Performance characteristics describe the structure of descriptors

I V Chervenchuk, V I Potapov, V V Makarov, O P Shafeeva, M S Doroshenko and A S Gritsay

Omsk State Technical University, 11, Mira ave., Omsk, 644050, Russia

Abstract. Monitoring systems, as systems for observing the state of objects, usually operate with data of some common nature, which can be successfully described further proposed hierarchical model. The initial data of observations, that is, the data of a certain subject area, which in the future presuppose the conduct of statistical studies over them, as a rule, are homogeneous in nature and, in most cases, are described in terms of descriptors and objects. On this basis, it is possible to build a fairly universal hierarchical model of data storage, focused on their subsequent statistical processing, which, in turn, can serve as an information core for building a specialized information system. The paper proposes a method of evaluating the effectiveness of descriptions of the structures having a natural hierarchy presents a study of the proposed features on the subject of production optimization problems.

1. Introduction
In the description of complex multi-level structures, which has a natural hierarchy, such as classification [3, 4], the communicative structure [8], administrative structures, and the like, the most acceptable is the use of graph structures such as a tree. With such a description excludes unnecessary repetition and redundancy.

2. Formulation of the problem
The aim is to find characteristics, quantitatively determining the winning of the hierarchical approach compared to conventional linear organization in the description of the structures, has a natural hierarchy to explore the data characteristics on the subject of production optimization problems.

3. Theory
We will assume that we are given a certain complex object with a hierarchical structure, therefore, a tree of objects is also specified, which must be described by the user-specialist through the proposed interface. The task is to construct for the given object the most optimal tree of indices or to evaluate any structure of descriptors, where it is proposed to take the characteristics of compactness and high-speed performance of descriptors [2, 5] as criteria for optimality; we have a two-criteria optimization problem.

In this case, by compactness we mean the characteristic

\[ \lambda(T_X) = \frac{l(T_X) + n}{m + n}, \]

where \( n = |X| \) - number of indices (of the hanging vertices of the metric tree \( T_X \)), \( m \) - the number of internal vertices of the tree \( T_X \), \( l(T_X) \) - total length of the tree description \( T_X \), \( k = m + n \).
Compactness of a tree of indices shows how many times the computer memory costs are reduced when describing the semantics of the system indicators due to the hierarchical structuring of this description.

The value

$$\tau = \frac{1+n}{2} \cdot \frac{n}{l(T_x)}$$

we will call further prior high-speed performance of search of an index. This characteristic shows the gain from hierarchical structuring when searching for an object [2].

These criteria it is necessary to maximize, that is, to construct such a tree of indices (find the structure of descriptors) for which these characteristics would be as much as possible.

However, the criteria of compactness and high-speed performance of search of an index are obviously contradictory, since the compactness of $\lambda$ increases with increasing $l(T_x)$, and the performance $\tau$ decreases.

4. Experimental results
Consider the set $\Theta(k)$ of all admissible indexes structures (that is, the set of all root trees) with a given total number of vertices $k = m + n$.

Consider the case when the total number of vertices $k$ and the number of hanging vertices $n$, that is, the number of descriptors are rigidly set.

In this case according to (1) and (2) if we are by means of some transformation moving from a structure $T_x^0$ to a structure $T_x'$ to increase the value by one criterion (by changing the value of the lengths of a tree of indices), we will necessarily lose by another criterion, and losing by one criterion we will necessarily win in another way. It is possible to give examples of different structures having the same numerical values of compactness and high-speed performance for constant $k$, it is necessary and sufficient for this that the structures have the same $l$ and $n$.

Consider the set of all effective structures $P(T_x,k)$, where by effective we will understand the Pareto optimality by the criteria of high-speed performance and compactness. Clearly in this case any structure is effective, i.e. $[\Theta(k) | n = n_0 = const] = P(T_x,k,n)$.

Thus, we have obtained that, in the case of given $k$ and $n$, any structure of the exponents is effective, i.e., there is no such structure $T_x'$, that would be inferior to another structure $T_x^0$ in terms of the compactness and high-speed performance characteristics simultaneously; as well as there is no such structure $T_x'$, that would outperform some structure $T_x^0$ at the same time in terms of compactness and high-speed performance.

