Time Reduction for Completion of a Civil Engineering Construction Using Fuzzy Clustering Techniques

Gaetano Bosurgi[1], Federico Carbone[1*], Orazio Pellegrino[1], Giuseppe Sollazzo[1]

Abstract

In the civil engineering field, there are usually unexpected troubles that can cause delays during execution. This situation involves numerous variables (resource number, execution time, costs, working area availability, etc.), mutually dependent, that complicate the definition of the problem analytical model and the related resolution. Consequently, the decision-maker may avoid rational methods to define the activities that could be conveniently modified, relying only on his personal experience or experts' advices. In order to improve this kind of decision from an objective point of view, the authors analysed the operation correction using a data mining technique, called Fuzzy Clustering. This allows the analysts to represent complex real scenarios and classify the various activities according to their influence on the reduction of the total execution time. The proposed procedure provides positive results that are also in compliance with significant operational constraints, such as the control of costs and areas needed by the workers to perform the tasks. Finally, it is possible to increase the input variable number preserving the algorithm simplicity and avoiding lacks of accuracy in the final numerical outcomes.

Keywords

project management, construction, road, scheduling, decision making

[1] Department of Engineering, University of Messina,
Contrada di Dio - Villaggio Sant'Agata, 98166 Messina, Italy
* Corresponding author email: fcarbone@unime.it

1 Introduction

The management of a working phase in the civil engineering field is a very complex process, because of the numerous factors affecting the uncertainty of the final result. The main goal of a building company is the production profit at the completion of the execution. This can be reached only by controlling many influencing elements, usually contrasting each other. Number and typology of the adopted resources (manpower and machines), their organization, available space, material costs, and available time should be all considered with accurate attention.

Since the time schedule defined in the design phase and generally represented by a Gantt (or Pert) chart is the main reference in the execution planning, the initial task of site organization is sufficiently simple. An expert operator, using this information in a proper way, assigns appropriate resource quantities to each operation to complete the working activities in the predefined and expected time (Fatemi Ghomi and Teimouri, 2002; White and Fortune, 2002).

Obviously, managing troubles occur when something disturbs the initial balanced situation. The final consequences could be very counterproductive: this can generate a cost increase, cause delays, require further operations and hazardous conditions for the workers. In these cases, according to the quality of the available information, the construction manager should identify the origin of the deviation. This may be due to variations of some material or manpower costs, reductions of the initial estimate of the machine efficiency, adverse weather conditions, errors in the executive design, etc. Unfortunately, correcting these disadvantageous situations is not an easy task, even if they are properly diagnosed.

In general, when delays occur, it is often practice to increase the number of workers or machines for the following activities. However, this choice can produce consequently critical complications, such as growth of construction costs or congestion of the working area generating unexpected diseconomies and potential hazards for workers' safety (Chretienne and Sourd, 2003; Castro et al., 2008a, 2008b).

It should be also considered that the activities efficiently resulting in a time recovery belong to the critical path, i.e. the

activity succession affecting the total execution time of the working operations. Then, assuming equal resource increases, the best choice is generally the correction and improvement of the activities determining the lowest growth of occupied areas and costs, and the highest time savings.

The scenario complexity could also lead the analyst through two different improper decision paths, producing unsatisfactory results:

- The decisions rely on the analyst's personal experience or his technical judgment ability; this is a rapid method useful in the emergency phase, but it is not much rational.
- The decision-maker adopts very sophisticated analytical methodologies (such as probabilistic analyses or optimization tools) that can be applied on real complex situation after extreme simplifications only (Hardie, 2001; Fatemi Ghomi and Rabani, 2003; Azaron et al., 2005; Mouhoub et al., 2011; Madadi and Iranmanesh, 2012); in this case, the final output is not correct either (Trietsch and Baker, 2012). It is robust, but the boundary conditions are different from the actual ones.

Then, there is a need for a novel objective method that can help the decision-maker and consider all the different variables in the resolution, even if not analytically independent (Arunava and Anand, 2008). For this reason, the clustering techniques are very interesting. In particular, they allow the analysts to perform classifications and indicizations of observed data according to conditions preliminarily fixed by the operator (Duda et al., 2001; Webb, 2002; McLachlan, 2004; Ripley, 2005; Theodoridis et al., 2006; Bosurgi et al., 2010; Pellegrino, 2011; 2012).

In particular, the Fuzzy C-means (FCM) method optimizes the criterion "sum of the squares errors" and it is generally used due to its simplicity and low computational cost. It provides a continuous (or fuzzy) classification that, unlike the hard clustering method, allows an object to belong to several classes with a certain membership degree (or, more improperly, probability) in a more realistic way. For these reasons, this technique has seemed the most suitable to represent an unsupervised classification, with samples consisting of a limited number of observations that can be assigned to one or another class only with a certain membership degree.

