A solution to the problem of clustered objects compact partitioning

D V Pogrebnoy, Al V Pogrebnoy, O V Deeva, I A Petrukhina

Tomsk Polytechnic University, 30, Lenina ave., Tomsk, 634050, Russia

E-mail: Pogrebnoy@tpu.ru

Abstract. The urgency of the study consists in the fact that an object arrangement topology of a distributed system is often nonuniform. Objects can be placed at different distances from each other, thus forming clusters. That is why solving the problem of compact partitioning into sets containing thousands of objects requires the most effective way to a better use of natural structuring of objects that form clusters. The aim of the study is the development of methods of compact partitioning of sets of objects presented as clusters. The research methods are based on applied theories of sets, theory of compact sets and compact partitions, and linear programming methods with Boolean variables. As a result, the paper offers the method necessary to analyze composition and content of clusters. It also evaluates cluster compactness, which results in the decision to include clusters into the sets of partitions. It addresses the problem of optimizing the rearrangement of objects between compact sets that form clusters, which is based on the criteria of maximizing the total compactness of sets. The problem is formulated in the class of objectives of linear programming methods with Boolean variables. It introduces the example of object rearrangement.

1. Introduction

The development of territorially distributed systems that incorporate hundreds and thousands of objects implies representing the system as a hierarchical structure. At the lower level of such structure, objects are united into separated sets and connected to their centers. These objects could be sensors of distributed control systems [1], monitoring systems [2] and sensor networks motes [3]. The closeness of object arrangement inside the set is evaluated by its compactness [4]. This evaluation can be calculated in two, almost the same, ways. The first way involves defining the coordinates of the center and evaluating the compactness as a sum of distances from the center to the objects of the set. The second way consists in evaluating the compactness as a sum of distances between all pairs of objects in the set.

The task is to divide all objects of the system into separated sets with a minimal sum of their compactness evaluations. This task is referred to as compact partitioning [4]. Approximate algorithms for solving this problem are proposed in [4, 5]. These algorithms work effectively with sizes up to 100 objects. When this number starts to increase we need to apply some additional procedures to reduce the size, which noticeably deteriorates the speed of solving the problem.

An object arrangement in the territory called a topological field is most of the time nonuniform [6]. Areas, where objects are arranged more compactly, are selected in this field. That is why a topological graph which represents a topological field can be presented as a group of relatively compact subsets of vertices called clusters. The awareness of the clusters structure and their parameters can help in forming restrictions for solving the problem of compact partitioning. First of all, it concerns the variety of subsets, as well as the interval of their cardinalities [7, 8].

The article proposes the solution to the problem of compact partitioning in such conditions when a topological graph of a big size consists of clusters, while limitations introduced to the number of partition sets and their cardinality are either just recommendations or not given at all. The resulting partition can be accepted by the system developer either as a final solution, or as information to make some other decisions concerning a structuring procedure.

Let us also note that if there are no restrictions to the number of sets and their cardinality, it is particularly useful to get a row of compact partitions for selecting the one that would fully satisfy the conditions of the project. Achieving such partition deserves a special study.

2. Clusters topology analysis

Solving the problem of compact partitioning will be considered for the case when the number of sets K and the range of their capacities [α, β] are just recommendations. It is also assumed that clusters representing a topological graph were obtained as a result of a continuous increase of the compact sets (*KMs*) cardinality. The incremental increase of cardinality g=2,3,…,β allows forming a *KM*(g) list which includes *KMs*, each containing g vertices. Resulting from the isolation effect [5], the *KM*(g) list is divided into many *KM* sets forming cluster s, where s=1,2,…,S.

A cluster topology that we denote as $T(g)$ describes the structure, content and arrangement of clusters in the topological field. We assume that the process of the *KM's* cardinality increase and formation of the *KM*(g) list were executed with limitations K and [α, β].

Let us consider the situation when $n(g) \le K$. It is obvious that such situation is quite possible, because the cardinality increase process, while growing from g to g=n, leads to $n(g)$=1. If the number of *KMs* in the *KM*(g) list is equal or close to K and value g is within the range of [α, β], then we can consider that the compact partitioning problem has been solved. We just need to rearrange the objects which, at the same time, belong to several *KMs*. This problem will be described in the next section.

The situation when $g \ge \beta$ and $n(g)$>K can also emerge, and it means that object arrangement in the topological field is getting closer to its being uniform and, as a consequence, the clusters isolation effect is weak. In this case, the number of objects owned by different *KMs* is increasing and the problem of object rearrangement between *KMs* is getting harder. The example of such situation is shown in Figure 2.

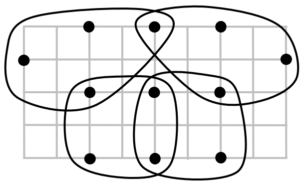

Figure 1. The example of the weak cluster isolation effect.

