УДК 519.234

B. Norkin

VISUAL EVOLUTIONARY SEARCH FOR THE PARETO-OPTIMAL DATA

The paper describes an information technology (and software) for interactive visual search for Pareto-optimal data in a large data set (data points). Each data element (point) is a vector with a number of components with values in completely ordered, probably different, spaces/sets. These components are treated as optimization criteria and can be maximized or minimized. The basic problem is to identify a non-dominated data subset with respect to the selected criteria/components with specified directions of optimization. The problem is solved interactively by graphical display of the data in different planes (pairs of coordinates). The second related problem is to order data elements with respect to their power of domination. The latter problem is solved by calculation of two numbers, the numbers of elements that dominate and are dominated by a given element, calculation of their difference and different sizes of data points on the displayed planes.

Keywords: Multi-criteria optimization, Pareto-optimality, data visualization, evolutionary search, interactive optimization.

Introduction

Multicriteria optimization is a generally applicable decision making methodology [7]. It uses a decision space to describe decisions and criteria space to display outcomes of decisions. Decisions and their outcomes are connected by a mathematical, computer or other type of model. The key concept in multicriteria optimization is Pareto optimality, which is based on the concept of domination of decision outcomes. Pareto optimal subset is defined as a set of nondominated decisions. In the criteria space Pareto optimal decisions form an efficient (Pareto optimal) frontier. Multicriteria optimization usually includes two steps: to find Pareto optimal subset and then to select an acceptable compromise point on it. A plenty of methods was developed for multicriteria optimization [11, 12], including the so called evolutionary methods [1, 4]. Data visualization is considered as a valuable tool for multi-criteria optimization. A number of methods and software were developed for this purpose [6, 8–10]. A deeper way to introduce structure into multivariate data is its partial ordering [2, 3].

In the present paper we present an information technology and corresponding software for interactive visual search for Pareto-optimal data in a large data set (data points, objects). Each data element (point, object) is a vector with a number of components with values in completely ordered, may be different, spaces/sets. These components are treated as optimization criteria and can be maximized or minimized. The basic problem is to identify a non-dominated data subset with respect to selected criteria/components and for specified directions of optimization. The problem is solved interactively by graphical display of the data in different planes (pares of coordinates), color selection of preferable points and tracing the colored search path. The second related problem is to order data elements with respect to their power of domination. The latter problem is solved by calculation of two numbers, the number of elements that are dominated by a given element and number of elements that dominate this element, calculation of the domination power as the difference of these two numbers and displaying data points in different colors and sizes according to the domination power. A novel element in our approach to multicriteria optimization is the use of $\vec{\varepsilon}$ -domination concept to regulate domination force of different coordinates/criteria, where $\vec{\varepsilon}$ is a tolerance vector.

A continuous multi-criteria optimization problem can be approximated by a discrete one [17], where a continuous decision space is approximated by a discrete set of points and then a discrete Pareto-optimal frontier is searched. The obtained discrete multi-criteria problem can, in particular, be analyzed by the method of the present paper. In [13, 15] the approach was validated in terms of convergence to approximate Pareto optimal solutions. In [14] this approach was further extended by interactive local random exploration of the Pareto-optimal frontier, i.e. by sampling test points around a selected reference point in the decision space and viewing the results in the criteria space, and so on.

Software system “A-ranger Online”

The software system “A-ranger Online” is designed to support interactive multi-criteria selection of the best from a large but finite set of objects (tens, hundreds, thousands of objects or options), evaluated by a large number of criteria (up to several tens of criteria). With a significant number of criteria *n* optimal choice of objects can be a problem for a user. For example, Saaty’s method [16] for quantitative criteria ranking assumes pairwise comparison of the criteria and filling in $n(n+1)$ matrix coefficients. In the proposed approach in one optimization cycle it is sufficient to specify, as maximum, *n* Boolean values (directions of optimization) and from 1 to *n* real parameters of optimization accuracy.

Problem formulation

It is assumed that there is $(m+1)\times(n+1)$ -table of data (see Table 1).