In scientific literature there are, of course, numerous methods of classification not suitable for the specific purposes of this research. For example, results from Artificial Neural Networks (ANN) strongly depend on the training database that must be large enough to avoid problems of overfitting (in this case, the network "learns by heart" the training set). The Support Vector Machines (SVM) method only works well when there are only two classes and these are linearly separable. Otherwise, it is necessary to introduce variants of the algorithm (so that it searches the hyperplane in higher dimensionality spaces) that greatly increases the computational costs. The Linear Discriminant Analysis (LDA) is generally used in the object recognition systems as linear classifier, but, more often, it finds application for dimensionality reduction and, therefore, decrease of the complexity of the original data set. Some research, however, did not show a better effectiveness of LDA than the Principal Component Analysis (PCA). Finally the problem of Bayesian classifiers is that their estimate is very complicated because of dependencies among attributes and, even with a few features, it is likely to become computationally unfeasible (Martinez A.M. and Kak, 2001; Abonyi J. and Feil, 2007).

Therefore, the aim of this paper is to identify, using a Fuzzy C-Means based methodology, the activities that can strongly accelerate the execution time during the building phase, in compliance with cost and safety constraints. This assures the achievement of the goal of execution time reduction avoiding the subjectivity of experts' decisions or the analytical difficulties of the actual optimization methodologies (Bosurgi and Trifirò, 2005a; 2005b; 2006; Faghihi et al., 2014).

In the following sections, after technical notices regarding clustering and Fuzzy C-Means methodology and the validation measures, the authors provide the characteristics of the proposed approach with details related to the specific research applications (Section 2). Later, the numerical results are presented (Section 3) and critically discussed (Section 4).

2 Method
2.1 Brief note about clustering

When it is necessary to identify hidden structures in raw data in order to classify them in appropriate ways, the wellknown techniques of clustering are used (Bezdek et al., 1984; Amadore et al., 2014). These data often come from surveys recorded during trials, where each observation includes nmeasured characteristics grouped into a n-dimensional vector (Abonyi and Feil, 2007):

$$x_k = [x_{k1}, x_{k2}, ..., x_{kn}]^T, \quad x_k \in R^n \tag{1}$$

All the N observations could be reported as a matrix N × n:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nn} \end{bmatrix} \tag{2}$$

As is known, a cluster attracts towards itself the elements of similar characteristics relative to other elements closer to other clusters. This property is quantified analytically as the normal distance between the cluster center and the observations that belong to it. In recent years, development of artificial intelligence has led researchers to the application of fuzzy techniques to clustering, so that an observation did not belong to a single cluster, but to all clusters with different degrees ranging from 0 to 1 and under the condition that their sum, referring to all the different clusters, is equal to 1.

The partition matrix N × c, therefore, has the following form:

$$U = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1c} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{N1} & \mu_{N2} & \cdots & \mu_{Nc} \end{bmatrix} \tag{3}$$

where c is the number of clusters and U have to respect the following:

$$\mu_{ik} \in [0,1], \ 1 \le i \le c, \ 1 \le k \le N \tag{4}$$

$$\sum_{i=1}^{c} \mu_{ik} = 1, \ 1 \le k \le N \tag{5}$$

$$0 < \sum_{k=1}^{N} \mu_{ik} < N, \ 1 \le i \le c \tag{6}$$

Then, the fuzzy partitioning space for X can be written through this expression:

$$M = \left\{ U \in R^{c \times N} \middle| \begin{array}{l} \mu_{ik} \in [0,1], \\ \forall i, k; \ \sum_{i=1}^{c} \mu_{ik} = 1, \ \forall k; \ 0 < \sum_{k=1}^{N} \mu_{ik} < N, \ \forall i \end{array} \right\} \tag{7}$$

As already said, in the Eq. (5), the sum of each column is 1 which means that the total membership of each x_k in X equals one.

In the Fuzzy C-means procedure, it is expected to perform the minimization of an objective function, called C-means functional, equal to:

$$J(X;U,V) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m \|x_k - v_i\|_A^2 \tag{8}$$

Where V is the vector containing the centers of the clusters and m is a weighting exponent >1:

$$V = [v_1, v_2, \ldots, v_c], \quad v_i \in R^n \tag{9}$$

The weighting factor $m \in \langle 1, \infty \rangle$ characterizes the fuzzyness of the clusters. When m=1 the membership tends to 0 or 1 implying a crisp partitioning where the vi are just the geometric centroids of the X_i.

The literature do not suggest any convenient value for m, since the best choice could derive from extensive trials, but a good compromise could be a selection in the range 1.5≤m≤3.0 (Bezdek et al., 1984).

The minimization J function is a problem of nonlinear optimization that can be passed through a Picard iteration through the first-order conditions for stationary points of J equation. It is opportune to apply some constraints (Eq. (5)) to J through the Lagrange multipliers and set the gradients $(\bar{J})$ respect to U, V and λ to zero:

$$\bar{J}(X;U,V,\lambda) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m D_{ikA}^2 + \sum_{k=1}^{N} \lambda_k \left(\sum_{i=1}^{c} \mu_{ik} - 1 \right) \tag{10}$$

Where the squared inner-product distance norm is represented by:

$$D_{ikA}^2 = \|x_k - v_i\|_A^2 = (x_k - v_i)^T A (x_k - v_i) \tag{11}$$

If

$$D_{ikA}^2 > 0, \ \forall i, \ k \text{ and } m > 1 \tag{12}$$

Then

$$(U,V) \in M \times R^{n \times c} \tag{13}$$

can minimize J only if:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(D_{ikA} / D_{jkA} \right)^{2/(m-1)}}, \quad 1 \le i \le c; \ 1 \le k \le N \tag{14}$$

and

$$v_i = \frac{\sum_{k=1}^{N} \mu_{ik}^m x_k}{\sum_{k=1}^{N} \mu_{ik}^m}, \quad 1 \le i \le c$$

2.2 The proposed procedure

In practical applications, some detailed analyses of the different activities involved in the building realization allow the operator to evaluate the cost of each resource (manpower, machines, materials), the general costs, and the profit. However, the total economic impact could be figured out only by considering the working time schedule. It represents, in a clear and operative way, the space-time distribution of the various resources used to realize the construction.

