O P E R A T I O N S R E S E A R C H A N D D E C I S I O N S

No. 4 2015

DOI: 10.5277/ord150405

Dmitri VIATTCHENIN[1]

REDUCING THE NUMBER OF PATHS IN A MINIMIZED PROJECT-NETWORK WITH GIVEN BOUNDS ON THE DURATIONS OF ACTIVITIES

This paper deals in a preliminary way with the problem of selecting the smallest possible number of dominant paths in a minimized project-network with given bounds on the permissible values of the durations of activities. For this purpose, a classification technique is proposed. This technique is based on a heuristic possibilistic clustering of interval-valued data. The basic concepts of heuristic possibilistic clustering are defined and methods for preprocessing interval-valued data are described. An illustrative example is considered in detail and some conclusions are formulated.

Keywords: *project-network, dominant path, heuristic possibilistic clustering, interval-valued data, stationary clustering structure, typical point*

1. Introduction

Project scheduling and project management are very relevant in a diverse range of problems. Applications can be found in such different areas as industry, engineering, research, defense, etc. Moreover, problems involving project scheduling and project management are very interesting for researchers, because they are difficult to solve.

It is often hard to obtain a priori exact values of the durations of activities in project management. Therefore, assuming that the duration of an activity may remain uncertain until it is completed is more realistic. Hence, it is appropriate to give lower and upper bounds on the possible durations of activities at the step of project planning. In this case, a critical path in the given project-network cannot be found a priori. However, arcs and vertices that cannot belong to a critical path for any possible realization of the project can be deleted from the project-network. Thus, the subset of dominant paths can be

[1]Department of Software Information Technology, Belarusian State University of Informatics and Radio-Electronics, P. Brovka St. 6, 220013 Minsk, Belarus, e-mail address: viattchenin@mail.ru

constructed and the initial project-network minimized. It should be noted, that the problem of uncertainty in project scheduling is considered, for example, in [1] and [7], where numerical data regarding upper and lower bounds on the durations of activities are considered.

Let us consider in brief some basic concepts and notation which were defined in [9] and [10]. Let $G^{(*)} = (X^{(*)t_1}, E^{(*)})$ be a graph without cycles. So, $X^{(*)t_1} = \{x^{(*)t_1} \mid t_1 = 1, \ldots, m_1\}$ is the set of vertices and $E^{(*)}$ is the set of arcs representing precedence constraints on the set $X^{(*)t_1}$. Let $X^* = \{x_1^*, \ldots, x_q^*\}$ be the set of all paths from the vertex $x^{(*)1} \in X^{(*)t_1}$ to the vertex $x^{(*)m_1} \in X^{(*)t_1}$ in the graph $G^{(*)} = (X^{(*)t_1}, E^{(*)})$. The graph $G^{(*)} = (X^{(*)t_1}, E^{(*)})$ is a project-network with start-vertex $x^{(*)1}$ and end-vertex $x^{(*)m_1}$. It is assumed that the durations of dummy activities $x^{(*)1}$ and $x^{(*)m_1}$ are equal to zero, and for any other activity $x^{(*)t_1}$, $1 < t_1 < m_1$, the closed interval $[x^{t_1(\min)}, x^{t_1(\max)}]$ of possible durations is given, where the condition $x^{t_1(\min)} \leq x^{t_1(\max)}$ is met.

Any longest path $x_k^* \in X^*$, $k \in \{1, \ldots, q\}$ from the start-vertex $x^{(*)1}$ to the end-vertex $x^{(*)m_1}$ is called a critical path in the network $G^{(*)}$. If the condition $x^{t_1(\min)} = x^{t_1(\max)}$ is met for each activity $x^{(*)t_1} \in X^{(*)t_1}$, then the graph $G^{(*)}$ is a deterministic project-network whose total duration is defined by the length of a critical path. If the condition $x^{t_1(\min)} = x^{t_1(\max)}$ is not met, then the length of a longest path in the realization of the project is unknown a priori. In other words, a deterministic project-network is a special case of the model considered in this paper in which the upper and lower bounds are equal for each random time.