In the example, proposed $T(g)$ consists of one cluster achieved for g=β=4 and $n(4)$>K. For analyzing the quality of cluster s containing several *KMs*, cluster density evaluation $P(g)_s$ is applied:

$$P(g)_s = (n(g)_s \cdot g - n_s) / n_s , \quad (1)$$

where n_s is the number of objects in cluster s and $n(g)_s$ is the number of *KMs* which form cluster s. Averaged density evaluation $P(g)$ is also used:

$$P(g) = 1/S \sum_{s=1}^{S} P(g)_s . \quad (2)$$

According to (1), density evaluation $P(g) \le 1$. Maximum density $P(g)_s$=1 refers to the situation when cluster s consists of one *KM*. This happens in case two *KM* clusters unite into one in going from

(g–1) to g. Those clusters are recognized in (2) with density $P(g)_s$=1. For the cluster in Figure 1, density $P(4)$ equals (4·4–11)/11=5/11.

While analyzing the structure of clusters in $T(g)$, we should consider the situation when $n(g)=K$, and g are within the range of [α, β] as both a rare and lucky case. The process of the *KM* cardinality increase generally ends when $n(g)$ maximum approximation is reached from up or down to value K. In any case, each cluster s in $T(g)$ is analyzed in terms of being included in partition sets. If a cluster consists only of one *KM*, then it is considered as a set and is excluded from the *KM*(g) list. Clusters of several *KMs* with high density $P(g)_s$, containing $n_s \le \beta$, are also considered as sets. For the residual clusters, the operation of object rearrangement between *KMs* is executed.

3. The problem of object rearrangement between *KMs*

It is assumed that following the analysis of the object structure, each cluster of *KMs* will correspond with the partition set. The analysis influences the objects owned by two or more *KMs*. It is necessary to decide to what *KMs* we will assign these objects in order to maximize the compactness of the resulting sets. Besides, the number of objects in each set should be kept within the range of [α, β].

The example of object rearrangement for the case when topology $T(G)$ has only one cluster is shown in Figure 2. The total number of objects is n_s=33. The recommended conditions of partitioning are K=5, [α, β]=[6, 8]. The cluster consists of five *KMs*. The density of cluster $P(8)$=(5·8–33)/33=7/33 is low, which means there are not many *KM*s intersections.

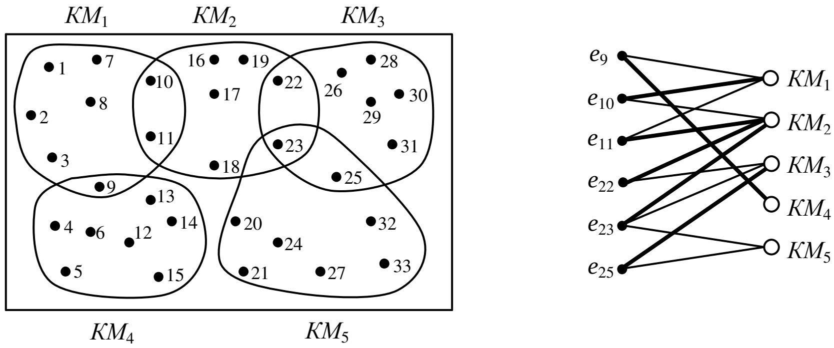

Figure 2. The example of object rearrangement between *KMs*.

Figure 2 shows that objects {e_9, e_{10}, e_{11}, e_{22}, e_{23}, e_{25}} should be rearranged. The solution to this problem can be achieved by conducting an incremental review of each object and assigning it to the closest one among all possible *KMs*. For example, if the distance from e_{23} to the objects in KM_5 is less than to the objects in KM_2 or in KM_3, then we should assign e_{23} to KM_5. While doing this we must not forget that the number of objects in KM_1, KM_2, KM_3 would not become less than α=6.

The given problem can be written as an optimization problem [9]. For this purpose, we introduce variable x_{ij}=1, if object e_i, which should be rearranged, is assigned to KM_j and x_{ij}=0, otherwise. The evaluation of the distance between object e_i and the objects in KM_j is denoted as c_{ij}. Then the problem of object rearrangement can be written as a linear programming problem with Boolean variables [10]:

$$\sum_{i=1}^{n_q}\sum_{j=1}^{m_q} c_{ij}x_{ij} \Rightarrow \min; \tag{3}$$

$$\sum_{j=1}^{m_q} x_{ij} = 1,\ i = 1,2,...n_q; \tag{4}$$

$$\sum_{i=1}^{n_q} x_{ij} \ge s_j,\ j = 1,2,..,m_q. \tag{5}$$

n_q is the number of objects which should be rearranged;
m_q is the number of *KMs* participating in the rearrangement process;

s_j is the minimal number of objects which is necessary to be assigned to KM_j to make it contain more than α objects.

Equations (4) ensure the assignment of each object e_i to one of KM_j. Inequality (5) keeps the number of objects in KM_j within the range of $[\alpha, \beta]$.