In modifying the ordinary state of the activities (for instance, accelerating some tasks), it is significant to estimate the potential advantages related to each one. By the way of example, if the working acceleration is pursued through a manpower increase, the most suitable operations to modify are those with the lowest cost/manpower and time/manpower ratios. Another strategical factor is the occupied area by each operation that increases with the workers' number. The growth of this area should be limited and controlled to not affect the other activities performed in the closest areas.

To handle the analytical problem, the authors defined a specific database including all the variables that influence and drive the manager's decision:

- Workers' number for performing a specific task; the upper limit depends on the fixed max number of workers or the maximum area actually available for the task; this variable influences the realization time;
- Number of machines for performing a specific task; each machine generally involves two workers – driver and assistant; this variable conditions the occupied area and the execution time;

- Activity cost for the company, with details concerning manpower, machines, materials, and general costs;
- Activity duration; at constant working quantity, an increase of resource produces a shorter realization time;
- Required area for the activity; the various zones dedicated to different activities do not have to be overlapped (also considering common space and paths), in order to avoid any interaction that could harm workers' efficiency and safety;
- dC/dMP, i.e. the cost variation for each further worker (Fig. 1,2);
- dA/dMP, i.e. the required area variation for each further worker;
- dD/dMP, i.e. the time variation for each further worker.
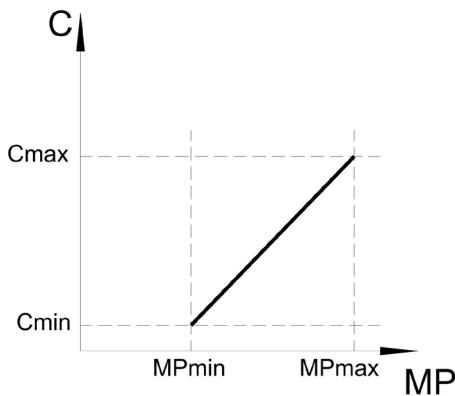
Fig. 1 dC/dMP is the variation of the cost (C) referred to the number of workers (MP). The maximum limits of MPmax are reached when it is not possible to further increase the number of workers (for example, because there is not more available area for the task). Similar laws can be deducted for dA/dMP and dD/dMP.

It is clear that the activities with a low value of dC/dMP should be preferred according to the research goal, because at constant workers' number their cost is lower (Fig. 2). Similar considerations can be stated for dA/dMP. Activities with low dA/dMP assure larger available areas and reduce interference risks between different simultaneous tasks. Finally, duration needs obviously opposite values of dD/dMP; the decision-maker should modify activities with a high duration variation at constant workers' number. Theoretically, the total cost of each activity should not vary if the manpower increases, because even if the daily cost growths it is balanced by the reduction of the days needed for the activity completion. In practice, when the number of workers rises, there is a negative interaction among them, producing a slightly decrease of their efficiency and production, in proportion to the added workers. As previously stated, it is important to modify operation included in a critical path to effectively reduce the working duration.
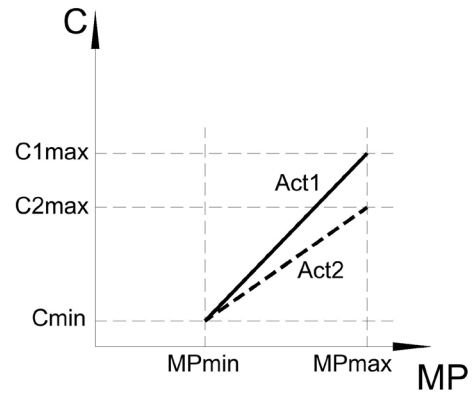
Fig. 2 In this figure it is possible to note the utility of the index dC/dMP when comparing two or more activities. In this case it is possible to see that the activity 2 (Act2) is more convenient than one called Act1 because, for the same number of workers, presents a lower cost. Similar considerations can be made for dA / dMP and dD / dMP.