The dominance relations $\preceq$ and $\prec$ on the set of paths in the network $G^{(*)}$ were defined in [9, 10] as follows: If the length of the path x_j^* is greater than the length of the path x_i^* for all the possible durations of the activities, then the dominance relation $x_i^* \prec x_j^*$ is met for paths x_i^* and x_j^*. If the length of the path x_j^* cannot be less than the length of the path x_i^* for all the possible durations of the activities, then the dominance relation $x_i^* \preceq x_j^*$ is met. If the dominance relation $x_i^* \preceq x_j^*$ holds, then the path x_j^* dominates the path x_i^*. If for any path $x_k^* \in X^{(*)}$, $k \in \{1, \ldots, q\}$, there exists a path $x_i \in X$, $i \in \{1, \ldots, n\}$ such that the dominance relation $x_k^* \preceq x_i$ is met, then any minimal set of paths $X = \{x_1, \ldots, x_n\}$, $X \subseteq X^{(*)}$ is called a dominant set for network $G^{(*)}$. Hence, $\operatorname{card}(X^{(*)}) = q$, $\operatorname{card}(X) = n$, and $n \leq q$.

The problem of constructing a minimal subgraph G of the given graph $G^{(*)}$, which is a network with the same start-vertex $x^{(*)1}$ and end-vertex $x^{(*)m_1}$, and for which $X \subseteq X^{(*)}$ is considered by Sotskov and Shilak [9, 10], who proposed an effective approach to the minimization of a project-network with given bounds on the durations of activities. In other words, the set of all paths X of the subgraph $G = (X^{t_1}, E)$ is a dominant set for the network $G^{(*)} = (X^{(*)t_1}, E^{(*)})$.

On the other hand, the number n of paths in the minimized project-network G can also be large and there still exists the problem of selecting one or several appropriate paths from the subset X of dominant paths. A heuristic technique to solve such problems is proposed in this paper. The proposed technique of deriving such a subset of paths is based on a heuristic approach to the possibilistic clustering of interval-valued data. The contents of this paper are as follows: in the second section a classification technique for choosing paths in the subset of dominant paths is proposed, in the third section an illustrative example of applying the proposed technique to an artificial project-network is given, in the fourth section final remarks are stated.

2. Discriminating paths in the minimized project-network

The first subsection considers the basic concepts of heuristic possibilistic clustering. Indexes for evaluating the results of clustering are described in the second subsection. Methods for the preprocessing of interval-valued data are described in the third subsection and a classification technique for selecting the minimal number of paths in the minimized project-network is presented in the fourth subsection of this section.

2.1. Basic concepts of a heuristic approach to possibilistic clustering

Let us recall the basic concepts of a heuristic method of possibilistic clustering which was proposed in [12]. The essence of this heuristic approach to possibilistic clustering is that the clustering structure of a set of objects is assumed to be based directly on the formal definition of a fuzzy cluster and possibilistic memberships are also directly determined from the values of the pairwise similarity of objects.

Let $X = \{x_1, ..., x_n\}$ be the initial set of objects. Let T be a fuzzy tolerance on X and α be the value defining the α-level of T, $\alpha \in (0,1]$. The columns or rows of the fuzzy tolerance matrix can be interpreted as fuzzy sets $\{A^1, ..., A^n\}$. Let $\{A^1, ..., A^n\}$ be fuzzy sets on X, which are generated by a fuzzy tolerance T. The α-level fuzzy set $A^l_{(\alpha)} = \{(x_i, \mu_{A^l}(x_i)) \mid \mu_{A^l}(x_i) \geq \alpha\}$, $l \in [1, n]$ is a fuzzy α-cluster or, simply, a fuzzy cluster. Thus $A^l_{(\alpha)} \subseteq A^l$, $\alpha \in (0,1]$, $A^l \in \{A^1, \ldots, A^n\}$ and μ_{l_i} is the degree of membership of the element $x_i \in X$ in some fuzzy cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1, n]$. The value of α is the tolerance threshold of the elements of fuzzy clusters.