Table. **Data table**

	Criterion 1	**Criterion 2**	…	**Criterion *n***
Object 1	X_{11}	X_{12}	…	X_{1n}
Object 2	X_{21}	X_{22}	…	X_{2n}
…	…	…	…	…
Object m	X_{m1}	X_{m2}	…	X_{mn}

In the first line of the table (in columns 2,..., $n+1$) the names of criteria (character strings) are written, in the first column of the table (lines 2,..., $m+1$) object names (character strings) are recorded, and in each cell of the table numerical estimate of the corresponding object with respect to the corresponding criterion is recorded. The values of each criterion in a column are comparable with each other, different criteria can compare objects in different measure units (unitless, percent, monetary or other units). It is assumed that for each criterion *k* the optimization direction (maximization or minimization) and a numerical parameter ε (or ε_k) of the tolerance of optimization can be set. Multi-criteria search problem consists in finding the so-called ε-Pareto-optimal (ε-nondominated) objects with respect to some subset of criteria. In addition, each object has three additional indicators, namely, the number of other objects that are ε-dominated by him and the number of objects that he ε-dominates, and the difference between these two indicators. The latter index (rating) can be a basis for the best object choice.

The optimality concept

Object *i* ε-dominates object *j* with respect to criterion *k* (from a set *K*), if $X_{ik} > X_{jk} + \varepsilon_k$ and criterion *k* is maximized (or $X_{ik} < X_{jk} - \varepsilon_k$ and criterion *k* is minimized*),* $\varepsilon = \{\varepsilon_k, k \in K\}$. Here, parameter ε_k controls the power of domination by the criterion *k*: the larger ε_k, the greater superiority in criterion *k* is necessary for the dominance by this criterion and the less significant is the dominance on this criterion compared with other criteria.

The object is called a ε-Pareto-optimal (ε-nondominated) in criteria *K*, if there is no other object that ε-dominates it by this set of criteria.

Accuracy optimization parameter ε_k on criterion *k* can be taken as $\varepsilon_k = \left(\max_i X_{ik} - \min_i X_{ik}\right)\varepsilon$, where $0 \le \varepsilon < 1$. Thus, for one optimization cycle it is enough to set directions for some optimization criteria and to set a common scalar relative optimization accuracy ε (the default value is zero).

Algorithm of the system

1. You need to select and upload a data file to the system with the data in csv-format (comma separated values, and the fractional parts of decimal numbers separated from the integer part by the period).

2. On the computer screen it appears a cloud of points/objects in the plane “Criterion 2 – Criterion 1”, and below it appears a table with the names of the criteria with windows to optimization directions (min, max), windows for optimization accuracies (epsilon), maximum and minimum values of criteria (see Fig. 2).

3. Plane representation of objects can be changed by selecting the criteria for horizontal and vertical axes from the pull-down menus (see Fig. 3).

4. The size of the display points/objects can be changed using two sliders at the top of the screen.

5. The point cloud can be moved sideways using the mouse by moving it to the cloud and holding down the right mouse button. In addition, you can change (zoom) the display scale of point clouds by rotating the mouse wheel.

6. When the mouse cursor is put on a specific point, information about the object and its characteristics of domination is displayed above the point, and in an additional table at the bottom of the screen in the column “Last Pointed Point” the complete set of values of all criteria for this point is shown (see Fig. 2).

7. Any point clouds can be marked by moving the cursor on it and pressing the left mouse button, the point will change its color, and in the table “Point Info” in the column “Optimization Point”

appears complete information (criteria values) on this point.

8. Column "OPTIMIZATION" is designated for choosing concrete optimization criteria and indicating optimization directions for these criteria (min or max) (see Fig. 4). They need not to coincide with displayed criteria. The optimization visually affects color and size of points (domination power of points with respect to optimization criteria).

