On Assessment of Performance of Systems by Combining On-the-Job and Expert Attributes Scales

Edward Pogossian

The Division of Computational and
Cognitive Networks
Institute for Informatics and
Automation Problems
National Academy of Sciences of
Armenia
e-mail: epogossi@aua.am

ABSTRACT

We develop a methodology of assessment of performance of systems based on the following two methods:
On-the-Job Competition Scales (JCS) method developed for measuring performance of competitive systems or their constituents by the original, used in practice criteria of their effectiveness and
Logic Scoring of Preference (LSP) method where for evaluation of systems elementary criteria, expert attributes are used for assessment followed by their aggregation into a global quality indicator called global preferences.

The methodology aimed to combine the strengths of LSP and JCS approaches while weakening some of their shortages to be applied to quantitative assessment of performances of competitive, education and software systems.

Keywords

performance assessment, tournaments, attribute scales

1. ON-THE-JOB COMPETITION SCALES

These JSC methods aimed to evaluate decision making systems (DMS) or their constituents in competitions. Given competition it allows to order the DMS by their on-the-job performance, or absolute scales, in accordance of comprehensive comparisons of performances of all competitors by the base criteria of success declared in the original definitions of the competitions.

For example, on-the-job performance in sport competitions, e.g., in chess, corresponds to robin-round tournaments that induce absolute scales determined by sum of scores of participants gained in those comprehensive competitions.

Those scales may be explicit or implicit, hypothetical or in a real usage.

Well known explicit and real usage scales are the ones in chess and tennis. Implicit scales are used for personnel assessment, for grading scientist and any professional skill, organizations, and in general, any human-based systems.

Those absolute scales possess *transitivity*, either *quasi-transitivity* characteristic doing possible by local tournaments for arbitrary groups of competitors to form their orderings consistent with, or isomorphically embedded into the absolute scale.

The quasi-transitivity tournaments are widely used in chess where Elo-based rating system is a world wide accepted scale to measure chess masters performance in an absolute sense, i.e., invariant to the location of chess masters, a time period when they are playing and number of participants in the local tournaments.

A framework for DMS software evaluation on the base of combinatorial games like chess was developed in [1,2]. Further development of those ideas resulted in development of an absolute scale for management skill assessment in oligopoly competitions understandable as a on-the-job performance of managers in all possible oligopoly competitions [3,4]. State-of- the-art of that approach is presented in [5-9].

Adequate simulation of the DMS is an essential precondition for constructing the JCS scales. That simulation includes the models of competing environment and knowledge-based (KB) game tree search procedures.

If alternative strategies are presented as a manageable table the best strategy may be found by the max min, max Sum, Condorcet' criteria.

Exhaustive search based methods and approximations to them to find acceptable strategies are based on the min max method with a variety of cutting search heuristics: evaluation function , branch and bound method, etc. Among them the Intermediate Goals at First (IGAF) and methods by [10-12], Common Planning and Dynamic Testing (CPDT) and PPIT (Personalized Planning and Integrated Testing) [13-15] are notable by their intensive inclusion of expert knowledge.

2. LOGIC SCORING OF PREFERENCE SCALES

The LSP methods are widely used for assessment of performances of people or systems in education and industry, accumulate the experience and knowledge of experts in applications and there is no restrictions on the new areas of applications.

For example, evaluation of software requires for a given user in some applications to select software systems, constituents of software or tools for their development that are the best among their competing alternatives.

LSP is applied to arbitrary software while evaluation criteria are determined by the view of experts what, in general is worth, to measure there.

According to [16] the LSP method can be classified as one of methods for multi-attribute and multiple-objective decisions based on cardinal ranking of alternatives under certainty and with prior articulation of preferences.