The degree of membership of the element $x_i \in X$ in some fuzzy cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1, n]$ can be defined as

$$\mu_{li} = \begin{cases} \mu_{A^l}(x_i), & x_i \in A^l_\alpha \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where the α-level $A^l_\alpha = \{x_i \in X \mid \mu_{A^l}(x_i) \geq \alpha\}$, $\alpha \in (0,1]$ of the fuzzy set A^l is the support of the fuzzy cluster $A^l_{(\alpha)}$. Hence, the condition $A^l_\alpha = \mathrm{Supp}(A^l_{(\alpha)})$ is met for each fuzzy cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1, n]$. The degree of membership can be interpreted as the degree of typicality of an element within a fuzzy cluster.

Let T be a fuzzy tolerance on X, where X is a set of objects, and $\{A^1_{(\alpha)}, \ldots, A^n_{(\alpha)}\}$ is a family of fuzzy clusters for some $\alpha \in (0,1]$. The point $\tau^l_e \in A^l_\alpha$, for which

$$\tau^l_e = \arg\max_{x_i} \mu_{li}, \qquad \forall x_i \in A^l_\alpha \tag{2}$$

is called a typical point of the fuzzy cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1, n]$. A fuzzy cluster $A^l_{(\alpha)}$ can have multiple typical points. That is why the symbol e is used as an index of a typical point.

Let $R^\alpha_{c(z)}(X) = \{A^l_{(\alpha)} \mid l = \overline{1, c},\ 2 \leq c \leq n,\ \alpha \in (0,1]\}$ be a family of fuzzy clusters for some value of the tolerance threshold α, $\alpha \in (0,1]$ which are generated by some fuzzy tolerance T on the initial set of elements $X = \{x_1, \ldots, x_n\}$. If the condition

$$\sum_{l=1}^{c} \mu_{li} > 0, \qquad \forall x_i \in X \tag{3}$$

is met for all fuzzy clusters $A^l_{(\alpha)} \in R^\alpha_{c(z)}(X)$, $l = \overline{1, c}$, $c \leq n$, then such a family is an allotment of the elements of the set $X = \{x_1, \ldots, x_n\}$ to the fuzzy clusters $\{A^l_{(\alpha)}, l = \overline{1, c}, 2 \leq c \leq n\}$ for some value of the tolerance threshold α. It should be noted that several allotments $R^\alpha_{c(z)}(X)$ can exist for a given tolerance threshold α. That is why the symbol z is used as an index of an allotment.

The allotment $R^\alpha_I(X) = \{A^l_{(\alpha)} \mid l = \overline{1, n},\ \alpha \in (0,1]\}$ of the set of objects to n fuzzy clusters for some tolerance threshold $\alpha \in (0,1]$ is an initial allotment of the set

$X = \{x_1, ..., x_n\}$. In other words, if the initial data are represented by a matrix generated by some fuzzy tolerance *T*, then the rows or columns of this matrix are the fuzzy sets $A^l \subseteq X$, $l = \overline{1, n}$ and the α-level fuzzy sets $A^l_{(\alpha)}$, $l = \overline{1, c}$, $\alpha \in (0, 1]$ are fuzzy clusters. These fuzzy clusters constitute an initial allotment corresponding to the tolerance threshold α and can be considered to be the components of a clustering.

If some allotment $R^{\alpha}_{c(z)}(X) = \{A^l_{(\alpha)} \mid l = \overline{1, c},\ c \le n,\ \alpha \in (0, 1]\}$ corresponds to the formulation of a concrete problem, then this allotment is an adequate allotment. In particular, if the condition

$$\bigcup_{l=1}^{c} A^l_{\alpha} = X \tag{4}$$

and the condition

$$\mathrm{card}(A^l_{\alpha} \cap A^m_{\alpha}) = 0, \qquad \forall A^l_{(\alpha)}, A^m_{(\alpha)}, \qquad l \ne m, \qquad \alpha \in (0, 1] \tag{5}$$

are met for all fuzzy clusters $A^l_{(\alpha)}$, $l = \overline{1, c}$ of some allotment $R^{\alpha}_{c(z)}(X) = \{A^l_{(\alpha)} \mid l = \overline{1, c}$, $c \le n\}$ for a value $\alpha \in (0, 1]$, then such an allotment is an allotment to fully separate fuzzy clusters.