3 Results

In order to evaluate the method efficiency, the authors applied the procedure on the building of a short road, involving 25 working activities (Table 1). The aim is to evaluate the activities to choose which accelerate in order to produce the maximum advantages for the company in terms of working duration. To actually perform the analysis, it is important to evaluate in detail not only all the factors affecting each single activity (manpower, machines, materials, costs, general expenses, etc.), but also the starting time schedule to properly estimate the resource distribution in time and space (Fig. 3). Although the goal of this application is the reduction of the whole construction time, the authors included in the analysis also the activities out of the critical path to generalize the methodology. In detail, the activities evidenced in Fig. 3 do not produce advantages in terms of total time reduction, but only of cost saving.
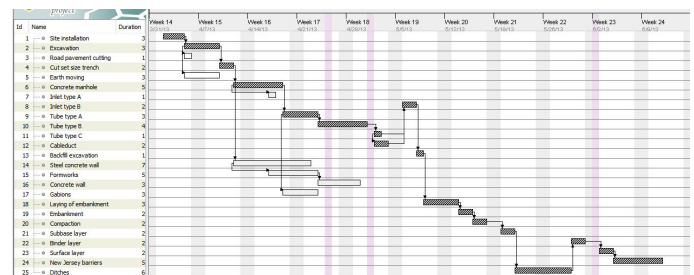
Fig. 3 Scheduling of the 25 activities with a Gantt chart. Tasks inside the critical path are highlighted with a hatch.

Table 1 List and duration of all the activities

ACTIVITY	DURATION (days)
Site installation	3
Excavation	5
Road pavement cutting	3
Cut set size trench	2
Earth moving	5
Concrete manhole	5
Inlet type A	1
Inlet type B	2
Tube type A	5
Tube type B	4
Tube type C	2
Cableduct	4
Backfill excavation	1
Steel concrete wall	7
Formworks	5
Concrete wall	5
Gabions	3
Laying of embankment	3
Embankment	2
Compaction	2
Subbase layer	2
Binder layer	2
Surface layer	2
New Jersey barrier	5
Ditches pose in work	11

The first step of procedure is the division of each activity in its basic or elementary components (use and cost of the adopted resources) in both the conditions of "starting data" (original time schedule) and "modified data" (changing manpower and machines to the maximum possible value), as shown for example in activity 22 (Table 2). The max boundary for manpower and machines is related to the actually available working area that makes unproductive and dangerous the task because of activity interferences.

Some of the common factors presented in the analysis of each activity are considered in the next steps, and they are: number of adopted workers;

- percentual incidence of manpower on the activity cost;
- activity unit cost (€/m³ o €/m²);
- activity duration (days);
- influence area for the activity (m²);
- dC/dMP, dA/dMP, and dD/dMP.

These factors are listed for each of the 25 activities in Table 3, including actual and max limits.

To achieve the research goal, the authors used only 2 clusters for the FCM analysis, representing 2 ideal conditions:
- cluster 1 represents a favourable situation since the activity can be effectively speeded up by increasing the number of workers without interferences with other activities;
- cluster 2 is the opposite condition; it represents a situation in which any modification is impossible because there are not available areas or the max number of workers has already been reached.

The values of the 8 variables for these alternative scenarios can be seen as the multidimensional coordinates of the clusters. For more efficiency, the authors normalized the data by the max value, generating values in the range [0, 1] (Table 4). Using the same approach, data listed in Table 3 are reported in Tab. 5 as standardized values for the starting condition. The result of the clustering procedure are listed in Table 6. For each of the 25 activities, the membership rate for the 2 clusters has been calculated. Moreover, Table 6 provides indications concerning the closer cluster for each activity and the critical path.

In Fig. 4, the authors have plotted the membership to cluster 1 of each activity: the values are represented in a logarithmic scale for more clarity.

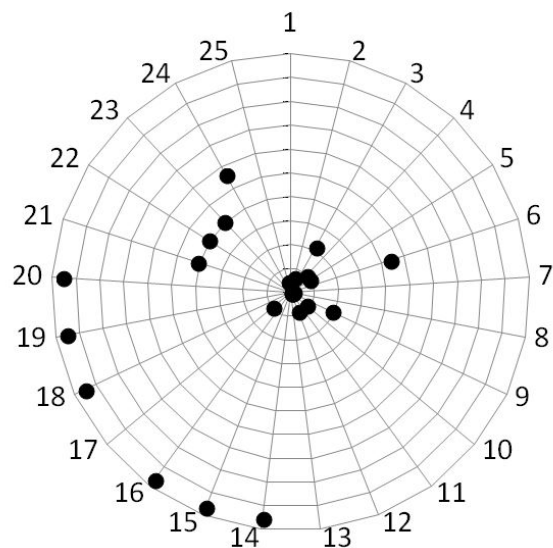**Fig. 4** Distance in log coordinates of the activities respect to the cluster 1 center.

4 Discussion

The analysis of the practical application reported in the previous section allows some general considerations.

Firstly, the size of the data set in Table 3 is not only proportional to the input variables (rows), but also depends on the number of working activities (columns). Considering this

Table 2a Basic components of each single activity
– example for activity 22 Binder layer. Starting Data

STARTING DATA	
Hourly manpower [€/h]	€ 24.00
Hourly excavator cost [€/h]	€ 75.00
Number of workers	2
% incidence manpower	5%
Number of excavator	1
% incidence excavator	8%
Amount of work [mc]	3800.00
Covered area [mq]	100.00
Manpower total cost	€ 259.86
Total cost	€ 5872.73
Received amount	€ 6460.00
Unit cost of manpower	€ 0.07
Unit cost of excavator	€ 0.11
Unit cost of materials	€ 1.19
TOTALE DIRECT COSTS	€ 1.37
Overhead costs (13%)	€ 0.18
TOTAL COSTS	€ 1.55
Profit (10%)	€ 0.15
TOTAL	€ 1.70
Unit duration [h/mc]	0.0014
Working hours in a day	8
Efficiency	3
Duration [days]	2.03

