HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Boris Mirkin, Michael Orlov

RESEARCH IMPACT: LEVEL OF RESULTS, CITATION, MERIT

Working Paper WP7/2014/09
Series WP7
Mathematical methods
for decision making in economics,
business and politics

Moscow
2014

УДК 001.89
ББК 73
М74

Mirkin, B., Orlov, M.

М74 Research impact: Level of results, citation, merit : Working paper WP7/2014/09 [Text] / Boris Mirkin, Michael Orlov ; National Research University Higher School of Economics. – Moscow : Higher School of Economics Publ. House, 2014. – 40 p. – 70 copies.

This report elaborates on an approach to measuring of the level of research results recently proposed by one of the co-authors. The approach involves a taxonomy of the research domain, that is, a hierarchy representing the domain's structure. The level of results is evaluated according to the taxonomy ranks of the subjects that have emerged or have been crucially transformed due to the results by the scientist under consideration. We also consider two more conventional approaches for scoring the research impact over (a) citation metrics and (b) merit metrics. To aggregate individual criteria in these approaches, we use an in-house automated criteria-weighting method oriented towards as tight a representation of the strata as possible. To compare – and combine – the three approaches empirically, we use a sample of publicly available data of scientists in the areas of data analysis and machine learning. As the domain's taxonomy, we use a corresponding part of the ACM Computing Classification System 2012 and slightly modify it to better reflect results by the scientists in our sample. The obtained ABC stratifications concur with intuition. Besides, they are rather far from each other. This supports the view that all the three approaches (citations, merits, taxonomic rank) should be considered as different aspects, and, therefore, a good method for scoring research impact should involve all the three.

УДК 001.89
ББК 73

Keywords: research impact, level of results, citation level, merit level, automatic weighting, taxonomy of the domain

Mirkin Boris – Department of Data Analysis and Machine Intelligence, National Research University Higher School of Economics, Moscow, RF; Department of Computer Science and Information Systems, Birkbeck University of London, UK.

Orlov Michael – Department of Data Analysis and Machine Intelligence, National Research University Higher School of Economics, Moscow, RF.

Препринты Национального исследовательского университета «Высшая школа экономики» размещаются по адресу: http://www.hse.ru/org/hse/wp

Contents

Acknowledgment This work was partially supported by the International Laboratory of Decision Choice and Analysis as part of a project within the Program for Fundamental Research of the National Research University Higher School of Economics Moscow.

1 Introduction: The problem and background

The issue of measuring research impact is attracting intense attention of scientists because metrics of research impact are being widely used by various managing bodies and by public at large as easy-to-get shortcuts for judging of comparative strengths among scientists, research centers, and universities. The citation index and such its derivatives as Hirsch index are produced by a number of organizations including the inventors, currently named Thomson Reuters [32], and Google. These indexes are used sometimes in evaluation and management in sciences, although this is subject to the ongoing debate because of over-simplifications that are immanent to bibliometrics [3]. There have been a number of proposals to amend the indexes, say, by using less extensive characteristics, such as centrality indexes in the intercitation graphs [6] or by following only citations in the "lead scientists" work [5]. Other proposals deny the usefulness of bibliometrics altogether; some propose even such drastic measures as the "careful socialization and selection of scholars, supplemented by periodic self-evaluations and awards" [26], that is, moving back to the closed orders of medieval monk-scientists. Other, more practical systems, such as the UK Research Assessment Exercise (RAE, recently rebranded as REF) intends to assess most significant contributions only, and in a much informal way, which seems a better option. Yet there have been criticisms of the RAE-like systems as well: on the one hand, in the absence of a citation index, the peer reviews do not manifest any consistency in evaluations [11], and, on the other hand, in the long run, the system itself seems a bit too short-sighted; it has cut off everything which is out of the mainstream [18]. Therefore, a recent initiative by a group of influential scientists DORA [30], while rejecting the bibliometrics as the only assessment source, proposes to switch from counting publications only, to checking for the whole list of scientific production including data sets, patents, and codes among others. The U.S. National Science Foundation already modified its instructions so that the outputs of scientific research include products rather than just publications [30]. Although there are less tangible dimensions of impact, that can be important in evaluations, this goes in line with what Alfred Nobel, the founder of the most prestigious science prize, has expressed in his will: the prize goes to those who "have conferred the greatest benefit on mankind" which is further detailed, say for physics, as " have made the most important discovery or invention within the field of physics" [23].

We adhere to this direction of thought. This paper is an attempt at exploring aspects of the concept of larger than papers researcher's productivity. Looking from a practical side, one can recognize that currently there are at least four types of products of scientific research:

1. producing novel scientific results to be described in papers and monographs;

2. participating in the organization of sciences such as being a journal editor or running a research conference;

3. transferring knowledge to and training of younger generations such as undergraduate and postgraduate students;

4. developing technology innovations including patents and other industry related products.

They all should be counted in as parts of the impact by a scientist.

Therefore we are going to explore how these can be reasonably measured and aggregated to derive a reasonable measure of research impact. We recognize the difficulties in measuring the last item, of technical innovations, for the currently living scientists because not so many of them ever get patents. To justfully abandon this item we restrict ourselves with university based researchers only, since academics normally are not required to get a practical use of their research results in engineering.

Another issue is in finding a direct measure to score the research results, item 1, which is so remarkably avoided in current mainstream efforts by using bibliometrics instead. Here we are going to employ the idea of using a hierarchical taxonomy of a research field for mapping research results in the field to those subjects that have been created or drastically revised in the light of these results. The ranks of the receiving nodes define the rank of the research results [19].

Another innovation reported in this paper is in the way of combining multiple criteria. Most popular approaches to multicriteria ranking rely on weighted combinations of criteria in such a way that the weights are defined either manually or in a supervised manner. For example, the former applies to computing university league tables, and the latter is characteristic for defining ABC-classifications of inventory items. Automatically deriving the weights have been pursued as well, mostly in the format of the eigenvector corresponding to the maximum eigenvalue for a similarity-between-criteria matrix such, as RankClus [31] and PCA [20]. This approach is much relevant when the criteria are well correlated so that a better entity over one criterion would be better over most other criteria. If, however, criteria are essentially inconsistent at different entities, the first eigenvector would take into account too little of the data scatter and, therefore, may appear somewhat inappropriate. We develop an approach which we think is adequate at both correlated and inconsistent criteria. According to our approach, the issue is to be solved by finding such a direction in the criteria space that all the entities are projected into compact well-separated clusters on the direction so that the orthogonal hyperplanes may be considered as boundaries between different multicriterial strata of entities. This approach was introduced and substantiated recently in [21, 24].

One more innovation described here is a case of practical implementation of our approaches. To be specific, we focus on the field of Computer Science related to data analysis, machine learning, cluster analysis and data mining. As a relevant taxonomy of the domain we take relevant parts of the ACM Computing Classification System 2012 [2]. We pick up a sample of 30 leading scientists in the field such that the information of their research results is publicly available. We consider three sets of criteria for research contributions: (a) one comprises three Google citation criteria, (b) the second, criteria for items of merit, 2 and 3 from the list above, and (c) the third utilizes adjusted ranks of research results within the taxonomy.

Our preliminary hypothesis is that the aggregate scales of both (i) citation and (ii) merit relate to popularity of scientists rather than anything else. Therefore, the combined scales for (i) and (ii) should have a rather high positive correlation between them. On the other hand, the level of results has no straightforward relation to popularity - the latter much depends on the scientist's character, communication skills, and being employed in a good university, whereas the former, on talent and luck. So any reasonable scale of the level of results should have rather low correlation with both the citation index and the merit index. Our computations do show that this is largely true at our data, although the level of correlation between (i) and

(ii) is not that high. To an extent, this observation supports the views expressed in DORA declaration [30]. Also, we may conclude that our method of mapping research results to a taxonomy of the field (MMRRTF) could be considered a good way forward. It does involve a great deal of manual component, of course. However, it is based on an agreed upon taxonomy of the domain and explicitly mapping the results to taxonomy nodes. Therefore, its results are explicitly expressed and admit public discussions of them, which leads to much less inconsistency in the assessments than just mere subjective evaluations by panel members.

The remainder is organized as follows. The next section provides an algorithmic background for our Linstrat method for aggregating criteria in the format of a weighted sum of them [24, 25]. Our method for mapping research results to a taxonomy of the fields is presented there too. The section 3 describes how our sample of scientists has been formed and how scientists' ranks have been defined by adapting an extract from the taxonomy in ACM-CCS [2]. Section 4 presents data related to features of (i) citation and (ii) merit for our sample. Our results in determining stratifications and criteria weights are presented here as well. Section 5 concludes with a summary and future work directions.