Consider the case where only the fixedly set number of common vertices $k$, and the number of dangling vertex $n$ (such as, respectively, and internal), freely variable parameter. This is possible when the classification of descriptors is not definitively defined and work is in progress to determine it, or when comparing a number of different indices structures based on performance and compactness.

In this case $\Theta(k) \supseteq P(T_x,k)$, with $k > 2$, inclusion will be strict. It will be shown below, how can one verify the affiliation of some structure to a set of effective structures $P(T_x,k)$.

For the following discussion, we introduce a certain class of trees (sub-class of root trees), which we call TR-Tree – trees with a trunk. TR-Tree have a trunk - not a closed sequence of vertices $< d_1, d_2, \ldots, d_r >$, bound ribs; from each such vertex $d_i$ only edges can come out to the hanging vertices or to the vertex $d_{i+1}$. A case is possible when $r = 1$. The set of hanging vertices issuing from the vertex $d_i$ forms a tier TR-Tree (Here, for convenience, we have moved a little from the traditional concept of "tier", we consider that the root has a 0-tier, and not the first.). Examples of TR-Tree are shown on Figures 1.
For convenience of the further presentation, contrary to tradition, we call a leaf as a hanging vertex in combination with an incident arc. We will assume that the lower tier has \(n_1\) leaves, the next (up) \(n_2\) leaves and the uppermost \(n_r\) of the leaves. It is obvious that TR-Tree can be represented by arrays of dimension \(r\) of the form \([n_1, n_2, \ldots, n_r]\), where \(r\) is the number of tiers. The following expressions are valid for TR-Tree

\[n = k - r = \sum_{i=1}^{r} n_i, \quad l(T_x) = \sum_{i=1}^{r} i \cdot n_{r-i+1}\]  \(3\)

It is possible to prove the following expressions (due to limited volume of the article scope of the article, the evidence is not given)

**Statement 1.** For each structure \(T_x^0\), that is not TR-Tree, there is a TR-Tree \(T_x^S\), that overlaps \(T_x^0\) by the criteria for speed and compactness.

**Evidence.** According to (1) and (2) in order for some structure \(T_x'\) overlapped some structure \(T_x^0\) by the speed and compactness at the same time, it is sufficient that the number of pendant vertices of the structure \(T_x'\) be larger than the number of pendant vertices of the structure \(T_x^0\) at a constant length, i.e. the following conditions would be fulfilled:

\[n(T_x') > n(T_x^0), l(T_x') = l(T_x^0).\]  \(4\)

Obviously, such structures exist.

We will take arbitrary root tree \(T_x^0\) that is not a CT-Tree.

We will transform this tree in corresponding to it TR-Tree \(T_x^S\). In \(T_x^0\) we will select a branch of the maximum length, beginning from root peak (if there are several such branches, one can take any), the whole branch except for the hanging vertex and the arc (leaf) incident to it will be assigned to the trunk of the future tree \(T_x^S\). In this case the original tree \(T_x^0\) we can consider as some a generalized TR-Tree from the stem vertices \(d_i\) of which not only leaves originate, but also complex subtrees (having more than one inner vertex).

We carry out the operation of disassembling of such subtrees \(R\), saving \(l(T_x) = l\), which consists in the following:

To tear off a leaf from some difficult subtree with the root \(d_i\),

a) if at the same time peak – the father of this leaf does not become trailing, then to transfer this leaf to a trunk with saving level;

b) if the peak – the father of this leaf becomes trailing, then to transfer this leaf to the first tier.

**Figure 1.** Examples of TR-Tree: a) one-level \([n]\), b) two-level \([n_1,n_2]\), c) three-level \([n_1,n_2,n_3]\).
Figures 2b, c show two successive transformations of $R$ to the structure of Figure 2a, which transform this structure into a TR-Tree.

Figure 2. Convert $R$ – disassemble complex subtrees: a) original tree, b) the first step in disassembly, c) the second step of the disassembly and conversion to TR-tree.

As for a trunk the branch of the maximum length is selected, similar disassembling is always possible. It is obvious that, applying to leaves of difficult subtrees, we will surely transform the initial tree to TR-Tree for the finite number of steps depending on complexity of the subtrees which are a part of the initial tree. At the same time, as we subject to disassembling the difficult subtrees (having more than one internal peak), $R$ -conversion for modification b, the increasing number of trailing peaks will be surely applied. As $R$ -conversion does not change total length of a tree, i.e. (4) is executed, in case of analysis of each difficult subtree the value of criteria of high-speed performance and compactness will increase.