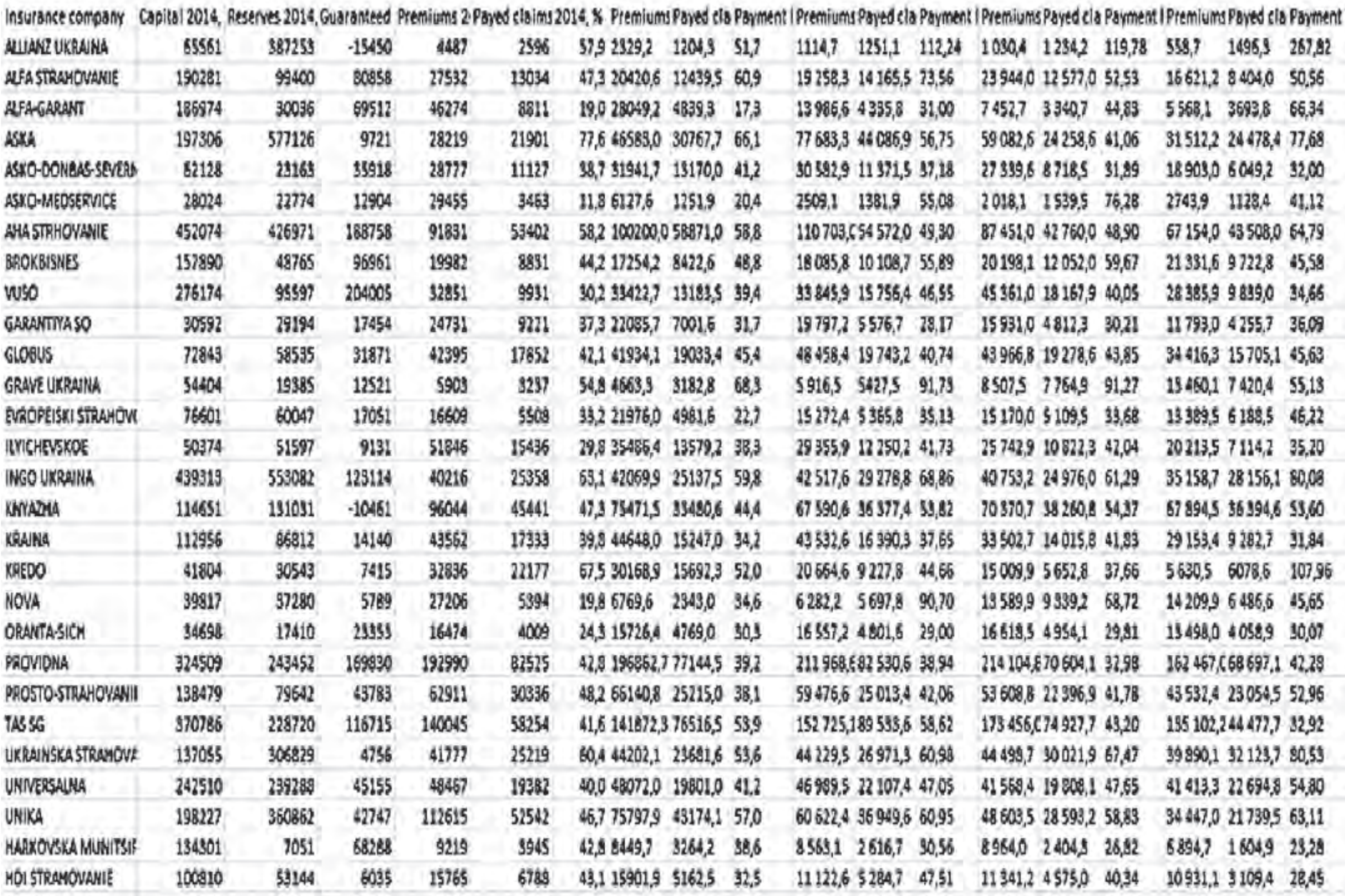
9. In the column "EPSILON" at the bottom table, you can specify a relative to all the criteria (the top box) or absolute optimization accuracy for each criterion.

10. To carry out an optimization loop one need to put cursor and right-click on the button "Optimize" under the cloud picture. After some period of time required to perform calculations on a remote server the cloud points change sizes and colors depending on their relative ranking, non-dominated points become brown. Dominated by only one point become green, the other points are blue.

11. For the next cycle of optimization one should visually analyze the differently colored cloud points in different planes (on various criteria pairs), mark the new reference point for the optimization, to select a new set of optimization criteria and optimization directions, and finally go to step 10 (or to earlier steps).