2 Methodology

2.1 The problem of stratification

There is general understanding that in the ranking problem one usually looks for an ordered partition in which entities in the same class are considered to be equivalent over a pre-specified set of criteria, rather than for just a linear ordering of the entities. Reasons for this may include a degree of indifference of the decision-makers (as reflected, say, in the concept of ABC ranking in inventories) or a degree of imprecision in the measurement of criteria or both. We refer to a partition, classes of which are linearly ordered by a relation of precedence, as a stratification. Such areas as sociology and mineralogy use this term exactly in this sense to express social inequality in the former and depth/time precedence in the latter.

Consider an example. Table 1 contains normalized food and housing prices for a foreigner in 10 cities [7]. The left part of Fig. 1 presents a three cluster partition found using k-means clustering method with cities Copenhagen, New-York and Peking taken as the initial centers. The right part of Fig. 1 presents a three strata stratification corresponding to the direction of a combined criterion $F = 0.4789 * HousingP + 0.5211 * FoodP$. This combined criterion can be interpreted as a measure of “cost of living” that takes into account the difference in the relative importance of the criteria.

Table 1: Prices of housing and food for a foreigner in ten cities normalized so that the minimum is zero and maximum, the hundred.

City	Housing	Food
Moscow	96.7284	56.0364
London	93.2099	62.4146
Tokyo	100.0000	44.4191
Copenhagen	42.7160	100.0000
New-York	96.7284	38.9522
Peking	59.9383	12.0729
Sydney	34.4444	19.5900
Vancouver	12.9630	10.2506
Johannesburg	0	5.2392
Buenos-Aires	14.1975	0

As expected, clusters consist of similar cities (see Fig. 1 on the left). Those labeled by a square have relatively low prices for both foods and housing. Cluster labeled by a circle is a singleton consisting of just Copenhagen, with a highest food price and moderate housing prices. The cluster of triangles on the right, in contrast, is of highest housing prices and moderate food prices. The strata, on the right side, are organized over a different principle. The first stratum, for example, is not a cluster but rather a Pareto boundary at highest prices. Each of the remaining cities is dominated, over both criteria, by a city from the first stratum. It is formed not according to similarity but rather according to the combined weighted criterion as a set of a higher cost of living. The second stratum is a set of a moderate living cost, and the third, of the lowest living cost in the set.

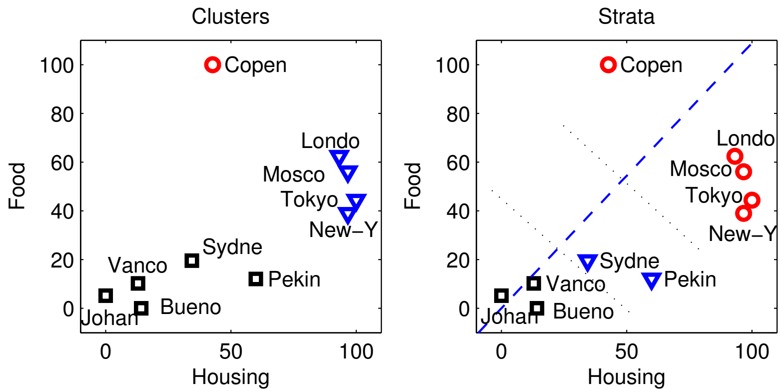

Figure 1: Ten cities over two normalized criteria: Housing price and Food price. They are partitioned in three clusters (on the left) and in three strata (on the right).

One can classify methods for multicriteria stratification according to the extent of the assumed elasticity of the criteria to each other. A constant elasticity e of criterion f_1 towards criterion f_2 would mean that a change of criterion f_2 by a unity is equivalent to the opposite change of f_1 in e units, independently of values of these and other criteria. That is, criteria f_1 and f_2 can be combined into weighted sum $f_1 + ef_2$ in this case. The case of a constant elasticity between all the criteria $f_1(x), f_2(x), \ldots, f_m(x)$ assumes that they can be equivalently substituted by an aggregate criterion $f(x)$ which is expressed as their weighted sum $f(x) = w_1 f_1(x) + w_2 f_2(x) + \ldots + w_m f_m(x)$, where $w_1, w_2, \ldots, w_m$ are non-negative constant weight coefficients summing to 1.

An opposite case is when all the criteria are mutually incomparable and there is no way that a change in one criterion can be equivalently represented by a change in another criterion. That is, each criterion must be taken into consideration whatever the other criteria's values are. The absence of interrelation among criteria leads to the multivariate relation ”better than”, that is ”better over every single criterion”, and the concept of Pareto boundary as the only solution that needs no interrelation between criteria at all.

Yet there is a kind of equivalence between these two extremes: under rather mild mathematical conditions on the criteria and the sets at which they are defined, every x maximizing the combined criterion

$$f(x) = \sum_{t=1}^{m} w_t f_t(x)$$

does belong to the Pareto boundary. And vice versa, any point x belonging to the Pareto boundary is a maximizer of the combined criterion $f(x) = \sum_{t=1}^{m} w_t f_t(x)$ for some x-specific set of weights w (see Fig. 2).

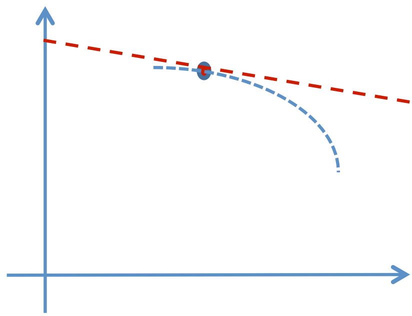

Figure 2: An illustration of the equivalence between two approaches; one of the weighted combined criteria and the other, of the Pareto boundary solutions.

For a detailed review of various interpretations of criteria weight coefficients one may refer to [9]. Much work on multicriterion ranking has been done along the lines of using an external information, say from a Decision Maker, to try to reveal as much information on comparability of criteria at various preference profiles (see, for example, Electre method [14]). Papers [22, 29] develop methods for dividing resources in ABC groups according to their importance for the company by using a criteria weighting system. The groupings are determined by using a combined weighted criterion in which weights are found by solving a linear programming problem. These weights are not constant but depend on the variants being compared.

As we concentrate on the case of a weighted combined criterion with constant weights, we should mention the following. In the real world, there are some applications in which weighted combined criteria are used in such a way that the weights are chosen manually by experts; such are methods applied in composition of university league ranking tables (see, for instance, [10]). In some works, weights are learned in a supervised or semi-supervised manner [17].

2.2 Linstrat criterion and method

We think of our Linstrat method as that inspired by the idea that Pareto boundaries, formed by consecutive "shaving" off the current Pareto boundary from the dataset, can be approximated as strata between parallel hyperplanes whose normal vector, that is, the vector of criteria weights, is taken such that the projections of entities under consideration within each stratum are as close to each other as possible. This idea leads to an optimization problem described below.

Consider a set of N items evaluated over M criteria so that the evaluation scores can be represented as a matrix (x_{ij}), where $i \in 1,...,N$ are the items or actions, $j \in 1,...,M$ criteria, and x_{ij} is the value of jth criterion at the ith item. Assume some criteria weights $w = (w_1, w_2, ..., w_M)$ such that $w_j \geq 0$ at every j and $\sum_j w_j = 1$. These weights are taken into account in the combined criterion $f = \sum_{j=1}^{M} w_j x_j$ where x_j is j-th column of matrix $X = (x_{ij})$. The problem is to divide the itemset in K disjoint subsets $S = \{S_1, ..S_k, ..., S_K\}, k = 1, ..., K$ referred to as strata, according to values of the combined criterion f. Each stratum is characterized by a value of the combined criterion c_k, referred to as the stratum value, or center. These values are ordered so that $c_k > c_l$ whenever $k < l$. That means that any item from k-th stratum is ranked higher, or is more preferable, than any item from stratum l if $k < l$.

Geometrically, strata are formed by layers between parallel hyperplanes in the space of criteria. At any stratum S_k, we assume that the value of the combined criterion $f_i = \sum_{j=1}^{M} w_j x_{ij}$ at any $i \in S_k$ approximates the stratum value c_k as much as possible. That is, in the equation $x_{i1}w_1 + x_{i2}w_2 + ... + x_{iM}w_M = c_k + e_i$, e_i is an error to be minimized over unknown weights w. The problem of finding an optimal vector $w = (w_j)$ can be formulated as the following optimization problem with respect to weights w_j, centers c_k and partition S:

$$
\begin{aligned}
\min_{w,c,S} \quad & \sum_{k=1}^{K} \sum_{i \in S_k} (\sum_{j=1}^{M} x_{ij} w_j - c_k)^2 \\
\text{such that} \quad & \sum_{j=1}^{M} w_j = 1 \\
& w_j \geq 0, j \in 1...M.
\end{aligned} \tag{1}
$$

At any given weight vector w, the criterion in (1) is but the conventional square-error clustering criterion of K-means clustering algorithm over a single feature, that is the combined criterion $f = \sum_{j=1}^{M} w_j x_j$. This implies that finding the optimal stratification S, at a pre-specified K, amounts to finding $K-1$ points dividing the f-axis in K intervals to minimize the within-cluster variance, and the optimal centers c_k are just within-cluster means of f. An optimal stratification over a single feature can be found by using Fisher's dynamic programming clustering algorithm [15]. Therefore, the difficulty in minimization of (1) is concentrated in the task of finding an appropriate w at a given stratification S. If an algorithm for this is specified, then one can proceed in the manner of an alternating minimization algorithm: starting from some weight vector $w(0)$, find optimal S and c. Based on these, find an appropriate weight vector $w(1)$, etc.