The statement is proved.

According to the proved statement we have that the set of TR-Trees with $k$ peaks includes a set of effective structures $P(T_x, k)$, that allows to narrow considerably a set of search effective on $\tau$ and $\lambda$ structures and to formulate sufficient sign of belonging of some structure to a set $P(T_x, k)$. However the set of ST trees, though much less set of all root trees nevertheless is rather great and its power sharply increases $(2^{k-2}, k > 1)$ with growth of $k$. The following statement gives the conditions allowing to narrow as much as possible a set of search of effective solutions and to formulate sign of belonging of some structure to a set$P(T_x, k)$.

**Statement 2.** In order that some structure $T_x^0$ belonged to a set of effective structure by the criteria of compactness and high-speed performance $P(T_x, k)$, is necessary also enough, that this structure $T_x^0$ was TR-Tree and the following inequalities were executed:

$$r = 1 \text{, or } \sum_{i=1}^{r-2} (r - i) \cdot n_{ri} > (r - 2) \cdot k - r \cdot (r - 3), \ r < \frac{k}{2}.$$  

where $r$ –numbers of tiers in this TR-Tree.

**Evidence.** We will consider structures of TR-Trees with different quantity of tiers $r$, upon transition to TR-Tree with a large number of tiers, the number of trailing peaks will decrease by unit.

1) We will consider structure of fan type – TR-Tree with $r = 1$ – Figure 1 a), this structure has maximum possible $n$ (i.e. $n_1 = k - 1$) and minimum possible $l(T_x) = k - 1$ – i.e. has the maximum on high-speed performance ($\tau$) also belongs $P(T_x, k)$. At the same time, as we have the two-criteria task, the value of compactness $\lambda(T_x^1)$ will be minimum among all structures $T_x \in P(T_x, k)$ (for example, [1]).

2) We will consider structures of two-tier TR-Tree with $r = 2$ – Figure. 1 b). These structures will have a little smaller values ($\tau$) because of smaller value $n$ (here $n_2 = k - 2 = n_1 - 1$). For this purpose, that structures of two-tier TR-Tree $T_{x1}^2$ (where $j$ – number of structure in case of lexicographic streamlining) exceeded structure $T_x^1$ on compactness according to (1) is necessary also
enough (since \(n\) is strictly set), that inequality was observed \(l(T_{2x}^2) > l(T_{x1}^3) + 1\), where unit in the right part compensates reduction of value \(n\) therefore \(n_{22} + 2n_{21} > k - 1\),

where the first index at \(n\) shows number of tiers in TR-Tree \(r\), the second — number of trailing peaks on a \(i\) tier, \(i = 1 \ldots r\). Taking into account that \(n_2 = k - 2 = n_{21} + n_{22}\) we have

\[n_{21} > 2. \tag{6}\]

that is in order that two-level ST tree superimposed fan structure on a compactness index it is necessary that the number of peaks on the lower tier (here will be only trailing) was strict more than two. As \(n_2 > n_1\), that any structure from \(T_{xj}^2\) will not exceed \(T_{x1}^1\) on compactness. The structure \(T_{xj}^2\) with the with the maximum total length will have total length \(l_{max}(T_{xj}^2) = 2 \cdot (k - 2)\), corresponds to structure with \(n_{22} = 0\).

3) We will consider structures of TR-Tree with \(r = 3\) — Figure 1 c). These structures will have smaller values (\(\tau\)) because of smaller value of \(n\) (here \(n_3 = k - 3 = n_2 - 1\)). For this purpose, that structures of TR-Tree \(T_{xj}^3\) exceeded all structures \(T_{xj}^2\) on compactness according to (1) it is necessary, that inequality \(l(T_{xj}^3) > l(T_{xj}^2) + 1\), where unit in the right part compensates reduction of value \(n\), for this case was observed

\[n_{33} + 2n_{32} + 3n_{31} > k - 1, \text{ и т. к. } n_3 = k - 3 = n_{31} + n_{32} + n_{33}\] we have

\[n_{32} + 2n_{31} > k, \tag{7}\]

and \(l_{max}(T_{xj}^3) = 2 \cdot (k - 3)\), corresponds to structure with \(n_{32} = n_{33} = 0\).