Table 2b – Basic components of each single activity
– example for activity 22 Binder layer. Modified Data

MODIFIED DATA		
Number of workers	4	
% incidence manpower	9%	
Number of excavator	2	
% incidence excavator	14%	
Incremental coeff. for interaction	1.11	
Minimum number of workers	2	
Maximum number of workers	4	
Minimum cost	€	1.55
Maximum cost	€	1.74
Minimum used area	100.00	
Maximum used area	200.00	
Minimum duration	1.02	
Maximum duration	2.03	
Unit cost of manpower	€	0.14
Unit cost of excavator	€	0.21
Unit cost of materials	€	1.19
TOTALE DIRECT COSTS	€	1.54
Overhead costs (13%)	€	0.20
TOTAL COSTS	€	1.74
Profit (10%)	€	0.17
TOTAL	€	1.92
Unit duration [h/mc]	0.001	
Working hours in a day	8	
Efficiency	3	
Duration [days]	1.02	
dC/dMP	0.10	
dA/dMP	98.00	
dD/dMP	0.51	

simple data set, the results show that the method is very satisfying and the related results are remarkable. Then, it is more productive to handle complex building with hundreds of activities through clustering techniques than through alternative methodologies, more fashionable but impractical to adopt. For instance, in the operative research techniques, the problem can be solved only if all the constraints are specifically defined and, in a real situation, they can be very numerous and analytically complex.

Despite the apparent simplicity of the proposed example, this procedure allows the manager to choose different kinds of workers, modify the manpower/machine ration, recombine the workers in the typical teams, change the unit cost of any resource. In general, this approach is very simple and rapid; the only relevant

task for the analyst is the examination of the basic component for each task, referring to unit duration and unit cost (Table 1). However, the project manager actually performs already this operation to verify eventual variations from the original plans during the executive phase and understand their reasons.

The advantages of the methodology are evident considering the goal of the research. The only exam of the data contained in Tables 3 and 5 do not permit to objectively define the relative convenience of all the activities when the manpower is increased, because of the numerous involved variables. Adopting the proposed clustering approach, analyst can classify all the activities and assign them to the 2 different clusters. Furthermore, the activities are not only ordered as convenient

Table 3a Actual and maximum condition of the main indicators of the 25 activities.

Act	No. of Workers		% Manpower		Unitary Cost	
	Actual	**MAX**	**Actual**	**MAX**	**Actual**	**MAX**
1	2	6	35%	49%	€ 13.64	€ 29.15
2	2	4	20%	28%	€ 3.65	€ 5.24
3	2	6	38%	41%	€ 3.44	€ 9.58
4	2	6	25%	36%	€ 4.15	€ 8.64
5	2	6	10%	23%	€ 4.91	€ 6.50
6	2	10	45%	57%	€ 113.64	€ 446.02
7	2	4	22%	32%	€ 111.09	€ 150.81
8	2	4	14%	23%	€ 136.36	€ 167.39
9	2	10	15%	38%	€ 10.36	€ 20.47
10	2	8	17%	37%	€ 12.09	€ 22.11
11	2	4	17%	27%	€ 8.91	€ 11.37
12	2	8	20%	41%	€ 15.91	€ 31.42
13	2	4	29%	36%	€ 2.82	€ 4.59
14	2	14	32%	67%	€ 1.49	€ 4.95
15	2	10	67%	79%	€ 17.91	€ 75.90
16	2	10	7%	24%	€ 139.64	€ 207.24
17	2	6	37%	54%	€ 120.91	€ 247.66
18	2	6	21%	35%	€ 1.41	€ 2.56
19	2	6	7%	17%	€ 14.82	€ 18.84
20	2	6	21%	35%	€ 3.66	€ 6.64
21	2	4	5%	9%	€ 23.00	€ 25.95
22	2	4	5%	9%	€ 1.55	€ 1.74
23	2	4	5%	9%	€ 1.88	€ 2.12
24	2	10	5%	19%	€ 181.82	€ 244.70
25	2	12	18%	42%	€ 100.00	€ 255.63

Table 3b Actual and maximum condition of the main indicators of the 25 activities.

Act	Duration		Area		dC/dMP	dA/dMP	dD/dMP
	Actual	MAX	Actual	**MAX**			
1	3.3	1.1	10	30	3.88	16	0.55
2	5.05	2.53	50	100	0.79	48	1.26
3	2.71	0.9	40	120	1.54	76	0.45
4	2.15	0.72	30	90	1.12	56	0.36
5	5.09	1.7	30	90	0.4	56	0.85
6	5.3	1.06	25	125	41.55	92	0.53
7	1.35	0.68	10	20	19.86	8	0.34
8	1.98	0.99	12	24	15.51	10	0.49
9	5.11	1.02	20	100	1.26	72	0.51
10	3.98	0.99	20	80	1.67	54	0.5
11	1.99	0.99	20	40	1.23	18	0.5
12	3.96	0.99	20	80	2.59	54	0.49
13	1.05	0.52	20	40	0.89	18	0.26
14	7.01	1	35	245	0.29	198	0.5
15	4.98	1	50	250	7.25	192	0.5
16	5.07	1.01	50	250	8.45	192	0.51
17	3.03	1.01	25	75	31.69	46	0.51
18	2.97	0.99	100	300	0.29	196	0.49
19	2.01	0.67	100	300	1	196	0.33
20	1.99	0.66	100	300	0.75	196	0.33
21	1.99	0.99	100	200	1.47	98	0.5
22	2.03	1.02	100	200	0.1	98	0.51
23	1.95	0.98	100	200	0.12	98	0.49
24	5.03	1.01	30	150	7.86	112	0.5
25	10.95	1.83	10	60	15.56	40	0.91

Table 4 Coordinates of the 2 clusters. The values of the cells have been normalized respect to the maximum value.