At first, we used an evolutionary algorithm for minimizing (1) with respect to w at a given S and c. However, such an algorithm as a whole leads to unstable solutions at some datasets and, moreover, the solutions at times are inferior to those found by using other approaches [21]. A modification based on a direct algorithm for solving the quadratic programming problem is proposed in [24]. It starts from a random w, but leads to a stable solution in most cases. Moreover, in our experiments with synthetic datasets it typically outperforms its competitors by a high margin [24, 25]. Therefore we use this version of Linstrat through the entire material reported in this report.

2.3 Taxonomic rank of a scientist

The concept of taxonomic rank is not uncommon in the sciences. Moreover, it is quite popular in biology as one of its fundamental structures:

”A Taxonomic Rank is the level that an organism is placed within the hierarchical level arrangement of life forms.”, according to a dictionary (see http://carm.org/dictionary-taxonomic-rank). Say, *Eucaryota* is a domain (rank 1) containing *Animals* kingdom (rank 2). The latter contains *Cordata* phylum (rank 3) which contains *Mammals* class (rank 4) which contains *Primates* order (rank 5) which contains *Hominidae* family (rank 6) which contains *Homo* genus (rank 7) which contains, at last, *Homo sapiens* species (rank 8).

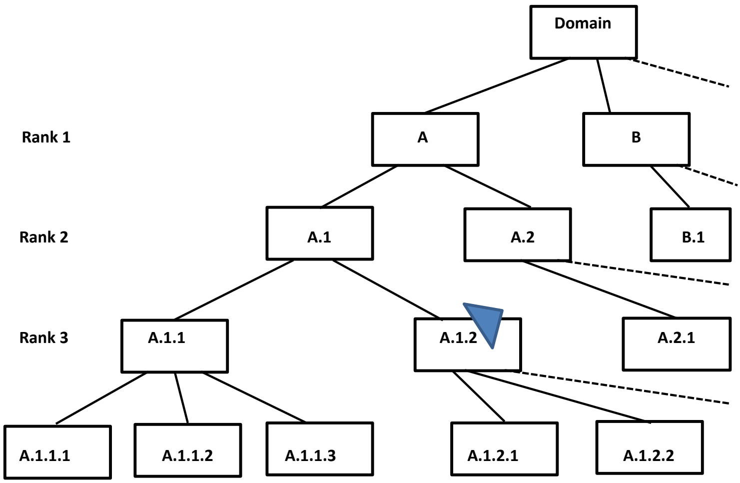

Figure 3: An illustrative taxonomy of a domain. The triangle shows that subdomain A.1.2 has been seriously affected by the results in example.

According to the proposal in [19], the taxonomic rank of a scientist should be defined in a similar way. The relevant science domain should be structured by a hierarchical taxonomy such as that on Fig. 3. The rank of a scientist is defined then as the rank of a subdomain which has appeared because of the scientist’s work or has been substantially transformed because of that. For example, if a domain has been structured as shown on Fig. 3 and a scientist’s work has highly affected the subdomain labelled as A.1.2 (see the triangle indicating that), then their rank would be 3, the number of characters, other than dot, in the code of the subdomain. Of course, this goes in the opposite direction: the higher the rank, the lower the level.

In a practical implementation, when scoring the level of results for a currently living scientist, it is much easier to map their individual papers to the taxonomy rather than the overall achievements. Indeed, the overall achievement is not easy to formulate, whereas an individual paper usually represents a single individual achievement which is not difficult to map to the taxonomy, even if onto two or more subdomains. Together with the plurality of one's results, this leads to the issue of multiple subdomains developed or transformed by a scientist. If the work of a scientist has affected a number of subdomains in a taxonomy, what rank should be assigned to them?

It seems natural that the contribution of an achievement at a lower layer to that of the highest layer achievement is less by an order of magnitude at scoring the taxonomic rank of a scientist. Therefore, of all the levels of the taxonomic hierarchy affected by them first and foremost the highest level is to be used. In the case that only one subdomain is considered as highly affected by the scientist, then their rank is defined as the taxonomy layer to which the subdomains belongs. Such is the case illustrated on Fig. 3 if the subdomain in question is A.1.2, then the scientist's rank is 3. In the case when two or more subdomains on the highest level are affected by a scientist, the rank should be further decreased within the unit interval separating the current rank from the higher one. The scale of the drop should depend on the range of numbers of possibly affected subdomains. In our empirical investigation, we considered, for each of the scientists in our sample, at most five papers leading to ground-breaking discoveries or methods within the taxonomy. Thus, we thought that each additional subdomain of the highest level affected should make a drop in the rank equal to 0.1. Then, an additional drop caused by a node of a lower layer should be about 0.01. For example, if a scientist's results highly affected 4 subdomains of rank 4 and 3 subdomains of rank 3, then the taxonomy rank of the scientist will be 2.76. Indeed, 4 subdomains of rank 4 contribute -0.01 each; one affected subdomain of rank 3 leads to the rank value 3, and each of the two remaining rank 3 subdomains decreases that by 0.1 so that the final rank is $3 - 2 * 0.1 - 4 * 0.01 = 2.76$. To make it simpler, we can assume that additionally 0.1 is subtracted from each of the ranks found – this will not affect the results of the data normalization to 0-100 scale, but the formula for computing the rank gets very simple. To formulate it, let us denote R

the set of nodes assigned to a scientist. Let it be partitioned in subsets R_h, $h \in H$, of the same rank where $H = (h_1, h_1+1, h_1+2, ..., h_1+p)$ where some R_h may be empty. Then the taxonomic rank of the scientist is defined as

$$r = h_1 - \sum_{k=1}^{p} (0.1)^k * |R_{h_1+k}|$$

where $|X|$ is the cardinality of set X.

This method for assigning a scientist their taxonomic rank as a measure of the level of their results suffers of issues of which the following three seem of importance.

First, the method is not automated. The mapping of a research paper to the taxonomy is done manually, so that the result is highly affected by the person(s) performing the mapping. Both the knowledge of the domain and its history as well as the extent of understanding of the result may vary dramatically from person to person. Still, any mapping decision is an explicitly stated judgement which can be discussed openly and corrected if needed. What important is that the subjective part in the decision can be made quite minor. This much differs from the currently available method of peer-reviewing. Indeed, peer-based results can be highly subjective and dependent on various external features such as citation scores [11, 12, 33].

Second, there can be no regular service for updating the taxonomy of the domain, which is the case at many domains. In this case, a ground-breaking paper can be assigned to a wrong sub-domain just because the proper one is not present yet in the taxonomy under consideration or because the taxonomy has not been revamped to fit into the newly acquired evidence. In our assignments reported in the next section this did happen more than once. Because of the presence of the senior co-author whose career spans for the past 50 years, we did not hesitate to expand the taxonomy with updated subdomains if we felt the existing structure was insufficient. Which means that this drawback can be dealt with, at least partly.

Third, and foremost feature is that, unlike in biology, the taxonomies of specific research domains, especially those being under development, cannot be specified exactly because of the changing structure of the domain and, therefore, are subject to much debate. Some popular concepts may be gone after a few years, some new concepts may emerge, some new links can be discovered, whereas some old links may become obsolete. This is especially true for such a dynamic area as computer-related computations and services in which the theoretical thinking is highly affected by the industrial progress. Say, initially computers were oriented at computations, then at data processing, and nowadays, it looks they are oriented mostly at networking. A change in the overall perspective necessarily leads to a drastic change in the taxonomy of the domain. For example, if one compares the current ACM Computing Classification System 2012 [2] with its previous version, the ACM Classification of Computing Subjects 1998 which is available at the same site, one cannot help but notice great differences in both the list of subdomains and the structure of their mutual arrangement. Yes indeed, the current taxonomies of domains can be not well structured and, thus, unstable. However, the appreciation of the level of results goes in line with the taxonomic structure of the domain. The more important is a subdomain currently, the more important deemed ground-breaking results of it. Indeed, unlike the levels of citations, the recognition of the relative importance of this or that subdomain is subject to change. This just shows that the domain taxonomy cannot be considered stable while the domain is being developed, so is the level of results.