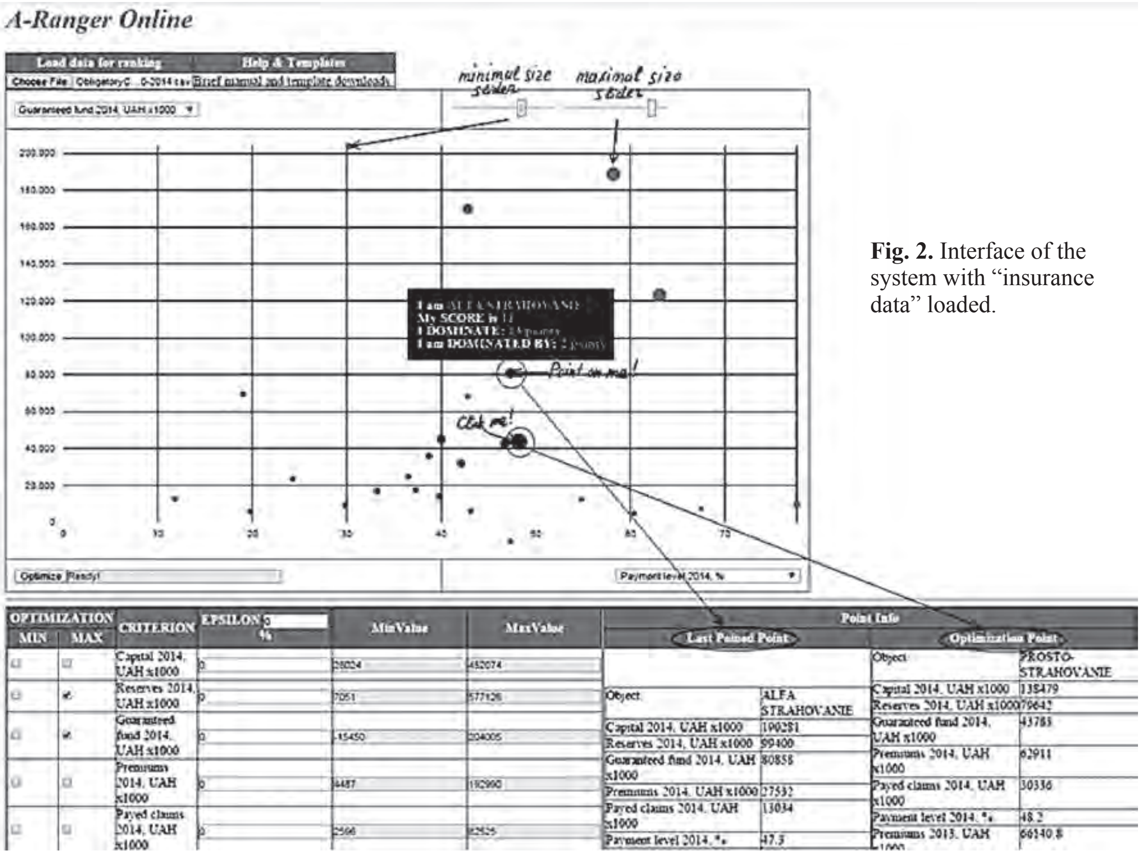
Example (choice of a company for car insurance)

The following pictures illustrate application of the system to selection of a company for car insurance. Insurance companies are characterized by the following indicators: Capital 2014, Reserves 2014, Guaranteed fund 2014, Premiums 2014, Paid claims 2014, Payment level 2014 (=Paid claims 2014/Premiums 2014) and similar indicators for years 2010 – 2013. The criteria of the most interest to maximize are "Payment levels 2010-2014" and "Reserves 2014".