First the classes of WE users are investigated and the type of users for which we want to develop the criterion function is selected. The next step is to develop a WE user model. Using a questionnaire developed in [17,18] the selected users specify value preferences for all main WE attributes. The attributes that are selected as relevant for WE evaluation are called performance variables. They are used

as inputs for the LSP criterion function. A statistical analysis of the user questionnaire answers is used to compute the best values of parameters of the LSP criterion function. Therefore, the LSP criterion function is calibrated to reflect the needs of the selected type of users. It includes all requirements that the evaluated WEs are expected to satisfy. For each competitive WE it is necessary to prepare the set of values of all performance variables (inputs for the LSP criterion function). For each competitive system the LSP criterion generates the overall performance indicator which is called the global preference and used for cardinal ranking of competitive systems. The global preference can be approximately interpreted as the percentage of satisfied requirements. Consequently, at the end of the evaluation process we generate a cardinal ranking of competitive systems using the scale from 0 to 100%.

3. REQUIREMENTS TO COMBINING LSP AND JSC SCALES

3.1. We aim to develop application areas context dependent assessment methodology, a shell based on simulation, methods of statistics and tournaments.

The methodology applied, particularly, to software assessment – the LSP+, could complement expert attribute based LSP scales with competitive performance measuring JSC ones and let us evaluate performance of constituents of software systems (but , yet, not the tools for their development).

Focusing software assessment LSP+ has to combine strengths of LSP method and JCS on-the-job performance scales.

We consider LSP as the base for the LSP+ which has to meet the following additional requirements.

1. It is natural to expect that the best alternative software chosen by the LSP+ method can be equal, or, at least, comparable with ones chosen by original, used in practice quality criteria.

Thus, the LSP+ has to clarify how the best alternative chosen by the LSP relates to one chosen by the original criteria of the performance of the system.

2. The LSP+ has to minimize human subjective intervention in the measurement process.

Particularly, for the LSP stages of selection the base criteria, say functionality, testability, etc., the constituents of the software to evaluate, the attributes to detail their characteristics and scales to quantify their values, and, finally, the selection of a method to integrate measurement of all constituents to get global estimates of the systems.

Ideally, orderings of the set of competing alternative software by different experts using the LSP+ have to coincide.

3.2. To meet the above requirements we modernize LSP by JCS scales as follows:

1. A software evaluation problem for DMS has to be chosen in a competition and for alternatives the JCS scale constructed which meets the following requirements:

- it uses the DMS success original criteria
- human intrusion and corresponding subjective factors are minimal
- JCS scale comparing the alternatives has to the result of a comprehensive comparison of all DMS in the competition.

2. The LSP and JCS scales are compared and the LSP method complemented by an iterative procedure to make the LSP scale consistent with the JCS one.

3. The tuning procedure will be adjusted by series of similar experiments for DMS in competitions to make it stable and reliable in constructing LSP scales consistent with the JCS ones

4. The results of experiments are generalized and the LSP method modernized comprising the new LSP+ method.

4. CONSTRUCTING JCS SCALES

4.1. Let M be a competitive environment, or a competition, say, chess, oligopoly, military, economic, others.

And let K be given criteria of the success in the competition,

{C} - the set of DMS in that competition M, e.g., organizations, human, other DMS, C - some DMS from the {C},

{T}- a variety of versions of a constituent T of the DMS that is the target for a measurement of a quality, e.g., a promotion instrument

We assume (**A1**) that performance of DMS in M can be measured by comprehensive tournaments or other regular ways of comparisons, thus, provide an absolute ordering, or a scale A(M,K) of measuring the quality of DMS in M.

And we assume (**A2**) that for fixed DMS C and varied constituents T in C the subscale A(M,K,C,T) of the scale A(M,K) can be considered for measuring the quality of T in terms of criteria defined for the quality of performances of C.

Finally, for creating the JCS scale we accept the following assumptions:

A3.There are adequate game models to simulate competitions and their determiners including the competitive environments and DMS, with all varieties of their constituents.

A4. The ordering of models of competitors is isomorphic to the ordering of real competitors.

A5. The assumptions 3 and 4 are sufficient to induce a scale of models of DMS in a model of some competitive environment which is isomorphic to the absolute scale of the original DMS in the original competitions and satisfy the quasi-transitivity constraint.