Let us note that solving the problem with condition $g>\beta$ requires an additional limitation:

$$\sum_{i=1}^{n_q} x_{ij} \le p_j,\ j=1,2,...,m_q\ . \qquad (6)$$

Here p_j is the maximum number of objects which can be assigned to KM_j to guarantee that it does not have more than β objects.

Limitation (6) implies that when condition $g > \beta$ is enabled, the cardinality of each cluster is $\left|KM_{j_1} \cup KM_{j_2}\right| \le 2\beta$. This ensures compatibility of the (3) – (6) problem. It is not obvious whether the above condition is followed while executing the algorithm of the *KM* cardinality increase. That is why, the question of compatibility of the (3) – (6) problem requires some more research and is not considered in the article. We can also assume that we need to introduce some additions to the problem description for the $g > \beta$ condition.

In Figure 2 on the right, we can see the illustration to the rearrangement problem. Edges (e_i, KM_j) match possible variants of assigning e_i to KM_j and have c_{ij} weight. Values s_j for the example are: s_1=s_3=1, s_2=2, s_4=s_5=0. Bold edges represent the allowed variant of assigning e_i to KM_j. As a result, 33 objects are divided into 5 subsets having 6, 7, 6, 8, 6 cardinality, respectively.

It is obvious, that the alternate solution for the given example even in formulation (3) – (5) is not the best. This is due to the fact that values c_{ij} are calculated relatively to initial KM_j and do not take into account objects rearrangement resulting from the problem solution. The attempts to continuously specify centers in new sets and solve the problem of object rearrangement relative to those centers can improve the compactness of sets, but this will cause the computation problem of the big size.

The computation problem depends on the number of *KMs* and the number of objects which have to be rearranged. The number of sets in practical applications even for thousands of objects is commonly restricted to tens. The number of objects in rearrangement depends on the excess of g over the average value of the range of $[\alpha, \beta]$. The number of such objects is limited, first of all, by excluding duplicate *KMs*. It is also important that the problem is solved for each cluster separately. Some clusters with high density can be included in compact partition sets in certain conditions.

All this eventually decreases the total size of the computation problem. Let us also note that the solution of the (3) – (6) problem is not required for the algorithm. It is possible to continuously rearrange the objects by analyzing each for including it to one of alternative *KMs*, while keeping to the conditions of the problem.

4. Conclusion

The article is devoted to the solution of the problem of compact set partitioning. The set contains hundreds and thousands of objects distributed within the given territory and is formed as a group of clusters. Each cluster consists of one or more intersected compact sets. The article examines the possibility of using nonuniformity of object arrangements in the topological field as a basis for achieving compact partitioning. In addition, the structure and cardinality of sets, relying on parameters of clusters and requirements to compact partitioning parameters, are considered as recommendations.

The article introduces the evaluation of cluster density, which is used for making decisions to include clusters to partition sets. It also proposes the technique used for clusters topology analysis. Besides, the article deals with the problem of object rearrangement between compact sets inside the cluster as a linear programming problem with Boolean variables.

The research showed that it is advisable to use the proposed analysis of the clusters topology while allocating the clusters, thus providing the project developer with different variants of compact partitions. Each variant has different partitioning parameters and describes the topology of nonuniform object arrangement. In this case, the selected variant of partitioning is more complete to comply with the goals and requirements to the system project.

References

[1] Botygin I A , Popov V N , Tartakovsky V A and Sherstnev V S 2015 *Proc. SPIE 21st Int. Symp. Atmospheric and Ocean Optics: Atmospheric Physics* **9680** 1–4
[2] Popov V N, Botygin I A and Koshelev N V 2016 *Key Eng. Mat.* **685** 925–929
[3] Reizlin V I 1981 *Sov. Phis. J.* **24(5)** 401–405
[4] Pogrebnoy V K and Pogrebnoy A V 2004 *In. Proc. 8th Korean-Russia International Symposium on Science and Technology* (*KORUS-2004*) Vol 1(Tomsk: Tomsk Polytechnic University) pp 137–141
[5] Reizlin V I and Nefedova A A 2016 *IOP Conf. Ser.: Mater. Sci. Eng.* **124** 012090
[6] Pogrebnoy Al V and Pogrebnoy An V 2016 *Key Eng. Mat.* **685** 952-956
[7] Tadao Murata 1989 *Proc. of the IEEE* **77(4)** 541-580
[8] Demin A Yu and Dorofeev V A 2014 *12th International Conference on Actual Problems of Electronics Instrument Engineering* (APEIE) (Novosibirsk) pp 624 –627
[9] Natalinova N. M 2016 *Proc. of the XII International Siberian Conference* (Moscow) p 6
[10] Richárd Dévai, Judit Jász, Csaba Nagy, and Rudolf Ferenc 2014 *Acta Cybernetica* **21** 419-437