3 Empirical testing base for the taxonomic rank evaluation

To test our method empirically, we need, first of all, to take a sample of scientists working in the same domain and score their contributions. The following steps should suffice:

1. Specify a knowledge domain
2. Take its appropriate taxonomy
3. Collect a representative sample of scientists with results in the domain
4. For each of the scientists in the sample, map their ground-breaking results to the taxonomy

5. Compute the taxonomic rank of each of the scientists in the sample

Further on we describe our work on implementation of these steps.

3.1 A taxonomy of data analysis subjects

For an empirical evaluation, we decided to focus on the domain of intelligent data analysis including what is referred to as machine learning and data mining areas. We know some of its history and the current state. We feel that our expertise in other domains is even worse. As to the taxonomy of the domain, we tried first the taxonomy from textbook [20], then from textbook [16] – both cover rather basic subjects only, and it remains entirely unclear at which places in them real-world research results should be mapped to. In this aspect, the ACM CCM 2012 taxonomy has proved to give a better guidance. Parts of ACM CCS 2012 related to the domain under consideration can be considered as composed of its branches presented in Table 2.

Table 2: ACM Computing Classification System (ACM CCS) 2012 high rank subjects related to data analysis, machine learning, and data mining.

Subject index	Subject name
1.	Theory of computation
1.1.	Theory and algorithms for application domains
2.	Mathematics of computing
2.1.	Probability and statistics
3.	Information systems
3.1.	Data management systems
3.2.	Information systems applications
3.3.	World Wide Web
3.4.	Information retrieval
4.	Human-centered computing
4.1.	Visualization
5.	Computing methodologies
5.1.	Artificial intelligence
5.2.	Machine learning

This part extended by finer concepts from ACM CCS 2012 and slightly updated is presented in Table 3. Parts of the hierarchy not affected by the mapping of research results are minimized. The update refers to adding items not covered in the taxonomy, yet concerning outsanding results by scientists from our sample. These concern, as a rule, only leaves of the tree, as can be seen in Table 3. This table represents that part of the taxonomy which has been used for mapping there outstanding results by scientists from our sample. The subdomains (taxonomy nodes) affected by these results are marked by one or two stars. A one star node refers to a subdomain from ACM-CCS 2012; a two star node refers to a subdomain added by us.

Table 3: ACM CCS 2012 based taxonomy of data analysis, machine learning and data mining. A star marks a taxon that has been seriously affected by a scientist from our sample. Two stars mark taxa added by the authors.

Index	Subject name
1.	Theory of computation
1.1.	Theory and algorithms for applications
1.1.1.	Machine learning theory
1.1.1.1.	Sample complexity
1.1.1.2.	Boolean function learning
1.1.1.3.*	Unsupervised learning and clustering
1.1.1.4.	Kernel methods
1.1.1.4.1.	Support vector machines
1.1.1.4.2.	Gaussian processes
1.1.1.4.3.**	Modelling
1.1.1.5.	Boosting
1.1.1.6.*	Bayesian analysis
1.1.1.7.-12.	...
2.	Mathematics of computing
2.1.	Probability and statistics
2.1.1.	Probabilistic representations
2.1.1.1.	Bayesian networks
2.1.1.2.*	Markov networks
2.1.1.3.-8.	...
2.1.1.8.1.	Kernel density estimators
2.1.1.8.2.	Spline models
2.1.1.8.3.*	Bayesian nonparametric models

Index	Subject name
2.1.2.	Probabilistic inference problems
2.1.2.1. - 3.6.	. . .
2.1.3.7.	Kalman filters and HMMs
2.1.3.7.1**	Factorial HMM
2.1.3.8. - 5.3.	. . .
2.1.5.3.1.*	Robust regression
2.1.5.4. - 10.	. . .
2.1.6. - 2.1.9.	. . .
3.	Information systems
3.1.	Data management systems
3.1.1.	Database design and models
3.1.1.1. - 5.	. . .
3.1.1.5.2.*	Data streams
3.1.1.5.3. - 7.	. . .
3.1.2.	Data structures
3.1.2.1.	Data access methods
3.1.2.1.1.*	Multidimensional range search
3.1.2.1.2.- 5.	. . .
3.1.2.2. - 5.9.	. . .
3.2.	Information systems applications
3.2.1.	Data mining
3.2.1.1.	Data cleaning
3.2.1.2.	Collaborative filtering
3.2.1.2.1**	Item-based
3.2.1.2.2**	Scalable
3.2.1.3.*	Association rules
3.2.1.3.1**	Types of association rules
3.2.1.3.2**	Interestingness
3.2.1.3.3**	Parallel computation
3.2.1.4.	Clustering
3.2.1.4.1**	Massive data clustering
3.2.1.4.2**	Consensus clustering
3.2.1.4.3**	Fuzzy clustering
3.2.1.4.4**	Additive clustering
3.2.1.4.5**	Feature weight clustering
3.2.1.4.6**	Conceptual clustering
3.2.1.4.7**	Biclustering
3.2.1.5.	Nearest-neighbor search

Index	Subject name
3.2.1.6.*	Data stream mining
3.2.1.7**	Graph mining
3.2.1.7.1**	Graph partitioning
3.2.1.7.2**	Frequent graph mining
3.2.1.7.3**	Graph based conceptual clustering
3.2.1.7.4**	Anomaly detection
3.2.1.7.5**	Critical nodes detection
3.2.1.8.**	Process mining
3.2.1.11**	Text mining
3.2.1.11.1**	Text categorization
3.2.1.11.2**	Key-phrase indexing
3.2.1.10.**	Data mining tools
3.2.1.9**	Sequence mining
3.2.1.9.1.**	Rule and pattern discovery
3.2.1.9.2.**	Trajectory clustering
3.2.1.9.3**	Market graph
3.2.1.12**	Formal concept analysis
3.3.	World Wide Web
3.3.1.	Web mining
3.3.1.1. - 5.	...
3.3.1.6**	Knowledge discovery
3.4.	Information retrieval
3.4.1.	Document representation
3.4.1.1. - 5.	...
3.4.1.6.*	Ontologies
3.4.1.7.	Dictionaries
3.4.1.8.	Thesauri
3.4.2. - 3.	...
3.4.4.	Retrieval models and ranking
3.4.4.1.*	Rank aggregation
3.4.4.2. - 4.	...
3.4.4.5.*	Learning to rank
3.4.4.6. - 7.3.	...
4.	Human-centered computing
4.1.	Visualization
4.1.2.	Visualization techniques
4.1.2.1. - 6.	...
4.1.2.7**	Elastic maps

Index	Subject name
4.1.3.	Visualization application domains
4.1.3.1.-4.	...
4.1.4. - 7.	...
5.	Computing methodologies
5.1.	Artificial intelligence
5.1.1.	Natural language processing
5.1.1.2. - 7.	...
5.1.1.7.1**	Wikipedia based semantics
5.1.1.8.	Phonology / morphology
5.1.1.9.	Language resources
5.1.2.	Knowledge representation and reasoning
5.1.2.1.- 3.	...
5.1.2.4.*	Probabilistic reasoning
5.1.2.5. -12.	...
5.1.3.	Computer vision
5.1.3.1.	Computer vision problems
5.1.3.1.1.	Interest point and salient region detections
5.1.3.1.2.	Image segmentation
5.1.3.1.3. - 10.	...
5.1.3.2.	Computer vision representations
5.1.3.2.1.	Image representations
5.1.3.2.1.1*	2D PCA
5.1.3.2.2.	Shape representations
5.1.3.2.3.	Appearance and texture representations
5.1.3.2.4.	Hierarchical representations
5.2.	Machine learning
5.2.1.	Learning paradigms
5.2.1.1.	Supervised learning
5.2.1.1.1.*	Ranking
5.2.1.1.2.	Learning to rank
5.2.1.1.3.*	Supervised learning by classification
5.2.1.1.4. - 6.	Supervised learning by regression
5.2.1.2.	Unsupervised learning
5.2.1.2.1.*	Cluster analysis
5.2.1.2.2.*	Anomaly detection
5.2.1.2.3.*	Mixture modeling
5.2.1.2.4.	Topic modeling
5.2.1.2.5.	Source separation