Assumption A5 comprises A3 and A4 ones and let measure performances by simulation.

Namely, using Assumption 3-5 the scale A(M,K) and its subscales for the original competitions will induce an isomorphic scales A(M,K)m and subscales A(M,K,C,T)m for the models of those competitions that can be run by simulation.

Thus, following A1-A5 the scales A(M,K,C,T)m will meet the quasi-transitivity constraints and, thus, will let order arbitrary pairs of target constituents T1 and T2 by local measurements in the metrics of the absolute scales A(M,K,C,T).

4.2. We argue the assumptions as follows.

1. *To preserve an adequacy* of simulation of competitions we are going to relay, particularly, on the models of business competitions widely used in educational business games [Markstrat, Brandmaps, see www.towson.edu/~absel], models of organizations and DMS constructed by the specialized simulation tools like Proforma [] as well as KB DMS , particularly, RGT (Reproducible Game Trees) Solvers [15] with CPDT or PPIT acceptable strategy search engines developed in [2,13-15,19-21].

Given variety of game models we are going to choose the best of them by comprehensive comparative experiments.

The first question to be answered is whether the game tree model contains all relevant for the competition strategies?

2. To answer this question game tree strategies exhaustive search procedure will be compared with the performances of a representative sample of experts in the field.

The experiments will be organized by the scheme successfully approbated in constructing game tree models for dynamic counteractions against network intrusions [Pogossian, Djavadyan03].

The second question to be resolved is whether strategy search algorithms are able to preserve the quality of search comparable with the exhaustive one. For that goal an experience gained in [13] and [5, 9] can be applied.

The third question relates to preserving diversity of the DMS models sufficient to create a scale of models with increasing strength in decision making. Increasing universality of the organizations simulation tools like Proforma is a good premise for it. Another leverage to achieve it is the KB embedded in the DMS that allows to relay on human ability to improvement of performance by gaining new knowledge.

3. *The assumption* A4 can be valid if, particularly, KB DMS were considered as the models of competitors, able to store knowledge and improve their performances correspondingly with the amount and quality of gained knowledge [2, 22].

Particularly, the organizational knowledge includes many measurable and communicable components that may be reproduced in their models along with human depended ones measurable, yet, only to some extent (*Moorman*).

It is worth to mention that DMS performance measurement issues are tightly interconnected with a fundamental issues of the nature of the knowledge and its processing [23].

5. CONCLUSION

The base of a methodology of assessment of performance of systems combining the strengths of methods measuring performance of competing systems or their constituents by original, used in practice criteria of effectiveness and ones evaluating systems by expert attributes followed by their aggregation into a global quality indicators called the global preferences was presented.

We have discussed the basics of those methods and the requirements to combine them for more effective assessment of performances.

An approach of using of absolute, i.e. real practice, scales of competing systems for assessing constituents of those systems was presented and argumentation of the validity of the approach was provided.

The methodology can be applied to quantitative assessment of performances of competitive, education and software systems as well as can be adapted to measuring mental performances by providing certain standard scales of human performances.

REFERENCES

[1] E. Pogossian (1977): Tournaments Approach in Comparing Effectiveness of Information Structures. *Proceedings of the conference "Mathematical Models in Complex Systems Simulation", Kiev, pp.16-19(in Russian).*

[2] E. Pogossian, *"Adaptation of combinatorial algorithms"*, Academy of Sci. of Armenia, Yerevan, (in Russian), p. 293, 1983.

[3] E. Pogossian (1997): Ability of Efficient Evaluation of Knowledge Based Management Strategies. *Proceedings of the ABSEL-97, New Orleans.*

[4] E. Pogossian, "Increasing efficiency of management skill assessment", *Proc. of the 26th annual conference of the Association for Business Simulation and Experiential Learning* (ABSEL-99), Philadelphia, March 1999, pp. 144-148.

