The Bi-partial Version of the p-median / p-center Facility Location Problem and Some Algorithmic Considerations

Submitted: 15th April 2014; accepted: 20th May 2014

Jan W. Owsinski

DOI: 10.14313/JAMRIS_3-2014/28

Abstract:
The paper introduces the bi-partial version of the well known p-median or p-center facility location problem. The bi-partial approach, developed by the author, primarily to deal with the clustering problems, is shown here to work for a problem that does not possess some of the essential properties, inherent to the bi-partial formulations. It is demonstrated that the classical objective function of the problem can be correctly interpreted in terms of the bi-partial approach, that it possesses the essential properties that are at the core of the bi-partial approach, and, finally, that the general algorithmic precepts of the bi-partial approach can also be applied to this problem. It is proposed that the use of bi-partial approach for similar problems can be beneficial from the point of view of flexibility and interpretation.

Keywords: *facility location, p-median, p-center, clustering, bi-partial approach*

1. Introducing the Bi-partial Approach

The bi-partial approach was developed by the present author at the beginning of the 1980s (see [5], [6]) primarily as a way of dealing with the problems of cluster analysis, its strongest point being the capacity of providing the solution to the clustering problem including the optimum number of clusters, without the need of referring to any external (usually statistical) criteria. The approach has been recently described in a formal manner in Owsiński [7], [8], and its application to some special task in data analysis was provided in Owsiński [9]. Dvoenko [1] applied the approach to the well-known k-means-type procedure.

The approach is based on the use of the bi-partial objective function, which is composed, according to the name, of two terms, which, in very general way, can be subsumed for clustering as representing the inner cohesion of the clusters and the outer separation of the clusters[1]. If cohesion within clusters is measured by some function of distances between the objects, or measurements, or samples, inside individual clusters, denoted $Q_D(P)$, where P is a partition of the set of n objects, indexed $i = 1,...,n$, into clusters A_q, $q = 1,...,p$, and subscript D means that we consider distances inside clusters, then we put as measure of separation of different clusters $Q^S(P)$, meaning a function of similarities of objects in different clusters, and the sum of the two, $Q_D{}^S(P)$, is minimised (possibly small distances inside clusters and possibly small similarities among clusters).

This function, $Q_D{}^S(P)$, has a natural dual, namely $Q_S{}^D(P)$, in which the two components represent, respectively, cohesion within clusters, measured with similarities (proximities) inside the particular clusters, $Q_S(P)$, and distances between different clusters, measured with distances between objects, belonging to different clusters, $Q^D(P)$. The function $Q_S{}^D(P)$ is, of course, maximised.

Even though this concept, at its general level, may appear to be close to trivial, there exist concrete implementations of the two dual objective functions, which form novel and interesting approaches, especially regarding cluster analysis. Moreover, if the components of the objective functions are endowed with definite, quite plausible properties, the approach leads to effective solution algorithms.

2. Problem Formulation

The problem we address here is different from the majority of problems taken as instances of application of the bi-partial approach. Namely, the problem we address is a classical question in operations research, related to location analysis. Not only, though, the interpretation of the problem is quite specific, but also the very form is in a way not appropriate for the treatment through the bi-partial formalism, as introduced here.

We deal, namely, in a very simplistic, but also very general manner, with the following problem

$$\min \Sigma_q(\Sigma_{i\in Aq}\, d(x_i,x^q) + c(q)) \qquad (1)$$

with minimisation being performed over the choice of the set of p points (objects) x_i that are selected as the central or median points x^q, $q = 1,...,p$. For our further considerations it is of no importance whether these points belong to the set X of objects (medians) or not – i.e. they are only required to be the elements of the space E_X (centers), to which all the objects, either actually observed, or potentially existing, belong. It is, however, highly important that the second component of the objective function, namely $\Sigma_q c(q)$, does not involve any notion of distance or proximity.

While $d(.,.)$ is some distance, like in the general formulation of the bi-partial approach, where it en-

ters either $Q_D(P)$ or $Q^D(P)$, $c(q)$ is a non-negative value, interpreted as some cost, related to a facility q. The problem regarding (1) is to find a set of p (q = 1,...,p) locations of facilities, such that the overall cost, composed of the sum of distances between points, assigned to the individual facilities, and these facilities, and the sum of costs, related to these facilities, is minimised. It is, of course, assumed that the costs $c(q)$ and distances $d(.,.)$ are appropriately scaled, in order for the whole to preserve interpretative sense.