Index	Subject name
5.2.1.2.6.	Motif discovery
5.2.1.2.7.*	Dimensionality reduction and manifold learning
5.2.1.2.7.1**	Graph embedding
5.2.1.2.7.2**	Supervised dimesionality reduction
5.2.1.3. - 2.6.	...
5.2.2.7.*	Semi-supervised learning settings
5.2.2.7.1.**	Kernel approach
5.2.3.	Machine learning approaches
5.2.3.1.	Classification and regression trees
5.2.3.1.1**	Parallel implementation
5.2.3.1.2**	Splittting criteria
5.2.3.1.3**	Model trees
5.2.3.2.	Kernel methods
5.2.3.2.1.**	Kernel support vector machines
5.2.3.2.1.1**	Dynamic kernel SVM
5.2.3.2.2.	Gaussian processes
5.2.3.2.3**	Kernel matrix
5.2.3.2.4**	Kernel independent components
5.2.3.2.5**	Kernel-based clustering
5.2.3.3.	Neural networks
5.2.3.3.1**	Self-organized map
5.2.3.3.2**	Training approaches
5.2.3.3.2.1**	Evolutionary approach
5.2.3.3.3**	Representation
5.2.3.3.3.1**	Rule-based netwok archirtecture
5.2.3.3.3.2**	Fuzzy representation
5.2.3.3.4**	Evolving NN
5.2.3.3.5**	Ensembling
5.2.3.4.	Logical and relational learning
5.2.3.4.1.	Inductive logic learning
5.2.3.4.2.	Statistical relational learning
5.2.3.5.*	Learning in probabilistic graphical models
5.2.3.5.1.*	Maximum likelihood modeling
5.2.3.5.2.	Maximum entropy modeling
5.2.3.5.3.	Maximum a posteriori modeling
5.2.3.5.4.*	Mixture models
5.2.3.5.5.	Latent variable models
5.2.3.5.6.*	Bayesian network models

Index	Subject name
5.2.3.5.7.**	Markov network models
5.2.3.6.	Learning linear models
5.2.3.6.1.	Perceptron algorithm
5.2.3.6.2**	Linear discriminant analysis
5.2.3.6.2.1**	Tensor representation
5.2.3.7.*	Factorization methods
5.2.3.7.1.*	Non-negative matrix factorization
5.2.3.7.2.	Factor analysis
5.2.3.7.3.	Principal component analysis
5.2.3.7.3.1**	2D PCA
5.2.3.7.3.2**	Sparse PCA
5.2.3.7.4.	Canonical correlation analysis
5.2.3.7.5.*	Latent Dirichlet allocation
5.2.3.7.6.**	Independent component analysis
5.2.3.7.7**	Nonlinear principal components
5.2.3.7.8**	Multidimentional scaling
5.2.3.7.8.1**	Least moduli
5.2.3.8.	Rule learning
5.2.3.8.1.**	Neuro-fuzzy approach
5.2.3.9. - 13.	...
5.2.3.13.1.*	Deep belief networks
5.2.3.14**	Multiresolution
5.2.3.15**	Support vector machines
5.2.4.	Machine learning algorithms
5.2.4.1.	Dynamic programming for Markov decision processes
5.2.4.1.1.- 2.2.	...
5.2.4.2.3.**	Fusion of classifiers
5.2.4.3.	Spectral methods
5.2.4.3.1**	Spectral clustering
5.2.4.4.	Feature selection
5.2.4.5.	Regularization
5.2.4.5.1**	Generalized eigenvalue
5.2.5.	Cross-validation

3.2 Sample of scientists and their taxonomic ranks

In our sampling, we rely on Google citation indexes and try to pick up those with maximum citations. Ideally, we wanted to take about 15-20 scientists from the USA and a couple of scientists from a country such as Australia, Canada, China, France, Germany, Netherlands, Russia, United Kingdom, etc., so that the relative contributions by countries would be reflected in the sample. This also would warrant a variation in citation levels: from many dozen thousands at some of the USA scientists to a very few thousands at those in Europe. This ideal composition, though, was difficult to achieve because for any scientist from the sample we needed data not only on citation and taxonomic rank, but on merit as well. The merit data was not always on display, so that we went as far as to contact those of the sampled scientists for whom the merit data was not easily available, asking them to fill in the slots of the numbers of successful PhDs supervised, journal editing positions, and chairing at conferences. Unfortunately, not all of the addressees replied to our messages, so that we had to remove from the sample those whose merit data were missing. In our final sample there are 30 active scientists in the domain.

Now comes a most controversial part of this project – establishing which areas of the domain have been developed or transformed by this or that scientist from the sample. One of the aspects under fire is crediting somebody for this or that result. Indeed, in the current era of globalization any idea of merit can be traced back to, usually, multiple origins. We accept an easy touch position so that a person is credited with an innovation if this is what they claim themselves, and an important part of the community does support the claim. Another issue is a correct interpretation of the set of main contributions by a person. How can one select the most important items from a few hundred publications? In no way can we claim that our selections have been correct in all the cases; we only hope that did not do much harm because we selected a number of publications, usually from 4 to 6, (co-)authored by each scientist from our sample. Another, even more controversial issue is of choosing subdomains in the taxonomy drastically affected by this or that publication. This is accompanied with a bunch of more-or-less arbitrary decisions starting from deciding was this or that effect drastic indeed and finishing by a decision to add this or that node to the taxonomy. Luckily, the AMS-CCS 2012 is flexible enough to admit different interpretations of the same term. For example, "Clustering" appears in it as part of 1.1 Theory and algorithms for application domains, as well as part of 3.2. Information systems applications, as well as part of 5.2. Machine learning. This allows to properly choose a location within the taxonomy for both algorithms, systems and applications.

All in all, our main argument for the usefulness of our approach is a clear visibility of the entire argument from a piece of work (paper) to formulation of a result to mapping that to a specific (set of) node(s). This gives to anybody an opportunity to operationally discuss and correct, if needed, any part of the picture. The only issue preventing us from presenting all the detail of the dataset and its mapping to the taxonomy is that the project involves scientists' names. We think that there is a kind of an implicit universal non-disclosure agreement making it inconvenient to collect a dataset about peer scientists for publicly ranking them without their consent or even their knowledge of that. The only exception from this "agreement" that can be admitted here are the names of Dr. Panos Pardalos and Dr. Boris Mirkin. There are two reasons for that. First, each of the two did want to be included into the sample, whatever the ranks found. Second, this disclosure makes an evidence that our data relate to real, not imaginary, scientists. Therefore, we report here that P. Pardalos is labeled as S19 and Boris Mirkin as S5, in our sample.

Table 4: Mapping main research results to the taxonomy: nodes affected; taxonomic ranks Tr; taxonomic ranks normalized Trn; three strata.

#	Mapping to taxonomy	Layers	Tr	Trn	Stratum
S1	4.1.2.7, 5.2.1.2.7, 5.2.3.7.7	4,5,5	3,88	73	1
S2	2.1.1.2, 2.1.1.2, 5.2.2.7, 5.2.3.5, 5.2.3.5	4,4,4,4,4	3.50	100	1
S3	3.2.1.4.2, 5.2.1.2.3, 5.2.1.2.7, 5.2.3.5.4 , 5.2.3.7.6	5,5,5,5,5	4.50	29	2
S4	1.1.1.4.3, 3.4.4.5, 5.2.1.1.1, 5.2.1.2.7, 5.2.3.2.1,5.2.3.7.8	5,4,5,5,5,5	3.90	71	1
S5	3.2.1.4.4, 3.2.1.4.4, 3.2.1.4.5, 3.2.1.4.6, 3.2.1.11.1	5,5,5,5,5	4.50	29	2

