a runtime of some superposition models on the artificial dataset has been evaluated.

3. Three models based on superposition principle have been proposed and applied to the search and tornado prediction problems as well as to the problem of areas allocation in the Arctic.

For tornado prediction problem, there has been proposed the superposition of super-threshold choice procedures that eliminates irrelevant objects using 1 or 2 parameters at each stage. Our model was tested on real data. The results showed higher efficiency comparing to the other existing techniques (logistic regression, support vector machines (SVM), random forest (RANF), etc.).

For the search problem, there has been proposed a model that defines the relevance of documents. The proposed model is based on the superposition of super-threshold choice procedures. It is distinguished by its simplicity and provides higher efficiency comparing to the standard techniques (decision tree and SVM).

Finally, several models of areas allocation to resolve the problem of mutual interests are proposed. It is believed that the presented models might ease the decision making process in international relations. Although the results of different models were almost identical (90%), the model based on the idea of superposition principle requires less iteration steps. One should also mention that the purpose of the study is to develop mathematical models of areas allocation. None of areas allocation scenarios is proposed in this work.

Thus, superposition models showed high performance in different fields of data analysis.

List of publications in the topic of the study

Publications from the list of peer-reviewed Q1 journals indexed by Web of Science and Scopus

1. Aleskerov F. T., Shvydun S. Allocation of Disputable Zones in the Arctic Region // Group Decision and Negotiation. 2019. Vol. 28. No. 1. P. 11-42.

Publications from the list of journals indexed by Web of Science and Scopus international citation databases

2. Shvydun S., Aleskerov F., "A Mathematical Approach to Conflict Resolution in the Arctic Region", 2017 3rd IEEE International Conference on Cybernetics (CYBCONF), Exeter, 2017, pp. 145-151.
3. Shvydun S. (2015) Properties and Complexity of Some Superposition Choice Procedures. In: Modelling, Computation and Optimization in Information Systems and Management Sciences. Advances in Intelligent Systems and Computing, vol 360. Springer, Cham. P. 475-486.

4. Aleskerov F., Mitichkin E., Shvydun S., Yakuba V. Super-threshold Procedures and Their Application to the Search Problem // Procedia Computer Science. 2013. No. 17. P. 1121-1124.

Other publications

5. Aleskerov, F., Demin, S. & Shvydun, S. Superposition of Choice Functions and Its Application to Tornado Prediction and Search Problems. SN COMPUT. SCI. 1, 68 (2020).
6. Shvydun S. V. Superposition Models of Conflict Resolution in the Arctic Region, in: Proceedings of the 10-th International Conference "Management of Large-Scale Systems Development" (MLSD'2017), M.: ICS RAS, 2017. P. 452-455.
7. Aleskerov F. T., Baiborodov N., Demin S. S., Richman M., Shvydun S. V., Trafalis T., Yakuba V. I. Constructing an efficient machine learning model for tornado prediction. / Working paper WP7/2016/05, M.: HSE Publishing House, 2016.
8. Shvydun S. V. Normative properties of multi-criteria choice procedures and their superpositions: I / Working paper WP7/2015/07, M.: HSE Publishing House, 2015.
9. Shvydun S. V. Normative properties of multi-criteria choice procedures and their superpositions: II / Working paper WP7/2015/07, M.: HSE Publishing House, 2015.
10. Shvydun S.V. Normative Properties of Two-Stage Procedures // In: 12th All-Russia Meeting on Control Problems (VSPU-2014). M.: ICS RAS, 2014. P. 7977-7985 (in Russian).

Copyright certificates, patents

1. Fuad T. Aleskerov, Evgeny O. Mitichkin, Vyacheslav V. Chistyakov, Sergey V. Shvydun, Viacheslav I. Iakuba USA Patent «Method for selecting

valid variants in search and recommendation systems» # US10275418 B2 of 04/30/2019.

2. Patent #2543315 Method for selecting valid variants in search and recommendation systems (variants), Priority 22.03.2013, Registered in the State Register of Russian Federation 27.01.2015. (coauthors: Aleskerov F.T., Mitichkin E.O., Chistyakov V.V., Iakuba V.I.)
3. Copyright certificate for the computer software 'Choice and ranking of alternatives using superposition of positional rules' (coauthors: Aleskerov F., Mitichkin E., Yakuba V.), Russian Federation, # RU 2013618228, 04.09.2013

The research results have been used grants and research projects

1. RFBR grant № 12-01-00226a «Superposition Choice models » (2012-2014).
2. «Modeling of decision making in social and economic systems», the research project of the Department of Mathematics is a division of the HSE Faculty of Economics, project supervisor: Aleskerov F.T., 2010.
3. «Design of economic mechanisms», the research project of the International Laboratory of Decision Choice and Analysis (HSE University), project supervisor: Aleskerov F.T., 2011.
4. «Research of new methods and approaches in the field of mathematical modeling and design of mechanisms in the social, economic and political spheres», the research project of the International Laboratory of Decision Choice and Analysis (HSE University), project supervisors: Aleskerov F.T., Maskin E., 2013.
5. «Theoretical and numerical study of advanced mathematical models in the social, economic, political and financial fields», the research project of the International Laboratory of Decision Choice and Analysis (HSE University), project supervisors: Aleskerov F.T., Maskin E., 2014.
6. «Data analysis and decision making in social, economic and political systems», the research project of the International Laboratory of Decision

Choice and Analysis (HSE University), project supervisors: Aleskerov F.T., Maskin E., 2015.

7. «Analysis, choice and decision making in socio-economic, political and financial spheres: new models, methods and algorithms», the research project of the International Laboratory of Decision Choice and Analysis (HSE University), project supervisors: Aleskerov F.T., Maskin E., 2017.