The costs may be given in a variety of manners: as equal constants for each arbitrary point from X or from E_X, i.e. c, so that the cost component in (1) is simply equal pc, or as (more realistically) the values, determined for each point separately, i.e. $c(i)$, or as a function, composed of the setup component (say, c_1, if this setup cost is equal for all locations) and the component that is proportional to the number of locations, assigned to the facility q, with the proportionality coefficient equal c_2 (i.e. the cost for a facility is then c_1 + card$A_q c_2$). Of course, more complex, nonlinear cost functions, also with c_1 replaced by $c_1(i)$, can, as well, be (and sometimes are) considered.

This problem has a very rich literature, with special numerical interest in its "pure" form, without the cost component, mainly devoted to mathematical and geometric properties and the respective (approximation) algorithms and their effectiveness. Notwithstanding this abundant tradition, the issues raised and the results obtained, we shall consider here the form of (1) in one of its basic variants.

3. Some Hints at Cluster Analysis

Any Reader with a knowledge in cluster analysis shall immediately recognise the first component of (1) as corresponding to the vast family of the so-called "k-means" algorithms, where such a form is taken as the minimised objective function. Indeed, this fact is the source of numerous studies, linking facility location problems with clustering approaches. One can cite in this context, for instance, the work of Pierre Hansen (e.g. [2]), but most to the point here is the recent proposal from Liao and Guo [3], this proposal explicitly linking k-means with facility location, similarly as this was done several decades ago by Mulvey and Beck [4].

The latter proposal by Liao and Guo [3] is insofar interesting as the facility of realisation of the basic k-means algorithm allows for the relatively straightforward accommodation of additional features of the facility location problem (e.g. definite constraints on facilities and their sets).

Thus, while the first component of the function (1) could be treated with some clustering approaches, e.g. those based on the k-means type of procedure, the issue is in the way the entire function (1) is to be minimised.

4. An Example

For the sake of illustration, we shall consider the problem (1) in the following more concrete, even though very simple, indeed, form:

$$\min_P \Sigma_q(\Sigma_{i\in Aq}\, d(x_i,x^q) + c_1 + c_2\text{card}(A_q)) \qquad (2)$$

where c_1 is the (constant) "facility setup cost", while c_2 is the (constant) unit cost, associated with the servicing of each object $i \in A_q$, except for the "first one", this cost being included in the setup cost. Such a formulation, even if still quite stylised, seems to be fully plausible as an approximation. It can, of course, be transformed to

$$\min\,(\Sigma_q\Sigma_{i\in Aq}\, d(x_i,x^q) + pc_1 + c_2 n), \qquad (2a)$$

where it is obvious that we could deal away with the component, associated with the unit cost c_2. We shall keep it, though, for illustrative purposes, since the part, related to unit costs may, and usually does, take more intricate, nonlinear forms.

The problem (2) can be, quite formally, and with all the obvious reservations, mentioned, anyway, before, moulded into the general bi-partial scheme, i.e.

$$\min_P Q^S{}_D(P) = Q_D(P) + Q^S(P), \qquad (3)$$

where partition P encompasses, in this case, both the composition of A_q, q = 1,...,p, taken together with the number p of facilities, and the location of these facilities, i.e. choice of locations from (say) X as the places for facilities q.

Consider the simple case, shown in Fig. 1, with $d(.,.)$ defined as Manhattan distance, the cost component of (2) being based on the parameter values c_1 = 3, c_2 = 1. Again, these numbers, if appropriately interpreted, can be considered plausible (e.g. distance, corresponding to annual transport cost, and c_1 corresponding to annual write-off value).