[5] T. Baghdasaryan, E. Danielyan, and E. Pogossian, "Testing oligopoly strategy plans by their on the job performance simulation", *Proceedings of the International Conference CSIT2005*, Yerevan, Armenia, 2005, pp.8.

[6] E. Danielyan, "Implementation of local tournaments methodology for evaluation of management strategies", *Proceedings of the International Conference CSIT2009*, Yerevan, Armenia, 2009.

[7] E. Pogossian "Ability of efficient evaluation of knowledge based management strategies", *Proceedings of the 24th Annual Conference of the Association for Business Simulation and Experiential Learning*, (ABSEL-97), New Orleans, March, 1997, pp. 93-97.

[8] E. Pogossian, "Business measurements by On-The-Job competition scale", *International Conference "Management of Small Business: Problems, Teaching, Future"*, Sevastopol, Ukraine, 2004, pp.47-51.

[9] E. Danielyan and E. Pogossian, "On expanding of method of local measurement to evaluation of management strategies", *Proceedings of SEUA*, 2009, 3p.

[10] M.M. Botvinnik, "Computers in chess: solving inexact search problems", *Springer Series in Symbolic Computation, with Appendixes*, Springer-Verlag, New York, 1984.

[11] D.E. Wilkins (1982) Using knowledge to control tree searching. *AI, v.18, 1-51.*

[12] E. Pogossian (2001) Focusing Management Strategy Provision Simulation, CSIT2001, Yerevan.

[13] E. Pogossian, A. Javadyan, and E. Ivanyan, **"**Effective discovery of intrusion protection strategies**",** *The International Workshop on Agents and Data Mining, St. Petersburg, Russia, Lecture Notes in Computer Science*, vol. 3505, 2005, pp. 263-274

[14] E. Pogossian, V. Vahradyan, and A. Grigoryan, "On competing agents consistent with expert Knowledge", *Lecture Notes in Computer Science, AIS-ADM-07: The Intern. Workshop on Autonomous Intelligent Systems - Agents and Data Mining*, St. Petersburg, June 5 -7, 2007, 11p.

[15] Pogossian E. Effectiveness Enhancing Knowledge Based Strategies for RGT Class of Defense Problems NATO ASI 2011 Prediction and Recognition of Piracy Efforts Using Collaborative Human-Centric Information Systems, Salamanca, Spain, September 19-30, 2011, pp.16.

[16] M. Mollaghasemi and J. Pet-Edwards, *Making Multiple-Objective Decision*s. IEEE Computer Society Press, 1997.

[17] J.J. Dujmovic, *A Method for Evaluation and Selection of Complex Hardware and Software System*s. The 22nd International Conference for the Resource Management and Performance Evaluation of Enterprise Computing Systems. CMG 96 Proceedings, Vol. 1, pp. 368-378, 1996.

[18] A.R. Bayucan, *Quantitative Evaluation of Windowed Environment*s. M.S. Thesis, San Francisco State University, 1996.

[19] K. Khachatryan and S. Grigoryan, "Java programs for presentation and acquisition of meanings in SSRGT games", *Proceedings of SEUA Annual conference*, pp. 127-135, Yerevan, Armenia, 2013.

[20] S. Grigoryan, "Structuring of Goals and Plans for Personalized Planning and Integrated Testing of Plans", *Mathematical Problems of Computer Science*, vol. 43, pp. 62-75, 2015.

[21] K. Khachatryan, S. Grigoryan and T. Baghdasaryan, "Experiments validating the Be-Have-Do meaning presentation model and matching algorithm for competing and combating problems", *International Conference in Computer Sciences and Information Technologies*, pp. 155-159, Yerevan, Armenia, 2013.

[22] E. Pogossian, "On measures of performance of functions of human mind", *6th International Conference in Computer Science and Information Technologies, CSIT2007*, Yerevan, Armenia, 2007, pp. 149-154.

[23] Pogossian E. Towards Adequate Constructive Models of Processing of Meanings, CSIT 2013 and in IEEE see http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6710347