#	Mapping to taxonomy	Layers	Tr	Trn	Stratum
S6	3.1.1.5.2, 3.1.2.1.1, 3.1.2.1.1 , 3.2.1.6., 3.2.1.7	5,5,5,4,4	3.77	81	1
S7	5.2.3.5.6, 5.2.3.5.7	5,5	4.80	7	3
S8	3.2.1.3.1, 3.2.1.4.1, 5.2.3.3.1, 5.1.3.2.1, 5.1.3.2.4	5,5,5,5,5	4.50	29	2
S9	5.2.1.2.3, 5.2.3.3.2, 5.2.3.5.1, 5.2.3.5.4, 5.2.3.6.2	5,5,5,5,5	4.50	29	2
S10	5.2.3.3.2, 5.2.3.13.1	5,5	4.80	7	3
S11	3.2.1.2, 3.2.1.2.1,3.2.1.3.3, 3.2.1.4.1, 3.2.1.7.2	4,5,5,5,5	3.86	74	1
S12	3.2.1.9.1.1,3.2.1.10,3.2.1.11.2, 5.1.1.7.1,5.2.3.1.3,5.2.3.4.1	6,4,5,5,5,5	3.86	74	1
S13	1.1.1.3, 5.2.1.2.1,5.2.1.2.1, 5.2.2.7.1,5.2.3.7.1	4,5,5,5,5	3.86	74	1
S14	3.2.1.3.1	5	4.90	0	3
S15	5.2.4.3.1	5	4.90	0	3
S16	5.2.4.2.3	5	4.90	0	3
S17	2.1.3.7.1, 5.2.4.3.1, 5.2.3.7.5, 5.2.1.2.4, 5.2.3.2.4, 5.2.3.7.3.2, 5.2.3.5.4., 5.2.4.3.1	5,5,5,5,6,5,5	4.39	36	2
S18	3.2.1.9.1,3.2.1.9.2,5.2.3.3.3.1	5,5,6	4.79	8	3
S19	3.2.1.7.5, 3.2.1.9.3, 5.2.3.2.1.1, 5.2.4.5.1	5,5,6,5	4.69	15	3
S20	3.2.1.4.3,5.2.3.7.7,5.2.3.7.8.1	5,5,6	4.79	8	3
S21	1.1.1.6, 2.1.1.2, 2.1.1.8.3, 3.2.1.6, 3.4.1.6, 5.1.2.4, 5.2.1.1.3	4,4,5,4,4,4,5	3.57	95	1
S22	3.2.1.2.2, 5.2.1.2.7.1, 5.2.3.1.2, 5.2.3.6.2.1	5,6,5,6	4.78	9	3
S23	3.2.1.3, 3.2.1.3.1, 3.4.4.1	4,5,4	3.79	79	1
S24	2.1.5.3.1	5	4.90	0	3
S25	5.2.3.3.3.2, 5.2.3.8.1	6,5	4.89	1	3
S26	3.2.1.11.1, 3.2.1.11.1, 3.3.1.6, 5.2.2.7,5.2.3.5.6	5,5,4,4,5	3.77	81	1
S27	3.2.1.3.2, 3.2.1.4.1, 5.2.1.2.1, 5.2.3.1.1	5,5,5,5	4.60	21	2
S28	3.2.1.8	4	3.90	71	1

#	Mapping to taxonomy	Layers	Tr	Trn	Stratum
S29	5.2.3.3.2.1, 5.2.3.3.3.3, 5.2.3.3.4	6,6,5	4.88	1	3
S30	5.1.3.2.1.1, 5.2.1.2.7.2, 5.2.3.3.5	6,6,5	4.88	1	3

The results of mapping of scientists from our sample to the taxonomy are presented in Table 4. The table also presents the derived taxonomic ranks and the same ranks, 0-100 normalized. The normalization went according to the accepted rule except that the minimum rank, 3.50, gets a 100 mark, and the maximum one, 4.89, gets a 0. By looking at the values of the taxonomic rank, it seems quite obvious that the number of strata should be set to 3, as most values concentrate around 0, 30 and 70 or more. This specifies the number of strata, three, to look for over all the criteria under consideration.

4 Citation and merit lining-up

4.1 Scoring citation and merit

There are a number of engines to score citation indexes of scientists. They are slightly differing over the databases of publications involved or the time periods used in evaluations or some formulaic modifications. Yet there are no verified claims of superiority or inferiority of ones over others. Therefore we limit ourselves with the citation indexes routinely available at Google Scholar. The three metrics readily available for every scientist who has arranged their Scholar Google profile are:

- Number of citations that the scientist has received (Citations);
- Number of their papers received at least 10 citations (#10);
- Hirsch index (H): The number *h* of papers that received at least *h* citations.

Table 5: Statistics for citation metrics: total number of citations, number of papers received 10 or more citations, and Hirsch index; at 2013 (real, on the left), at 2014 (normalized, on the right), gains in 2014 (in the middle).

#	In 2013			Gains, %			Normalized 2014		
	Citations	#10	Hirsch	Citat.	#10	Hirsch	Citat.	#10	Hirsch
S1	5138	101	32	11	6	3	0	8	9
S2	37371	175	78	15	4	4	20	20	46
S3	113240	476	144	14	6	4	68	70	100
S4	70932	292	98	17	15	5	41	40	63
S5	5205	61	31	16	7	3	0	2	8
S6	47844	316	96	15	10	8	27	44	61
S7	38862	299	97	16	44	4	21	41	62
S8	9400	119	46	14	7	2	3	11	20
S9	26630	134	42	18	12	8	14	14	17
S10	92538	239	102	32	4	15	55	31	66
S11	39468	182	73	13	6	6	22	22	42
S12	55831	220	65	16	4	5	32	28	36
S13	14653	104	53	18	12	6	6	9	26
S14	95598	608	122	19	40	7	57	91	82
S15	84127	179	83	25	7	4	50	21	50
S16	12028	86	45	17	10	7	4	6	20
S17	77512	342	116	19	12	9	45	48	77
S18	30009	150	65	14	8	7	16	16	36
S19	26220	402	76	7	7	1	13	58	45
S20	5408	50	21	2	6	-9	0	0	0
S21	24117	121	70	14	7	9	12	12	40
S22	18665	260	70	26	12	11	9	34	40
S23	82781	203	89	10	4	1	49	25	55
S24	164251	280	108	16	10	7	100	38	71
S25	5530	50	29	16	11	7	0	0	7
S26	29334	155	65	11	8	5	15	17	36
S27	54579	661	87	11	23	4	31	100	54
S28	54098	472	111	1	1	0	31	69	73
S29	23773	309	69	16	14	10	12	42	39
S30	14954	179	61	31	20	13	6	21	33

Table 6: Three merit criteria: PDS – number of successful PhDs supervised, CC - number of conferences (co)-chaired, EJ - the number of journals (co)-edited, both real and 0 -100 normalized.

	Merits			Normalized values		
#	PDS	CC	EJ	PDS	CC	EJ
S1	28	5	2	49	6	3
S2	15	12	4	22	16	8
S3	38	24	9	69	31	22
S4	9	5	8	10	6	19
S5	16	21	4	24	27	8
S6	18	6	1	29	8	0
S7	4	0	1	0	0	0
S8	7	19	6	6	25	14
S9	11	5	16	14	6	42
S10	30	36	2	53	47	3
S11	12	7	5	16	9	11
S12	5	20	6	2	26	14
S13	8	7	5	8	9	11
S14	8	11	2	8	14	3
S15	31	3	2	55	4	3
S16	5	1	2	2	1	3
S17	34	2	8	61	3	19
S18	12	6	6	16	8	14
S19	53	77	27	100	100	72
S20	10	2	5	12	3	11
S21	9	7	1	10	9	0
S22	6	18	8	4	23	19
S23	9	9	9	10	12	22
S24	17	3	8	27	4	19
S25	7	7	3	6	9	6
S26	30	30	6	53	39	14
S27	25	28	12	43	36	31
S28	16	29	37	24	38	100
S29	13	28	15	18	36	39
S30	7	16	17	6	21	44

Table 5 contains values of the three criteria in July 2013 as well as the gain values, per cent, showing how much they increased to September 2014. Three columns on the right in Table 5 present criteria values in 2014 normalized so that the minimum is 0 and maximum, 100. Although some empirical proof of stability of the Linstrat stratification method has been described in [24], these two data sets can be used to further check for the stability of the method.

Merit of a scientist is a rather vague concept to represent the level of services to and appreciation of the scientist by the "research community". Of many possible criteria we select those related to the success of the "research school" established by the scientist and the level of recognition of them. Of course, the levels of citations reflect both. Yet here we are going to use measures related to personal efforts made and personal positions taken by a scientist.

The success manifests itself both scientifically and administratively. The former can be measured by the number of successful PhD students (co)-supervised by the scientist. The latter can be measured by the number of research publishing journals at which the scientist has a role. The level of recognition can be measured by the number of conferences at which the scientist has been invited to give a plenary presentation or to participate in organization of. With some adjustment, these three can be expressed, for a scientist, as

- Number of successful PhD students supervised (PDS);
- Number of scientific journals in which they have been chief or associate editor (at any time) or a member of the editorial board currently (EJ);
- Number of conferences at which they have participated as either chair or co-chair or program-chair or keynote-chair or deputy chair or global chair (CC).

These data over our sample of 30 scientists are presented in Table 6.

4.2 Combined criteria and stratifications obtained

Here are the results of the analyzes over the data in Tables 4, 5, 6:

1. Found a 3-strata stratification over three citation features in Table 5. The combined criterion is formed with weights 0.5, for Citations, 0.5, for #10, and 0 for Hirsch over the data at 2014. For the data of 2013, the respective weights are 0.44 (Citations), 0.56 (#10), 0 (Hirsch). Given that the Citations criterion grew by two-digit percentage points from 2013 to 2014 at 90% of the sample while the #10 criterion by only a one-digit per cent value in most cases, the change of the weights between the two criteria from 2013 to 2014 is consistent. The fact that the Hirsch index criterion's weight is 0 in both cases goes in line with the overwhelming critiques the criterion has been exposed to recently, see [3, 26, 30, 33].