Fuzzy clusters in the sense of definition (1) can intersect. If the intersection of any pair of different fuzzy clusters is an empty set, then conditions (4) and (5) are met and fuzzy clusters are called fully separate fuzzy clusters. Otherwise, fuzzy clusters are called particularly separate fuzzy clusters and $w \in \{0, \dots, n\}$ is the maximum number of elements in the intersection of different fuzzy clusters. When $w = 0$, fuzzy clusters are fully separate fuzzy clusters. Thus, conditions (4) and (5) can be generalized to the case of particularly separate fuzzy clusters. Hence, condition

$$\sum_{l=1}^{c} \mathrm{card}(A^l_{\alpha}) \ge \mathrm{card}(X), \qquad \forall A^l_{(\alpha)} \in R^{\alpha}_{c(z)}(X), \qquad \alpha \in (0, 1], \qquad \mathrm{card}(R^{\alpha}_{c(z)}(X)) = c \tag{6}$$

and condition

$$\mathrm{card}(A^l_{\alpha} \cap A^m_{\alpha}) \le w, \qquad \forall A^l_{(\alpha)}, A^m_{(\alpha)}, \qquad l \ne m, \qquad \alpha \in (0, 1] \tag{7}$$

are generalizations of conditions (4) and (5). Obviously, if $w = 0$ in conditions (6) and (7), then conditions (4) and (5) are met. The adequate allotment $R^{\alpha}_{c(z)}(X)$ for some value of the tolerance threshold $\alpha \in (0,\ 1]$ is a family of fuzzy clusters which are elements of the initial allotment $R_I^\alpha(X)$ for the given value of α and the family of fuzzy clusters should satisfy conditions (6) and (7). Hence, the construction of adequate allotments $R_{c(z)}^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c},\ c \le n\}$ for any α is a trivial combinatorical problem.

The allotment $R_P^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c}\}$ of a set of objects to the minimal number c, $2 \le c \le n$ of fully separate fuzzy clusters for some tolerance threshold $\alpha \in (0,1]$ is called the principal allotment of the set $X = \{x_1, ..., x_n\}$.

Several adequate allotments can exist. Thus, the problem consists of selecting the unique adequate allotment $R_c^*(X)$ from the set B of adequate allotments, $B = \{R_{c(z)}^\alpha(X)\}$, which is the class of possible solutions of the given classification problem. The selection of the unique adequate allotment $R_c^*(X)$ from the set $B = \{R_{c(z)}^\alpha(X)\}$ of adequate allotments must be made on the basis of evaluating the allotments. In particular, the criterion

$$F(R_{c(z)}^\alpha(X), \alpha) = \sum_{l=1}^{c} \frac{1}{n_l} \sum_{i=1}^{n_l} \mu_{li} - \alpha c \tag{8}$$

where c is the number of fuzzy clusters in the allotment $R_{c(z)}^\alpha(X)$ and $n_l = \mathrm{card}(A_\alpha^l)$, $A_{(\alpha)}^l \in R_{c(z)}^\alpha(X)$ is the number of elements in the support of the fuzzy cluster $A_{(\alpha)}^l$, can be used to evaluate allotments. The maximum value of function (8) corresponds to the best allotment of objects to c fuzzy clusters. Hence, such a classification problem can be characterized formally as determining the solution $R_c^*(X)$ satisfying

$$R_c^*(X) = \arg \max_{R_{c(z)}^\alpha(X) \in B} F(R_{c(z)}^\alpha(X), \alpha) \tag{9}$$

The problem of cluster analysis can be defined in general as the problem of deriving the unique allotment $R_c^*(X)$ resulting from the classification process and defining a fixed or unknown number c of fuzzy clusters can be considered as the aim of classification.