Insurance company	Capital 2014,	Reserves 2014,	Guaranteed	Premiums 2	Payed claims	2014, %	Premiums	Payed cla	Payment l	Premiums	Payed cla	Payment l	Premiums	Payed cla	Payment l	Premiums	Payed cla	Payment
ALLIANZ UKRAINA	65561	387253	-15450	4487	2596	57,9	2329,2	1204,3	51,7	1114,7	1251,1	112,24	1 030,4	1 234,2	119,78	558,7	1496,3	267,82
ALFA STRAHOVANIE	190281	99400	80858	27532	13034	47,3	20420,6	12439,5	60,9	19 258,3	14 165,5	73,56	23 944,0	12 577,0	52,53	16 621,2	8 404,0	50,56
ALFA-GARANT	186974	30036	69512	46274	8811	19,0	28049,2	4839,3	17,3	13 986,6	4 335,8	31,00	7 452,7	3 340,7	44,83	5 568,1	3693,8	66,34
ASKA	197306	577126	9721	28219	21901	77,6	46585,0	30767,7	66,1	77 683,3	44 086,9	56,75	59 082,6	24 258,6	41,06	31 512,2	24 478,4	77,68
ASKO-DONBAS-SEVERI	62128	23163	35918	28777	11127	38,7	31941,7	13170,0	41,2	30 582,9	11 371,5	37,18	27 339,6	8 718,5	31,89	18 903,0	6 049,2	32,00
ASKO-MEDSERVICE	28024	22774	12904	29455	3463	11,8	6127,6	1251,9	20,4	2509,1	1381,9	55,08	2 018,1	1 539,5	76,28	2743,9	1128,4	41,12
AHA STRHOVANIE	452074	426971	188758	91831	53402	58,2	100200,0	58871,0	58,8	110 703,0	54 572,0	49,30	87 451,0	42 760,0	48,90	67 154,0	43 508,0	64,79
BROKBISNES	157890	48765	96961	19982	8831	44,2	17254,2	8422,6	48,8	18 085,8	10 108,7	55,89	20 198,1	12 052,0	59,67	21 331,6	9 722,8	45,58
VUSO	276174	95597	204005	32851	9931	30,2	33422,7	13183,5	39,4	33 845,9	15 756,4	46,55	45 361,0	18 167,9	40,05	28 385,9	9 839,0	34,66
GARANTIYA SO	30592	29194	17454	24731	9221	37,3	22085,7	7001,6	31,7	19 797,2	5 576,7	28,17	15 931,0	4 812,3	30,21	11 793,0	4 255,7	36,09
GLOBUS	72843	58535	31871	42395	17852	42,1	41934,1	19033,4	45,4	48 458,4	19 743,2	40,74	43 966,8	19 278,6	43,85	34 416,3	15 705,1	45,63
GRAVE UKRAINA	54404	19385	12521	5903	3237	54,8	4663,3	3182,8	68,3	5 916,5	5427,5	91,73	8 507,5	7 764,9	91,27	13 460,1	7 420,4	55,13
EVROPEISKI STRAHOVI	76601	60047	17051	16609	5508	33,2	21976,0	4981,6	22,7	15 272,4	5 365,8	35,13	15 170,0	5 109,5	33,68	13 389,5	6 188,5	46,22
ILYICHEVSKOE	50374	51597	9131	51846	15436	29,8	35486,4	13579,2	38,3	29 355,9	12 250,2	41,73	25 742,9	10 822,3	42,04	20 213,5	7 114,2	35,20
INGO UKRAINA	439313	553082	123114	40216	25358	63,1	42069,9	25137,5	59,8	42 517,6	29 278,8	68,86	40 753,2	24 976,0	61,29	35 158,7	28 156,1	80,08
KNYAZHA	114651	131031	-10461	96044	45441	47,3	75471,5	33480,6	44,4	67 590,6	36 377,4	53,82	70 370,7	38 260,8	54,37	67 894,5	36 394,6	53,60
KRAINA	112956	86812	14140	43562	17333	39,8	44648,0	15247,0	34,2	43 532,6	16 390,3	37,65	33 502,7	14 015,8	41,83	29 153,4	9 282,7	31,84
KREDO	41804	30543	7415	32836	22177	67,5	30168,9	15692,3	52,0	20 664,6	9 227,8	44,66	15 009,9	5 652,8	37,66	5 630,5	6078,6	107,96
NOVA	39817	37280	5789	27206	5394	19,8	6769,6	2343,0	34,6	6 282,2	5 697,8	90,70	13 589,9	9 339,2	68,72	14 209,9	6 486,6	45,65
ORANTA-SICH	34698	17410	23353	16474	4009	24,3	15726,4	4769,0	30,3	16 557,2	4 801,6	29,00	16 618,5	4 954,1	29,81	13 498,0	4 058,9	30,07
PROVIDNA	324509	243452	169830	192990	82525	42,8	196862,7	77144,5	39,2	211 968,6	82 530,6	38,94	214 104,8	70 604,1	32,98	162 467,0	68 697,1	42,28
PROSTO-STRAHOVANII	138479	79642	43783	62911	30336	48,2	66140,8	25215,0	38,1	59 476,6	25 013,4	42,06	53 608,8	22 396,9	41,78	43 532,4	23 054,5	52,96
TAS SG	370786	228720	116715	140045	58254	41,6	141872,3	76516,5	53,9	152 725,1	89 533,6	58,62	173 456,0	74 927,7	43,20	135 102,2	44 477,7	32,92
UKRAINSKA STRAHOVA	137055	306829	4756	41777	25219	60,4	44202,1	23681,6	53,6	44 229,5	26 971,3	60,98	44 498,7	30 021,9	67,47	39 890,1	32 123,7	80,53
UNIVERSALNA	242510	239288	45155	48467	19382	40,0	48072,0	19801,0	41,2	46 989,5	22 107,4	47,05	41 568,4	19 808,1	47,65	41 413,3	22 694,8	54,80
UNIKA	198227	360862	42747	112615	52542	46,7	75797,9	43174,1	57,0	60 622,4	36 949,6	60,95	48 603,5	28 593,2	58,83	34 447,0	21 739,5	63,11
HARKOVSKA MUNITSIP	134301	7051	68288	9219	3945	42,8	8449,7	3264,2	38,6	8 563,1	2 616,7	30,56	8 964,0	2 404,5	26,82	6 894,7	1 604,9	23,28
HOI STRAHOVANIE	100810	53144	6035	15765	6789	43,1	15901,9	5162,5	32,5	11 122,6	5 284,7	47,51	11 341,2	4 575,0	40,34	10 931,1	3 109,4	28,45
UNIVERS	55291	15169	24918	7521	2748	36,5	7728,8	1704,3	22,1	5 250,4	1751,6	33,36	4 474,0	964,0	21,55	2528	477	18,87