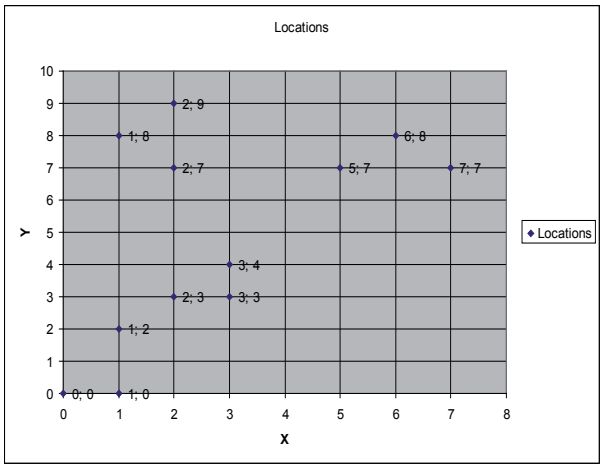

Figure 1. A simple academic example for the facility location problem

Table 1 shows the exemplary values of $Q^S{}_D(P)$ = $Q_D(P)$ + $Q^S(P)$, according to (2), for a series of partitions P. This is a nested set of partitions, i.e. in each consecutive partition in the series one of the subsets of objects, a cluster A_q, is the sum of some of the clusters from the preceding partition, with all the other clusters being preserved. Such a nested sequence of

Table 1. Values of $Q^S_D(P) = Q_D(P) + Q^S(P)$ for a series of partitions, according to (2)

$Q_D(P)$	$Q^S(P)$ – calculation	$Q^S(P)$ - value	$Q^S_D(P)$	Partitions (facility locations in bold)	p
0	12*3+12*1	48	48	All locations are facility locations	12
1	11*3+10*1+1*2	45	46	Merger of (0,0) and (1,0)	11
2	10*3+8*1+2*2	42	44	Merger of (2,3) and (3,3)	10
3	9*3+7*1+2+3	39	42	Addition of (3,4) to (2,3) and (3,3)	9
13	4*3+4*3	24	37	{(0,0), (**1,0**), (1,2)} {(2,3), (**3,3**), (3,4)}, {(5,7), (**6,8**), (7,7)}, {(**1,8**), (2,7), (2,9)}	4
22	3*3+6+3+3	21	43	{(0,0), (1,0), (**1,2**), (2,3), (3,3), (3,4)}, {(5,7), (**6,8**), (7,7)}, {(**1,8**), (2,7), (2,9)}	3
55	1*3+12	15	70	{(0,0), (1,0), (1,2), (2,3), (3,3), (**3,4**), (5,7), (6,8), (7,7), (1,8), (2,7), (2,9)}	1

partitions is characteristic for a very broad family of cluster algorithms – the progressive merger or progressive split algorithms.

The character of results from Table 1, even if close to trivial, is quite telling, and indeed constitutes a repetition of the observations made for other cases, in which the bi-partial approach has been applied. Note that the values of $Q_D(P)$ increase along the series of partitions, while the values of $Q^S(S)$ – decrease, and $Q^S_D(P)$ has a minimum, which, for his simple case, corresponds, indeed, to the solution to the problem.

5. Some Algorithmic Considerations: the Use of the k-means Procedure

As indicated before, the problem lends itself to the k-means-like procedure, which, in general and quite rough terms, at that, takes the following course:

0° Generate p^2 points as initial (facility location) seeds (in this case, the case of p-centers, the points generated belong to X), usually $p << n$

1° Assign to the facility location points all the n points from the set X, based on minimum distance, establishing thereby clusters A_q, $q = 1,\ldots, p$

2° If the stop condition is not fulfilled, determine the representatives (facility locations) for the clusters A_q, otherwise STOP

3° Go to 1°.

This procedure, as we know, converges very quickly, although it can get stuck in a local minimum. Yet, owing to its positive numerical features, it can be restarted from various initial sets of p points many times over, and the minimum values of the objective function obtained indicate the proper solution.

In the here analysed problem of facility location, since such problems rarely are really large in the standard sense of data analysis problems, it is quite feasible to run the k-means procedure, as outlined above, for consecutive values of p in order to check whether a minimum over p can be found for a definite formulation of the facility-location-related $Q^S_D(P)$. Although we shall not be demonstrating this here, in view of the opposite monotonicity of both components of $Q^S_D(P)$ along p, the minimum found over p is a global minimum (although, of course, it is not necessarily the solution to the problem considered, since we deal here only with an approximation of the actual objective function). This procedure can be simplified so as to encompass only a part of the sequence of values of p, starting, say from $p = 2$ upwards, until a minimum is encountered.