So, structures \(T_{xj}^3\), fitting (7) will superimpose structures \(T_{xj}^2\), and furthermore \(T_{x1}^1\) (in view of transitivity “\(>\)”by criterion of compactness. At the same time with building of total length of value of criterion of compactness will decrease (since the denominator (2) grows), it is obvious as according to (7) of structures \(T_{xj}^3\) we will take only those structures which have big total length, than structures \(T_{xj}^2\), and the number of trailing peaks \(T_{xj}^3\) are less than \(T_{xj}^2\), at what such structure from \(T_{xj}^2\) will superimpose what structure from \(T_{xj}^3\), by criterion of compactness. Thus, all structures \(T_{xj}^3\), fitting (6) and \(T_{xj}^3\) fitting (7) are effective.

Similarly arguing for TR-Trees with any number of tiers we will come to the first part of inequalities (5). However growth of an index of compactness with growth of number of tiers of TR-Trees is limited, adding the greatest possible values of total length of trees for each \(r\), we will find a limit of this growth.

\[\lambda_r > \lambda_{r-1},\]

\[r \cdot (k - r) + k - r > (r - 1)(k - r + 1) + k - r + 1,\]

\[r < \frac{k}{2}.\]

The structure with the maximum number of tiers, \(r_{max}\) satisfying to the last inequality at which all trailing peaks are located at the bottom (first) level will have the maximum total length among effective structures, and respectively, at most on compactness. Such structure reminding a broom is given in a Figure 3 a).

Thus, the sufficiency is proved.

For the proof of need we will assume that there is some structure \(T_{xj0}^0 \in P(T_{x}, k)\), and at the same time inequalities (5) are not executed. If \(r0 = 1\), that (5) can be executed; if \(r0 > 1\), that in case of not execution (5) among TR-Trees with \(r = r0 - 1\), there will be a tree with the same total length (or, at least, with total length smaller on unit), but at the same time as such tree will have quantity of
trailing peaks, bigger on unit, then at $T_{xj0}^r$, it will superimpose $T_{xj0}^r$ by compactness (or, at least, to be the equivalent) and to superimpose on high-speed performance. Thus, $T_{xj0}^r$ cannot be effective structure.

The statement is proved.

**Remark.** If to replace signs of a strict inequality (5) with signs of mild inequality we will receive necessary and sufficient conditions for a set of poorly effective structures (optimum according to Slater) by criteria of compactness and high-speed performance.

**Evidence.** Really, if in (5) to replace signs of a strict inequality with signs of mild inequality, then structures meeting such condition with number of tiers of $r$ will include also structure $T_{xj1}^r$ (or a several the equivalent by criteria) the equivalent on compactness with some structure $T_{xj0}^{r-1}$, having the maximum index of compactness for structures with number of tiers of $r - 1$. At the same time among structures $T_{xj}^{r-1}$, obviously, there will be no structures which are strictly superimposing (i.e. on “>”) $T_{xj1}^r$ by criteria of high-speed performance and compactness at the same time.

For the final proof of the remark we will prove that among poorly effective estimates there are no structures, the equivalent on high-speed performance and the having different values of criterion of compactness. Let there are two structures $T_{x1}^{r1}$ with number of trailing peaks of $n_1$ and total length of a tree of $l_1$, and $T_{x2}^{r2}$ with $n_2, l_2$ respectively. According to (2) in order that these structures were equivalent on high-speed performance and had different values of criterion of compactness, it is necessary that $n_2 - n_1 > 0$, $l_2 - l_1 > 0$ or $n_2 - n_1 < 0$, $l_2 - l_1 < 0$. Let's say that $T_{x1}^{r1}$ are effective, then if $n_2 - n_1 > 0$, then $l_2 - l_1 > 0$ therefore, $\lambda_2 > \lambda_1$ what contradicts efficiency $T_{x1}^{r1}$. If $n_2 - n_1 < 0$, $l_2 - l_1 < 0$, therefore, $r2 > r1$, then among structures with number of tiers of $r1$ we will take structure $T_{x3}^{r1}$ with total length $l(T_{x3}^{r1}) = l_2$, such structure exists as minimum possible value of length of a measure description is less for structures with smaller number of tiers (with $r2 = 1$ (5) it is automatically executed). According to (1) and (2) structure $T_{x3}^{r1}$ will superimpose structure $T_{x2}^{r2}$ on high-speed performance and compactness, i.e. $T_{x2}^{r2}$ cannot be poorly effective.