Fig. 1. Insurance data (taken from [5]).



I am ALFA STRAHOVANIE
My SCORE is 11
I DOMINATE: 13 points
I am DOMINATED BY: 2 points

OPTIMIZATION MIN	OPTIMIZATION MAX	CRITERION	EPSILON 0 %	MinValue	MaxValue
□	□	Capital 2014, UAH x1000	0	28024	452074
□	☑	Reserves 2014, UAH x1000	0	7051	577126
□	☑	Guaranteed fund 2014, UAH x1000	0	-15450	204005
□	□	Premiums 2014, UAH x1000	0	4487	192990
□	□	Payed claims 2014, UAH x1000	0	2566	82525

Point Info

Last Poined Point	
Object	ALFA STRAHOVANIE
Capital 2014, UAH x1000	190281
Reserves 2014, UAH x1000	99400
Guaranteed fund 2014, UAH x1000	80858
Premiums 2014, UAH x1000	27532
Payed claims 2014, UAH x1000	13034
Payment level 2014, %	47.3

Optimization Point	
Object	PROSTO-STRAHOVANIE
Capital 2014, UAH x1000	138479
Reserves 2014, UAH x1000	79642
Guaranteed fund 2014, UAH x1000	43783
Premiums 2014, UAH x1000	62911
Payed claims 2014, UAH x1000	30336
Payment level 2014, %	48.2
Premiums 2013, UAH x1000	66140.8

Fig. 2. Interface of the system with “insurance data” loaded.