6. Algorithmic Considerations Based on the Bi-partial Approach

We shall now present the algorithmic approach that is based on the basic precepts of the bi-partial approach. Assuming, namely, the property that we have observed for the case of the concrete objective function (2), that is – the opposite monotonicity of the two components of the objective function, we can reformulate it, obtaining, in the general case, the following parametric problem:

$$\min_P Q^S_D(P,r) = rQ_D(P) + (1\text{-}r)Q^S(P), \qquad (4)$$

where the parameter $r \in [0,1]$ corresponds to the weights we may attach to the two components of the objective function. Actually, it is used only for algorithmic purposes, and not to express any sort of weight, and we assume that we weigh equally the two components ($r = ½$). Here, we make no a priori assumptions as to the value of p, in distinction from the approach, outlined above, based on the k-means procedure. The form (4) enables the construction of a suboptimisation algorithm, provided the two components of the objective function are endowed with certain properties. We shall outline the construction of this algorithm for the case of the objective function (2).

Thus, the above general form is equivalent, for (2), to the following one:

$$\min_P (r\Sigma_q\Sigma_{i\in Aq} d(x_i,x^q) + (1\text{-}r)\Sigma_q(c_1 + c_2\text{card}A_q)). \quad (5)$$

Now, take the iteration step index, t, starting with $t = 0$. Consider (5) for $r^0 = 1$. We obtain

$$\min_P (1\cdot\Sigma_q\Sigma_{i\in Aq} d(x_i,x^q) + 0\cdot\Sigma_q(c_1 + c_2\text{card}A_q) = \Sigma_q\Sigma_{i\in Aq} d(x_i,x^q)). \quad (6)$$

Since we did not make any assumptions, concerning the value of p, we can easily see that the global minimum for (6) is obtained for $p = n$, i.e. when each object (location) contains a facility (each location constitutes a separate cluster). Denote this particular, extreme partition by P^0. The situation described is illustrated in the first line of Table 1. The value of the original objective function is, therefore, equal $n(c_1 + c_2)$, since the first component disappears, we deal with n facilities, and all $\text{card}A_q = \text{card}A_i$ are equal 1.

Then, we decrease the value of r from $r^0 = 1$ down. At some point, for r^1, the value of the parameter is low enough to make the value of the second component of the objective function, $(1\text{-}r)\Sigma_q(c_1 + c_2\text{card}A_q)$, weigh sufficiently to warrant aggregation of two locations into one cluster, with one facility, serving the two locations. This happens when the following equality holds:

$$Q^S{}_D(P^0,r^1) = Q^S{}_D(P^1,r^1), \quad (7)$$

where P^1 is the partition, which corresponds to the aggregation operation mentioned, the equality from (7) being equivalent, in the case here considered, to

$$r^1\cdot 0 + (1\text{-}r^1)n(c_1 + c_2) = r^1 d(i^*,j^*) + (1\text{-}r^1)(n(c_1 + c_2) - c_1) \quad (8)$$

where i^*,j^* is the pair of locations, for which the value of r^1 is determined. This value, conform to (8) equals

$$r^1(i^*,j^*) = c_1/(d(i^*,j^*) + c_1). \quad (9)$$

This relation is justified by the fact that for each passage from p to p-1, accompanying aggregation, the value of the second component decreases by c_1, while a value of distance, or a more complex function of distances, is added to the first component.

As we look for the highest possible r^1, which follows $r^0 = 1$, it is obvious, that the $d(i^*,j^*)$ we look for must be smallest one among those not yet contained inside the clusters (i.e., for this step – among all distances). In the subsequent steps t we use the equation (7) in its more general form, i.e.

$$Q^S{}_D(P^{t\text{-}1},r^t) = Q^S{}_D(P^t,r^t), \quad (10)$$

and derive from it the expression analogous to (9). In this particular case – which is, anyway, quite similar to several of the implementations of the bi-partial approach for clustering – the equation, analogous to (9) is obtained from (10), meaning that at each step t the minimum of distance is being sought, exactly as in the classical progressive merger procedures, like single link, complete link etc.

The procedure stops when, for the first time, r^t is obtained in the decreasing sequence of r^0, r^1, r^2,..., lower than ½ (the sequence of r^t, if realised until the aggregation of all locations into one cluster, will, of course, end at $t = n$-1). Falling below ½ means, namely, that "on the way" the partition P^t was obtained, which was generated by the algorithm for r = ½, corresponding to the equal weights of the two components of the objective function.

Thus, we deal with a procedure that is entirely analogous to the simple progressive merger algorithms, but has an inherent capacity of indicating the "solution" to the problem, without any reference to an external criterion. We used the quotation marks, when speaking of "solution", because the procedure does not guarantee in any way the actual minimum of (2), since the operations, performed at each step, are limited to aggregation. The experience with other cases shows that a simple search in the neighbourhood of the suboptimal solution found suffices for finding the actual solution, if it differs from the suboptimal one.