	No. of Workers	**% Manpower**	**Unitary Cost**	**Duration**
1	0.143	0.269	0.236	0.143
2	1	1	1	1
	Area	**dC/dMP**	**dA/dMP**	**dD/dMP**
1	0.143	0.002	0.04	0.207
2	1	1	1	1

Table 5a Data set regarding all the activities where the values of the columns have been normalized respect to the maximum value.

	No. of Workers	% Manpower	Unitary Cost	Duration
1	0.333	0.713	0.468	0.333
2	0.5	0.717	0.698	0.5
3	0.333	0.93	0.359	0.333
4	0.333	0.694	0.48	0.333
5	0.333	0.442	0.755	0.333
6	0.2	0.785	0.255	0.2
7	0.5	0.679	0.737	0.5
8	0.5	0.614	0.815	0.5
9	0.2	0.395	0.506	0.2
10	0.25	0.457	0.547	0.25
11	0.5	0.638	0.784	0.5
12	0.25	0.494	0.506	0.25
13	0.5	0.814	0.614	0.5
14	0.143	0.474	0.301	0.143
15	0.2	0.848	0.236	0.2
16	0.2	0.297	0.674	0.2
17	0.333	0.683	0.488	0.333
18	0.333	0.605	0.551	0.333
19	0.333	0.424	0.787	0.333
20	0.333	0.605	0.551	0.333
21	0.5	0.564	0.886	0.5
22	0.5	0.564	0.886	0.5
23	0.5	0.564	0.886	0.5
24	0.2	0.269	0.743	0.2
25	0.167	0.426	0.391	0.167

Table 5b Data set regarding all the activities where the values of the columns have been normalized respect to the maximum value.

	Area	dC/dMP	dA/dMP	dD/dMP
1	0.333	0.093	0.081	0.435
2	0.5	0.019	0.242	1
3	0.333	0.037	0.384	0.357
4	0.333	0.027	0.283	0.284
5	0.333	0.01	0.283	0.672
6	0.2	1	0.465	0.42
7	0.5	0.478	0.04	0.268
8	0.5	0.373	0.051	0.392
9	0.2	0.03	0.364	0.404
10	0.25	0.04	0.273	0.394
11	0.5	0.03	0.091	0.394
12	0.25	0.062	0.273	0.392
13	0.5	0.021	0.091	0.207
14	0.143	0.007	1	0.396
15	0.2	0.174	0.97	0.394
16	0.2	0.203	0.97	0.401
17	0.333	0.763	0.232	0.4
18	0.333	0.007	0.99	0.391
19	0.333	0.024	0.99	0.265
20	0.333	0.018	0.99	0.263
21	0.5	0.035	0.495	0.393
22	0.5	0.002	0.495	0.402
23	0.5	0.003	0.495	0.386
24	0.2	0.189	0.566	0.398
25	0.167	0.375	0.202	0.722

and not convenient, but the Fuzzy method attributes a numerical evaluation to their membership. Obviously, if the variables are few, the solution is trivial and there is no need to apply this approach. Otherwise, when variables are numerous as in the example, this method is able to provide a priority order to optimize the working operations. It should be noticed that the considerations concerning the available areas have not been reported in the paper for length limitations. In addition, although they include features related both to the occupied area by each activity and the interferences of simultaneous activities, the authors preferred to focus only on the novel procedure proposed in this paper.

In the original configuration of the time schedule reported in Table 5, the 25 activities show that their values is close to 1 only when they are near to the cluster 2, i.e. when they are not susceptible to be accelerated. In the same table, the values in the last 3 columns have information about the max value then

properly vary in the range [0, 1]. Only in this specific application, then, the first 5 columns are not very selective in terms of classification, because the values are similar for the each variable. Consequently, the classification was driven by the last 3 variables (dC/dMP, dA/dMP, dD/dMP). For example, it is significant to compare activity 1 (the closest to cluster 1 – the most convenient) to activity 20 (the closest to cluster 2 – the least convenient). Although variables 1-5 show similar values, strong differences are evidenced for variables 6 and 7. In particular, considering dC/dMP the difference is 0.0751, while for dA/dMP is larger and equal to 0.909. For activity 20, dA/dMP is 0.970; this means that the working area rises up too much if new workers are hired for it, producing critical consequences:

- Impossibility to perform other simultaneous tasks;
- Hazard growth for simultaneous tasks;
- High decrease of efficiency and production.

Even dC/dMP shows a relatively high value for task 20 (compared to other activities). The decision-maker can easily understand to not accelerate this activity for not increasing excessively the total costs. Similar considerations can be extended to other tasks and other variables.