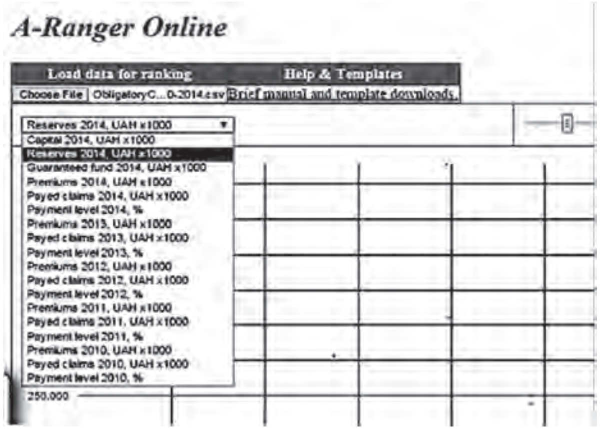
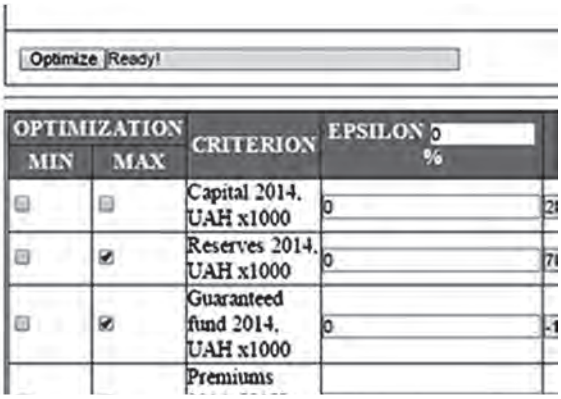


Fig. 3. Selection of a criterion for display



OPTIMIZATION MIN	OPTIMIZATION MAX	CRITERION	EPSILON 0 %
□	□	Capital 2014, UAH x1000	0
□	☑	Reserves 2014, UAH x1000	0
□	☑	Guaranteed fund 2014, UAH x1000	0
		Premiums	

Fig. 4. Selection of criteria to optimize

Conclusions

The paper describes an information technology for multicriteria data mining. The data are presented as a (very) large table, where rows correspond to objects and columns correspond to quantitative and/or qualitative criteria. The problem is to mine Pareto-optimal objects, i. e. to select a subset of nondominated objects with respect to a (changeable) set of criteria. In case of a large table the problem becomes hard for mental solving and needs computer assistance. If only one criterion is chosen, then the problem is reduced to sorting and admits fast solution algorithms. In case of two criteria chosen the problem can be quickly solved by displaying data on the screen and by visual selection of the desired point. If more criteria are involved some information technology is needed to compare, sort, view and analyze data. The proposed technology and computer system suggests interactive visual graphical and table tools for selection of the best data, namely Pareto-optimal data.