2. Found a 3-strata stratification over three merit features in Table 6. The combined criterion is formed with weights 0.22 at PDS, 0.10 at CC, and 0.69 at EJ. The relative weight values are consistent with our intuition based upon the prevailing practice of mantaining a heavy and just submission reviewing process in leading journals.

3. Found a panoramic stratification embracing all the three aspects of the researcher's impact combined: level of results, level of citation, and level of merit. The combined panoramic criterion according to Linstrat is formed by summing those three with the weights 0.80 (Taxonomy rank), 0.04 (Combined citation), and 0.16 (Combined merit), which also corresponds to our intuition.

We summarize these results in Table 7, for the weights, and in Table 8, for the combined criteria and stratifications.

Table 7: Weights of individual criteria in: Citation combined, Merit combined, and Research impact panoramic.

CCitation		CMerit		Panoramic	
Citations	0.5	PDS	0.22	Taxonomic rank	0.80
#10	0.5	CC	0.10	Citation combined	0.04
Hirsch	0.0	EJ	0.69	Merit combined	0.16

Table 8: Stratifications and combined criteria values at the sample of scientists over various sets of criteria.

#t	Cc	Ccn	Mc	Mcn	Trn	P	Pn	Cs	Ms	Ts	Ps
S1	4	5	13	17	73	61	73	3	3	1	1
S2	20	27	12	15	100	84	100	3	3	1	1
S3	69	93	33	41	29	33	39	1	2	2	2
S4	41	55	16	19	71	62	74	2	3	1	1
S5	1	1	13	17	29	26	30	3	3	2	2
S6	35	48	7	9	81	68	81	2	3	1	1
S7	31	42	0	0	7	7	8	2	3	3	3
S8	7	9	13	16	29	26	31	3	3	2	2
S9	14	19	32	40	29	30	36	3	2	2	2
S10	43	58	18	23	7	11	13	2	3	3	3
S11	22	30	12	15	74	63	75	3	3	1	1
S12	30	41	12	15	74	63	76	2	3	1	1
S13	7	10	10	13	74	62	74	3	3	1	1
S14	74	100	5	6	0	5	5	1	3	3	3
S15	36	48	15	18	0	5	5	2	3	3	3
S16	5	7	3	3	0	1	0	3	3	3	3
S17	46	63	27	33	36	37	43	2	2	2	2
S18	16	22	14	17	8	10	11	3	3	3	3
S19	35	48	81	100	15	30	35	2	1	3	2
S20	0	0	10	13	8	8	9	3	3	3	3
S21	12	16	3	4	95	78	93	3	3	1	1
S22	21	29	16	20	9	11	13	3	3	3	3
S23	37	50	18	23	79	70	83	2	3	1	1
S24	69	93	19	24	0	7	8	1	3	3	3
S25	0	0	6	8	1	2	2	3	3	3	3
S26	16	22	25	31	81	71	84	3	2	1	1
S27	65	88	34	42	21	27	32	1	2	2	2
S28	50	68	77	96	71	75	89	2	1	1	1
S29	27	36	34	42	1	9	10	2	2	3	3
S30	13	18	33	41	1	8	9	3	2	3	3

The Table 8 presents both stratifications and combined criteria values at the sample of scientists over various sets of criteria. Specifically, Cc and Ccn are the citation combined criterion values as computed and normalized to 0-100 scale, respectively; Mc and Mcn are the merit combined criterion values as computed and normalized to 0-100 scale, respectively; Trn is the taxonomic rank normalized; P and Pn is the panoramic combined criterion values as computed and normalized to 0-100 scale, respectively. The right part of the table contains three-strata stratifications Cs, Ms, Ts, and Ps over combined criteria in the normalized to 0-100 scale format, Ccn, Mcn, Trn, and Pn, respectively.

To further summarize these results, let us take Pearson correlation coefficients between the four criteria, Cc, Mc,T, and P, as well as Spearman correlation coefficients between the stratification rankings, Cs,Ms,Ts, and Ps. They are presented in Table 9.

Table 9: Pairwise correlation values between both the four criteria, Pearson coefficients, and between the four stratifications, Spearman coefficients.

Criteria				Stratifications			
	Ccn	Mcn	Pn		Cs	Ms	Ps
Tr	-0.12	-0.04	0.99	Ts	-0.12	-0.02	0.98
Cc		0.31	-0.04	Cs		0.25	- 0.10
Mc			0.10	Ms			0.06

As one can see, the three aspects under consideration, Citation, Merit, and Taxonomy rank, are rather uncorrelated pair-wise, which justifies, up to the extent of the representativeness of our sample, the choice for measurement scales of these aspects. Yet the two indirect scales, Citation and Merit, are somewhat positively correlated, probably to that extent at which they both relate to the popularity of a scientist. Of course, the comprehensive Panoramic criterion much correlates with its major constituent, the Taxonomy rank. Especially impressive this correlation is at the stratifications: Ps almost coincides with Ts, differing from Ts by just one scientist's move from stratum 3 to stratum 2.

On the level of individual researchers, S5 and S19, their lot put them into the middle lane, stratum 2, of the Panoramic scale. Yet the trajectories are different. Scientist S5, Boris Mirkin, makes very little on both, Citation and Merit, scales, yet falls in stratum 2 over the Taxonomy. In contrast, scientist S19 is good on both Citation and Merit, especially on the latter, where he is the best of the entire sample and shares the stratum Ms=1 with just one other researcher. He falls within Ps=2 just because the papers that have been published by him on data analysis, although quite fine from the optimality point of view, did not pay much attention to the structure of the data analysis area. It seems rather obvious that with the publication of his more recent results on deriving deep hidden features [13, 28], P. Pardalos will be getting a higher rank at the ACM-CCS taxonomy, which should propel him to much higher scores on that in a very near future.

5 Conclusion

The described is an attempt at taking a more rounded view on the problem of evaluating impact of a researcher than it is considered usually. Rather than dwelling on conventional citation scoring or more recent network related scoring or even somewhat controversial peer-review evaluations, we come up with an idea that the impact cannot be properly evaluated without looking at the meaning and level of the research results obtained by scientists. We realize that the idea is not quite novel philosophically, so to speak. Yet it is quite novel computationally, as we develop an operational approach to implement the idea by mapping the published research results to a taxonomy of the domain and, moreover, we explicitly show how this can be done by presenting an example of such an evaluation. The example concerns the very area at which we conduct our research projects ourselves, the domain of data analysis, data mining, and machine learning. We take a small sample of scientists in this area so that we are able to manually map their research results to a suitable taxonomy, which is an adaptation of the ACM CCS 2012 taxonomy.

We also tackle two other aspects of the impact, citation and merit, by using three operationally defined criteria for each. To combine the criteria, we find such a weighting of them that approximates the Pareto slices with between-hyperplane layers. Although rather unconventional, this approach has been found competitive [21, 24, 25].

Our empirical results are well matching the conventional wisdom, especially in the following:

1) The controversial Hirsh index has disappeared in our project by itself, that is, its automatically derived weight versus two other citation criteria appears to be 0, at least at our sample.

2) When developing a most comprehensive, the Panoramic, criterion, its constituting combined criteria – Citation, Merit, and Level of results – get respective weights 0.04, 0.016, and 0.80, which is far from uniform, yet much consistent with intuition.

3) The three combined criteria are not correlated, except for a small positive correlation value between the Citation and Merit combined, probably because both reflect popularity of a scientist.

These conclusions go in line with the common wisdom, which may make them looking somewhat suspicious. But all the results have been computed from the data without any attempt at trimming them. We make our data available in the Tables 4, 5, 6, and 8 so that everybody could make their own computations. And we can open the identities of the members of our sample if needed indeed.

The results suggest directions for future work. First of all, we would like to further verify our stratification method. For example, quite different but well justified methods, such as those in [4] and [27], should be applied to these data so that one can take a look at how much the results are similar and dissimilar and, depending on that, take next steps in this direction. Another direction would be in extending the empirical research both in getting larger samples and tackling on other research domains. Next, we should try to automate the task of mapping one's research results to the taxonomy. Moreover, we should take a look whether other uncorrelated dimensions for research impact exist and, if yes, what are they and how one could measure them. Making these and similar steps will bring us closer to the final goal of developing a comprehensive measure of research impact.

References

[1] Abramo, G., Cicero, T., DAngelo, C. A. (2013) National peer-review research assessment exercises for the hard sciences can be a complete waste of money: the Italian case. *Scientometrics*, 95(1), 311-324.