Considering the critical path, tasks 5, 14, 15, 16, and 17 are external to it (Fig. 3), and thus the total working duration is not influenced by their eventual acceleration (Table 6). Examining membership data, only tasks 5 and 17 are close to cluster 1 (0.0954 and 0.0952), but if they had been in the critical path, other 9 activities would have been more convenient.

Table 6 Membership of the 25 activities to one of the two cluster.

Act	μ1	μ2	Cluster	Critical Path
1	0.0023	0.9977	1	Yes
2	0.0622	0.9378	1	Yes
3	0.2191	0.7809	1	**NO**
4	0.0956	0.9044	1	Yes
5	0.0954	0.9046	1	**NO**
6	0.4384	0.5616	1	Yes
7	0.0106	0.9894	1	NO
8	0.0067	0.9933	1	Yes
9	0.1897	0.8103	1	Yes
10	0.0867	0.9133	1	Yes
11	0.0030	0.9970	1	Yes
12	0.0871	0.9129	1	Yes
13	0.0030	0.9970	1	Yes
14	0.9549	0.0451	2	**NO**
15	0.9654	0.0346	2	**NO**
16	0.9676	0.0324	2	**NO**
17	0.0952	0.9048	1	**NO**
18	0.9539	0.0461	2	Yes
19	0.9554	0.0446	2	Yes
20	0.9549	0.0451	2	Yes
21	0.4111	0.5889	1	Yes
22	0.4087	0.5913	1	Yes
23	0.4087	0.5913	1	Yes
24	0.5603	0.4397	2	Yes
25	0.0470	0.9530	1	Yes

Because of this, it is not difficult to highlight the advantages of the fuzzy clustering technique compared to the traditional hard clustering. The latter would have only divided the activities into the two clusters, without any order. This does not represent a real situation and may not actually help the analyst, since it would have been really hard for him to subjectively choose the most convenient tasks to accelerate among those in a single cluster. The hard clustering can simplify the selection problem, but not solve it. On the contrary, the fuzzy method provides an objective classification of the cluster membership rates that is very useful for the analyst (Table 6). The shade classification is useful for rates close to the threshold. For instance, two tasks with values of 0.49 and 0.51 would be classified by the hard clustering as not member and member to a cluster respectively. This would lead the decision-maker to complex and not correct choices that can be avoided using the fuzzy technique. The numerical results have been also represented in the radial plot of Fig. 4 in which the distance to cluster 1 is reported in a logarithmic scale. In this way the authors have evidenced not only activities with a membership rate lower than 0.5, but also those really close to 0, producing a visual classification of the activities that are more convenient.

Finally, 2 clusters are enough for the proposed procedure, because a higher number of clusters would have not improved the classification of the activities, but only complicated the analytical examination. Obviously, the analyst should split the cumulate delay to the various activities according to the final ranking proposed by the fuzzy procedure. The repartition strategies are numerous, but generally the decision-maker can refer to the following two approaches:

- Max acceleration of the activities starting from the 1st one, increasing the number of workers up to the max limit;
- Delay repartition on different tasks, slightly increasing the number of workers.

5 Conclusion

This paper has analysed a situation very frequent in the field of civil engineering constructions: when the actual execution is not in compliance with the initial design plans, the decision-maker needs to introduce modifications in the work organization for a time recovery. In this case, the simple addition of manpower or machines does not assure a real and substantial acceleration of the execution. For instance, increasing resources for activities external to the critical path does not produce a decrease of the working duration, but only an unjustified cost growth. However, it is not enough to choose activities in the critical path, because the same improvement in terms of manpower or machines can imply different efficiency values and acceleration rates.

Sometimes, upgrading resources for not really productive tasks can determine cost rises that are almost completely ineffective. In general, the resource should be directed to the different activities not only according to cost and time considerations, but also evaluating the areas occupied to perform them. These values may collide with those actually available; then, a specific activity has an upgrade limitation due to the max number of workers or machines that can produce negative interactions among the tasks in terms of safety.

The proposed procedure, based on the Fuzzy-C Means clustering technique, provided very optimistic results. Although the example application was characterized by a restricted number of variables (8) and activities (25), the procedure actually evidenced the most convenient activities to accelerate for improving the time schedule. Furthermore, increasing the data set would not imply analytical complications or errors in the final outcomes, due to the implicit features of the method. In conclusion, the advantages of the method can be listed in the following:

- Possibility to delete already performed tasks or add new activities without modifying the algorithm.
- Possibility to increase the number of variables if needed, adding for example machine production and teams' efficiency data.
- Generation of an ordered list of the activities, according to the membership to the clusters.
- Application simplicity, even for operators not able to handles complex analytical methodologies.

References

Abonyi, J., Feil, B. (2007). *Cluster Analysis for Data Mining and System Identification*, Birkhäuser, Basel.