References

1. Глибовець М. М. Еволюційні алгоритми : підручник / М. М. Глибовець, Н. М. Гулаєва. – К.: Києво-Могилян. акад., 2013. – 828 с. (= Glybovets M. M. Evolutionary algorithms: textbook / M. M. Glybovets, N. M. Gulaeva. – Kyiv : Kyiv-Mohyla Academy, 2013. – 828 p.)
2. Ляшко С. И. Многомерное ранжирование с помощью эллиптического пилинга / С. И. Ляшко, Д. А. Клюшин, В. В. Алексеенко // Кибернетика и системный анализ. – 2013. – № 4. – С. 29–36. (= Lyashko S. I. Multivariate ranking using elliptical peeling / S. I. Lyashko, D. A. Klyushin, V. V. Alexeyenko // Cybern. Syst. Anal. – 2013. – Vol. 49. – P. 511–516. – Translated from Kibernetika i Sistemnyi Analiz, No. 4, July–August, 2013, pp. 29–36.)
3. Barnett V. The ordering of multivariate data / V. Barnett // J. Royal Statist. Soc. Ser. A (General). – 1976. – Vol. 39. – No. 3. – P. 318–355.
4. Deb K. Multi-objective optimization using evolutionary algorithms / K. Deb. – Chichester: John Willey & Sons, 2001. – 497 p.
5. Internet resource: http://forinsurer.com/stat.
6. Inselberg A. Parallel Coordinates. Visual Multidimensional Geometry and Its Applications / A. Inselberg. – Springer, 2009. – 554 p.
7. Köksalan M. Multiple Criteria Decision Making: From Early History to the 21st Century / M. Köksalan, J. Wallenius, and S. Zionts. – World Scientific, 2011. – 212 p.
8. Kohonen T. Self-Organizing Maps / T. Kohonen. – Berlin, Heidelberg: Springer-Verlag, 3rd edition, 2001. – 501 p.
9. Korhonen P. Visualization in the multiple objective decision-making framework / P. Korhonen, J. Wallenius // Multiobjective Optimization / Eds. J. Branke, K. Deb, K. Miettinen, R. Słowiński // Lecture Notes in Computer Science. – Vol. 5252. – Berlin, Heidelberg: Springer-Verlag, 2008. – P. 195–212.
10. Lotov A. V. Interactive Decision Maps: Approximation and Visualization of Pareto Frontier / A. V. Lotov, V. A. Bushenkov, and G. K. Kamenev. – Norwell, MA: Kluwer Academic Publishers, 2004.
11. Multiobjective Optimization. Interactive and Evolutionary Approaches / Eds. J. Branke, K. Deb, K. Miettinen, R. Słowiński // Lecture Notes in Computer Science. – Vol. 5252. – Berlin, Heidelberg: Springer-Verlag, 2008. – 470 p.
12. Multiple criteria optimization: State of the art annotated bibliographic surveys / Eds. M. Ehrgott, X. Gandibleux. – New York ; Boston ; Dordrecht ; London ; Moscow : Kluwer Academic Publishers, 2003. – 496 p.
13. Norkin B. V. On the approximation of vector optimization problems / B. V. Norkin // Кибернетика и вычисл. техника. – 2015. – Вып. 179. – С. 35–42. (= Norkin B. V. On the approximation of vector optimization problems / B. V. Norkin // Kibernetika i vychislitelnaya tehnika. – 2015. – Issue 179. – P. 35–42.)
14. Norkin B.V. Random Search of Multicriterion Optimum in Insurance / B. V. Norkin // Theoretical and Applied Aspects of Cybernetics. Proceedings of the 4-th International Scientific Conference of Students and Young Scientists. – Kyiv : Bukrek, 2014. – P. 176–187.
15. Norkin B. V. Statistical approximation of multicriteria stochastic optimization problems / B. V. Norkin // Доповіді НАНУ. 2015. – № 4. – С. 35–41. (= Norkin B. V. Statistical approximation of multicriteria stochastic optimization problems / B. V. Norkin // Dopovidi NANU. 2015. – № 4. – P. 35–41.)
16. Saaty T. L. Relative Measurement and Its Generalization in Decision Making. Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors. The Analytic Hierarchy/Network Process / T. L. Saaty // RACSAM (Rev. R. Acad. Cien. Serie A. Mat.) – 2008. – Vol. 102 (2). – P. 251–318.
17. Sobol I. M. Vybor optimalnyh parametrov v zadachah so mnogimi kriteriyami (= Selection of optimal parameters in problems with multiple criteria) / I. M. Sobol, R. B. Statnikov. – 2-nd ed, revised and supplemented. – Moscow : Drofa, 2006. – 176 p.
18. Zuo Y. General notions of statistical depth function / Y. Zuo, R. Serfling // Annals Statict. – 2000. – Vol. 28. – P. 461–482.

Норкін Б. В.

ВІЗУАЛІЗОВАНИЙ ЕВОЛЮЦІЙНИЙ ПОШУК ПАРЕТО-ОПТИМАЛЬНИХ ДАНИХ

У статті описується інформаційна технологія (та програмне забезпечення) для інтерактивного візуального пошуку Парето-оптимальних даних у великому наборі даних (точок даних). Кожен елемент даних (точка) є вектором із набором компонентів зі значеннями в повністю впорядкованих, можливо різних, просторах або множинах. Ці компоненти розглядаються як критерії оптимізації, які можуть бути максимізовані або мінімізовані. Основна проблема полягає у визначенні недомінуємої підмножини даних щодо обраних критеріїв/компонентів з заданими напрямками оптимізації. Задача вирішується в інтерактивному режимі за допомогою графічного відображення даних у різних площинах (парах координат). Друга проблема полягає у впорядкуванні даних по відношенню до їх сили домінування. Остання задача вирішується шляхом розрахунку двох чисел, кількості елементів, які домінуються даним елементом, і кількості елементів, яких домінує даний елемент, обчислення їх різниці та відображенням різних розмірів точок даних на дисплеї.

Ключові слова: мультикритеріальна оптимізація, Парето-оптимальність, еволюційний пошук, системи підтримки прийняття рішень.

Матеріал надійшов 27.10.2016