[2] The 2012 ACM Computing Classification System, http://www.acm.org/about/class/2012, accessed 17 October 2014.

[3] Albert, B. (2013) Impact factor distortions, *Science*, 340, no. 6134, 787.

[4] Aleskerov, F. T., Chistyakov, V. V., Kalyagin, V. A. (2010) Multiple criteria threshold decision making algorithms, Working paper WP7/2010/02. Moscow: State University HSE. 40 p.

[5] Aragon, A.M. (2013) A measure for the impact of research, *Scientific Reports*, 3, Article number: 1649 doi:10.1038/srep01649.

[6] Bollen, J., Van de Sompel, H., Hagberg, A., Chute, R. (2009) A principal component analysis of 39 scientific impact measures, *PloS one*, 4(6), e6022.

[7] Burgess A., Davies U., Doyle M., Gilbert A, Heine C., Howard C, Jones S., McKelvey D., Potter K., Wright S. (2006) The Economists pocket world in figures: 2007 Edition *The Economist in association with Profile Books Ltd.* London. 2006. 254 p.

[8] Canavan, J., Gillen, A., Shaw, A. (2009) Measuring research impact: developing practical and cost-effective approaches. *Evidence and Policy: A Journal of Research, Debate and Practice*, 5.2, 167-177.

[9] Choo, Eng U., Schoner, B., Wedley W.C. (1999) Interpretation of criteria weights in multicriteria decision making. *Computers and Industrial Engineering*, 37, no. 3, 527-541.

[10] The Complete University League Guide, http://www.thecompleteuniversityguide.co.uk/league-tables/methodology, accessed 25 October 2014.

[11] Eisen, J.A., MacCallum, C.J., Neylon, C (2013) Expert failure: Re-evaluating Research Assessment. *PLoS Biol*, 11(10): e1001677. doi:10.1371/journal.pbio.1001677

[12] Engels, T. C., Goos, P., Dexters, N., Spruyt, E. H. (2013) Group size, h-index, and efficiency in publishing in top journals explain expert panel assessments of research group quality and productivity. *Research Evaluation*, rvt013.

[13] Fan N., Sadeghi, E., Pardalos, P.M. (2014) Robust support vector machines with polyhedral uncertainty of the input data, in *Lecture Notes in Computer Science*, 8426, Springer, 291-305.

[14] Figueira, J. R., Greco, S., Roy, B., Slowinski, R. (2013). An overview of ELECTRE methods and their recent extensions. *Journal of Multi-Criteria Decision Analysis*, 20(1-2), 61-85.

[15] Fisher, W.D. (1969) *Clustering and Aggregation in Economics*. Baltimore: The Johns Hopkins Press.

[16] Han, J., Kamber, M., Jian Pei (2011) *Data Mining: Concepts and Techniques*, Third Edition, Morgan Kaufmann.

[17] Koksalan, M., Mousseau, V., Ozpeynirci, O., Ozpeynirci S.B. (2009) An outranking-based approach for assigning alternatives to ordered classes. *Naval Research Logistics*, 56(1), 74-85.

[18] Lee, F. S., Pham, X., Gu, G. (2013) The UK research assessment exercise and the narrowing of UK economics, *Cambridge Journal of Economics*, 37(4), 693-717.

[19] Mirkin, B. (2013) On the notion of research impact and its measurement, *Control in Large Systems, Special Issue: Scientometry and Experts in Managing Science*, 44, 292–307, Institute of Control Problems, Moscow (in Russian)

[20] Mirkin, B. (2011) *Core Concepts in Data Analysis: Correlation, Summarization, Visualization*, Springer, London.

[21] Mirkin, B., Orlov, M. (2013) *Methods for Multicriteria Stratification and Experimental Comparison of Them*, Preprint WP7/2013/06, Higher School of Economics, Moscow, 31 p. (in Russian)

[22] Ng, W.L. (2007) A simple classifier for multiple criteria ABC analysis, *European Journal of Operational Research*, 177, 344353.

[23] Nobel Prize page, http://nobelprize.org/alfrednobel/will/will-full.html, accessed 16 October 2014.

[24] Orlov, M. (2014) An algorithm for deriving a multicriterion stratification, *Business Informatics*, no. 4 (In Russian).

[25] Orlov, M., Mirkin, B. (2014) A concept of multicriteria stratification: A definition and solution, *Procedia Computer Science*, 31, 273-280.

[26] Osterloh, M., Frey, B. S. (2014) Ranking games, *Evaluation Review*, Sage, 0193841X14524957.

[27] Page, L. , Brin, S. , Motwani, R., Winograd, T. (1999) *The PageRank Citation Ranking: Bringing Order to the Web*, Technical Report. Stanford InfoLab.

[28] Pappu, V., Pardalos, P.M. (2014) High-dimensional data classification, in *Clusters, Orders, and Trees: Methods and Applications*, Springer Series on Optimization and Its Applications, 92, 119-150.

[29] Ramanathan R. (2006) Inventory classification with multiple criteria using weighted linear optimization *Computers and Operations Research*, 33. 695-700.

[30] San Francisco Declaration on Research Assessment (DORA), am.ascb.org/dora/, accessed 16 October 2014.

[31] Sun, Y., Han, J., Zhao, P., Yin, Z., Cheng, H., Wu, T. (2009) RankClus: integrating clustering with ranking for heterogeneous information network analysis, *Proc. EDBT*, 565-576.

[32] Thompson Reuters intellectual property and science, http://ip-science.thomsonreuters.com/, accessed 16 October 2016.

[33] Van Raan, A. F. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491-502.

Миркин, Б. Г., Орлов, М. А.

Научный вклад: уровень результатов, цитирование, заслуги : препринт WP7/2014/09 [Текст] / Б. Г. Миркин, М. А. Орлов ; Нац. исслед. ун-т «Высшая школа экономики». – М. : Изд. дом Высшей школы экономики, 2014. – 40 с. – 70 экз. (на англ. яз.)

Данная работа развивает подход к изменению уровня научных результатов, недавно предложенный одним из авторов. Этот подход использует таксономию предметной области, т.е. иерархию её понятий. Уровень результатов оценивается в соответствии с рангами тех понятий таксономии, которые возникли или были существенно преобразованы благодаря этим результатам. Рассматриваются также два существующих подхода к оценке научного вклада, (а) по уровню цитируемости и (б) по уровню заслуг. Для агрегирования отдельных критериев внутри этих подходов используется разработанный авторами метод автоматического отыскания весовых коэффициентов критериев. Согласно этому методу веса критериев назначаются таким образом, чтобы возникающие страты были как можно более компактными. Для того, чтобы сравнить и использовать все три подхода на реальных данных, мы используем выборку данных, доступных в Интернете, о научных сотрудниках – специалистах в области анализа данных и машинного обучения. В качестве таксономии предметной области мы используем соответствующую часть многоуровневой классификации компьютерных наук, разработанной всемирной Ассоциацией вычислительных машин в 2012 г. и несколько модифицированной нами, чтобы точнее отобразить результаты, полученные сотрудниками из нашей выборки. Полученные стратификации согласуются с интуицией. Более того, они оказались слабо коррелированными, так что рассматриваемые три подхода (цитирование, заслуги и уровень результатов) следует рассматривать в качестве разных аспектов общего понятия научного вклада. Это означает, что любой разумный метод оценки научного вклада должен включать все три аспекта.

Ключевые слова: научный вклад, уровень научных результатов, уровень цитируемости, уровень заслуг, автоматическое определение весов критериев, таксономия предметной области

Миркин Б.Г. – Департамент анализа данных и искусственного интеллекта, Национальный исследовательский университет «Высшая школа экономики», Москва, РФ; Департамент компьютерных наук и информационных систем, Биркбек Колледж, Лондонский Университет, Лондон, СК.

Орлов М.А. – Департамент анализа данных и искусственного интеллекта, Национальный исследовательский университет «Высшая школа экономики», Москва, РФ.

Препринт WP7/2014/09
Серия WP7
Математические методы анализа решений
в экономике, бизнесе и политике

Миркин Б.Г., Орлов М.А.

Научный вклад: уровень результатов, цитирование, заслуги
(на английском языке)

Зав. редакцией оперативного выпуска *А.В. Заиченко*
Технический редактор *Ю.Н. Петрина*

Отпечатано в типографии
Национального исследовательского университета
«Высшая школа экономики» с представленного оригинал-макета

Формат 60×84 1/16. Тираж 70 экз. Уч.-изд. л. 2,4.
Усл . печ. л. 2,3. Заказ № . Изд. № 1911

Национальный исследовательский университет
«Высшая школа экономики»
125319, Москва, Кочновский проезд, 3
Типография Национального исследовательского университета
«Высшая школа экономики»