Amadore, A., Bosurgi, G., Pellegrino, O. (2014). Classification of measures from deflection tests by means of fuzzy clustering techniques. *Construction and Building Materials*. 53, pp. 173-181. https://doi.org/10.1016/j.conbuildmat.2013.11.094

Arunava, B., Anand, P. (2008). On path correlation and PERT bias. *European Journal of Operational Research*. 189, pp. 1208-1216. https://doi.org/10.1016/j.ejor.2007.01.061

Azaron, A., Perkgoz, C., Sakawa, M. (2005). A genetic algorithm approach for the time-cost trade-off in PERT networks. *Applied Mathematics and Computation*. 168, pp. 1317-1339. https://doi.org/10.1016/j.amc.2004.10.021

Bezdek, J. C., Ehrlich, R., Full, W. (1984). FCM: The Fuzzy c-Means Clustering Algorithm. *Computer& Geosciences*. 10(2-3), pp. 191-203. https://doi.org/10.1016/0098-3004(84)90020-7

Bosurgi, G, Trifirò, F. (2005a). A model based on artificial neural network and genetic algorithms for pavement maintenance management. *International Journal of Pavement Engineering*. 6, pp. 201–209. https://doi.org/10.1080/10298430500195432

Bosurgi, G, Trifirò, F. (2005b) A hybrid approach for pavement maintenance management of an Italian motorway. In: International Conference on Computing in Civil Engineering, Cancun, Mexico, July 12–15, 2005. https://doi.org/10.1061/40794(179)69

Bosurgi, G, Trifirò, F. (2006). A genetic algorithm approach for pavement maintenance program optimization using safety evaluation data. *International Journal of Pavements*. 5, pp. 164–174.

Bosurgi, G., D'Andrea, A., Pellegrino, O. (2010). Could drivers' visual behaviour influence road design?. *Advances in Transportation Studies*. 22, pp. 17-30.

Castro, J., Gomez, D., Tejada, J. (2008a). A polynomial rule for the problem of sharing delay costs in PERT networks. *Computers & Operations Research*. 35, pp. 2376-2387. https://doi.org/10.1016/j.cor.2006.11.003

Castro, J., Gomez, D., Tejada, J. (2008b). A rule for slack allocation proportional to the duration in a PERT network. *European Journal of Operational Research*. 187, pp. 556-570. https://doi.org/10.1016/j.ejor.2007.03.039

Chreitienne, P., Sourd, F. (2003). PERT scheduling with convex cost functions. *Theoretical Computer Science*. 292, pp. 145-164. https://doi.org/10.1016/S0304-3975(01)00220-1

Duda, R. O., Hart, P. E., Stork, D. G. (2001). *Pattern Classification*. John Wiley and Sons, USA, 2nd edition.

Faghihi, V., Reinschmidt, K. F., Kang, J. K. (2014). Construction scheduling using Genetic Algorithm based on Building Information Model. *Expert Systems with Applications*. 41, pp. 7565-7578. https://doi.org/10.1016/j.eswa.2014.05.047

Fatemi Ghomi, S. T., Rabbani, M. (2003). A new structural mechanism for reducibility of stochastic PERT networks. *European Journal of Operational Research*. 145, pp. 394-402. https://doi.org/10.1016/S0377-2217(02)00155-8

Fatemi Ghomi, S. T., Teimouri, E. (2002). Path critical index and activity critical index in PERT networks. *European Journal of Operational Research*. 141, pp. 147-152. https://doi.org/10.1016/S0377-2217(01)00268-5

Hardie, N. (2001). The prediction and control of project duration: a recursive model. *International Journal of Project Management*. 19, pp. 401-409. https://doi.org/10.1016/S0263-7863(00)00027-2

Madadi, M., Iranmanesh, H. (2012). A management oriented approach to reduce a project duration and its risk (variability). *European Journal of Operational Research*. 219, pp. 751-761. http://dx.doi.org/10.1016/j.ejor.2012.01.006

Martinez, A. M., Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 23(2), pp. 228-233. https://doi.org/10.1109/34.908974

McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley – IEEE.

Mouhoub, N. E., Benhocine, A., Belouadah, H. (2011). A new method for constructing a minimal PERT network. *Applied Mathematical Modelling*. 35, pp. 4575-4588. https://doi.org/10.1016/j.apm.2011.03.031

Pellegrino, O. (2011). Road context evaluated by means of fuzzy interval. *Cognition, Technology and Work*. 13(1), pp. 67-79. https://doi.org/10.1007/s10111-010-0155-2

Pellegrino, O. (2012). The application of drivers' cognitive reliability using fuzzy techniques. *Applied Artificial Intelligence*. 26(6), pp. 517-540. https://doi.org/10.1080/08839514.2012.670973

Ripley, B. D. (2005). *Pattern Recognition and Neural Networks*. Cambridge University Press, 8th edition.

Theodoridis, S., Koutroumbas, K. (2006). *Pattern Recognition*. Academic Press, USA, 3rd edition.

Trietsch, D., Baker, K. R. (2012). PERT 21: Fitting PERT/CPM for use in the 21st century. *International Journal of Project Management*. 30, pp. 490-502. https://doi.org/10.1016/j.ijproman.2011.09.004

Webb, A. R. (2002). *Statistical Pattern Recognition*. John Wiley and Sons Ltd, 2nd edition.

White, D., Fortune, J. (2002). Current practice in project management – an empirical study. *International Journal of Project Management*. 20, pp. 1-11. http://dx.doi.org/10.1016/S0263-7863(00)00029-6