7. Some Comments and the Outlook

The illustration, here provided, even though extremely simple, is sufficient to highlight the capacity of the bi-partial approach to deal with the p-median / p-center type of facility location problems. In fact, for (slightly) more complex formulations of the problem, like

$$\min_P \Sigma_q(\Sigma_{i\in Aq} d(x_i,x^q) + c_1(q) + c_2 f(\text{card}(A_q))) \quad (11)$$

i.e. where setup costs are calculated for each potential facility location separately, and $f(.)$ is an increasing concave function, the relation analogous to (10) yields only marginally more intricate procedure, analogous to that based on (9), where for each aggregation the minimum has to be found for the two locations or clusters aggregated.

The issue, worth investigation, which arises there from is: what realistic class of the facility location problems can be dealt with through the bi-partial approach?

Concerning the comparison with the here proposed procedure, based on the classical k-means, the following points must be raised:

– k-means outperform progressive merger procedures for data sets with numerous objects (locations), but not too many dimensions (here: by virtue of definition, either very few, or just two), when storing of the distance matrix and operating on it is heavier than calculating np (much less than n^2) distances at each iteration; in the cases envisaged n would not exceed thousands, and p is expected not to be higher than 100, so that the two types of procedures might be quite comparable;

– there exists a possibility of constructing a hybrid procedure, in which k-means would be performed for a sequence of values of p at the later stages of the bi-partial procedure, with the result of the aggregation, performed by the bi-partial procedure being the starting point for the k-means algorithm;

– given the proposal by Dvoenko [1], there *e*xists also a possibility of imple*m*enting directly the bi-partial

version of k-means, with specially designed form of the two components of the objective function; this, however, would require, indeed, additional studies.

ACKNOWLEDGMENT

This research has been partially supported by the National Centre of Science of the Republic of Poland under Grant No. UMO-2012/05/B/SsT6/03068.

Notes

[1] In some other circumstances the two can be referred to as "precision" and "distinguishability", which brings us quite close, indeed, to the standard oppositions, known from various domains of data analysis, such as "fit" and "generalisation" or "precision" and "recall".

[2] We use the classical name of the k-means algorithm, although the number of clusters, referred to in this name as "k", is denoted in the present paper, conform to the notation adopted in the bi-partial approach, by *p*.

AUTHOR

Jan W. Owsinski – Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01–447 Warszawa, Poland.
E-mail: owsinski@ibspan.waw.pl.

REFERENCES

[1] Dvoenko S., "Meanless k-means as k-meanless clustering with the bi-partial approach". In: *Proceedings of PRIP 2014 Conference*, Minsk, May 2014.

[2] Hansen P., Brimberg J., Urosević D., Mladenović, N., "Solving large p-median clustering problems by primal-dual variable neighbourhood search", *Data Mining and Knowledge Discovery*, vol. 19, 2009, 351–375.

[3] Liao K., Guo D., "A clustering-based approach to the capacitated facility location problem", *Transactions in GIS*, vol. 12, no. 3, 2008, 323–339.

[4] Mulvey J. M., Beck M. P., "Solving capacitated clustering problems", *European Journal of Operational Research*, vol., 18, no. 3, 1984, 339–348. DOI: http://dx.doi.org/10.1016/0377-2217(84)90155-3.

[5] Owsiński J.W., *Regionalization revisited: an explicit optimization approach*, CP-80-26. IIASA, Laxenburg 1980.

[6] Owsiński J.W., "Intuition vs. formalization: local and global criteria of grouping", *Control and Cybernetics*, vol. 10, no. 1–2, 1981, 73–88.

[7] Owsiński, J.W., "The bi-partial approach in clustering and ordering: the model and the algorithms,. *Statistica & Applicazioni*, 2011, Special Issue, 43–59.

[8] Owsiński J. W., "Clustering and ordering via the bi-partial approach: the rationale, the model and some algorithmic considerations". In: J. Pociecha & Reinhold Decker, eds., *Data Analysis Methods and its Applications*, Wydawnictwo C.H. Beck, Warszawa, 2012a, 109–124.

[9] Owsiński J. W., "On the optimal division of an empirical distribution (and some related problems)", *Przegląd Statystyczny*, Special Issue 1, 2012b, 109–122.