5. The discussion of the results

For convenience of execution of the procedure of check of some structure regarding belonging to effective boundary $P(T_x, k)$ by the accepted criteria of inequality (5) it is possible to expand. From inequalities (5) it is possible to pass to the system of inequalities which in a general view looks as

$$
\begin{cases}
(jn_{r1} + (j - 1)n_{r2} + \cdots + n_{rj}) > (j - 1)k - (j - 2)r; \\
\quad j = 1 \ldots r; \\
r < \frac{k}{2}.
\end{cases}$$

(8)

The system of inequalities (8) allows to realize quickly check regarding belonging of some structure to a set of effective solutions $P(T_x, k)$. If necessary, using these inequalities, it is possible to write easily the procedure allowing to find all effective structures in case of the given $k$ (at the same time also you should not forget that all effective structures belong to a class TR-Trees). In figures 3 a-g the set of $P(T_x, 8)$ is given.
Figure 3. Effective structures for $k = 8$: a) configuration 1, b) configuration 2, c) configuration 3, d) configuration 4, e) configuration 5, f) configuration 6, g) configuration 7.

6. Conclusion

The use of hierarchical structures in the description of objects that have a natural hierarchy is the most appropriate and effective, the characteristics proposed in this article confirm this. In addition, it is possible to identify a number of structures for which such a hierarchical description gives the greatest benefit, to formulate and solve a number of optimization problems.

In view of feature of a structure, use of effective structures in practice is limited, however they, being theoretical result, represent some limit structures in an optimization sense by criteria of compactness and high-speed performance, and can serve as a starting point (a standard for comparing) in case of assessment of arbitrary structures of indices by criteria of compactness and high-speed performance.

Such a hierarchical model descriptions have large scope of application, for example, can be used to describe historical data of electricity load [6, 7].

References

[1] Chao Niu, Yongshun Zhang and Jianrong Guo 2018 Pareto optimal layout of multistatic radar, Signal Processing 142 pp 152–156 https://doi.org/10.1016/j.sigpro.2017.07.017.
[2] Michael James Winter, Masayuki Hyodo, Yang Wu, Norimasa Yoshimoto, Muzamir Bin Hasan and Kanta Matsui 2017 Influences of particle characteristic and compaction degree on the shear response of clinker ash Engineering Geology 230 2017 pp 32–45 https://doi.org/10.1016/j.enggeo.2017.09.019.
[3] Cervenchuk I 2015 Model of statistical monitoring data of tap water Natural and intellectual resources of Siberia 1 pp 187–191
[4] Haoliang Yuan, Xuecong Li, Fangyuan Xu, Yifei Wang, Loi Lei Lai and Yuan Yan Tang, 2018 A collaborative-competitive representation based classifier model Neurocomputing 275 pp 627-635, https://doi.org/10.1016/j.neucom.2017.09.022.
[5] Khamitov R N, Gritsay A S, Tyunkov D A, Sinitsin G E 2017 Day-Ahead Short-Term Forecasting Electricity Load via Approximation IOP Conf. Ser.: Mater. Sci. Eng. 189 https://doi.org/10.1088/1757-899X/189/1/012005
[6] Potapov V I, Gritsay A S, Tyunkov D A and Sinitsin G E 2016 Using neural network for building short term forecast of electricity load of LLC “Omsk energy retail company” Bulletin of the Tomsk Polytechnic University Geo Assets Engineering 327(8) pp 44–51
[7] Gritsay A S, Khamitov R N, Sinitsyn G E, Cervenchuk I V 2016 The study of methods of approximation for solving the problem of short-term forecasting of the daily electricity consumption Vestnik Samgtu. Ser. Tech. science 4(52) pp 91–97
[8] Shpynov S Pozdnichenko N and Gumenuk A 2015 Application of Formal Order Analysis (FOA) to Higher Order Grouping of Bacteria in the Genera Rickettsia and Orientia Microbes and Infection 17 pp 839–844