Direct heuristic algorithms of possibilistic clustering can be divided into two types: relational versus prototype-based. The matrix of initial data for direct heuristic relational algorithms of possibilistic clustering is a fuzzy tolerance relation matrix and a matrix of attributes is the input matrix for prototype-based algorithms. In particular, the group of relational direct heuristic algorithms of possibilistic clustering includes:

- D-AFC(c) algorithm constructing an allotment among a given number c of partially separate fuzzy clusters,

• D-PAFC algorithm constructing the principal allotment among an unknown number of at least c fully separate fuzzy clusters,

• D-AFC-PS(c) algorithm – a partially supervised construction of an allotment among a given number c of partially separate fuzzy clusters.

On the other hand, the family of direct prototype-based heuristic algorithms of possibilistic clustering includes:

• D-AFC-TC algorithm constructing an allotment among an unknown number c of fully separate fuzzy clusters,

• D-PAFC-TC algorithm constructing the principal allotment among an unknown number of at least c fully separate fuzzy clusters,

• D-AFC-TC(α) algorithm constructing an allotment among an unknown number c of fully separate fuzzy clusters with respect to the minimal value α of the tolerance threshold.

It should be noted that these direct prototype-based heuristic possibilistic clustering algorithms are based on a transitive closure of an initial fuzzy tolerance relation.

2.2. Evaluating fuzzy clusters

The results of classification should be assessed. Some formal criteria can be useful for this aim. For example, the most appropriate distance between fuzzy sets for data preprocessing can be selected on the basis of evaluating the results of classification. The problem of evaluating fuzzy clusters was considered in [12].

Qualitative inspection of the results of fuzzy clustering can be done, e.g., using a linear or quadratic index of fuzziness . These two indexes are considered by Kaufmann [5]. A modification of the linear index of fuzziness is defined in [12] as

$$I_L(A^l_{(\alpha)}) = \frac{2}{n_l} d_H(A^l_{(\alpha)}, \underline{A}^l_{(\alpha)}) \tag{10}$$

where $n_l = \mathrm{card}(A^l_\alpha)$, $A^l_{(\alpha)} \in R^*_c(X)$ is the number of objects in the fuzzy cluster $A^l_{(\alpha)}$ and $d_H(A^l_{(\alpha)}, \underline{A}^l_{(\alpha)})$ is the Hamming distance,

$$d_H(A^l_{(\alpha)}, \underline{A}^l_{(\alpha)}) = \sum_{x_i \in A^l_\alpha} \left| \mu_{li} - \mu_{\underline{A}^l_{(\alpha)}}(x_i) \right| \tag{11}$$

between the fuzzy cluster $A^l_{(\alpha)}$ and the crisp set $\underline{A}^l_{(\alpha)}$ nearest to the fuzzy cluster $A^l_{(\alpha)}$. The membership function of the crisp set $\underline{A}^l_{(\alpha)}$ can be defined as

$$\mu_{\underline{A}^l_{(\alpha)}}(x_i) = \begin{cases} 0, & \mu_{A^l_{(\alpha)}}(x_i) \le 0.5 \\ 1, & \mu_{A^l_{(\alpha)}}(x_i) > 0.5 \end{cases}, \qquad \forall x_i \in A^l_\alpha \tag{12}$$

where $\alpha \in (0,1]$.

The modified quadratic index of fuzziness is defined as [12]

$$I_Q(A^l_{(\alpha)}) = \frac{2}{\sqrt{n_l}} d_E(A^l_{(\alpha)}, \underline{A}^l_{(\alpha)}) \tag{13}$$

where $n_l = \mathrm{card}(A^l_\alpha)$, $A^l_{(\alpha)} \in R^*_c(X)$ and $d_E(A^l_{(\alpha)}, \underline{A}^l_{(\alpha)})$ is the Euclidean distance,

$$d_E(A^l_{(\alpha)}, \underline{A}^l_{(\alpha)}) = \sqrt{\sum_{x_i \in A^l_\alpha} \left(\mu_{li} - \mu_{\underline{A}^l_{(\alpha)}}(x_i)\right)^2} \tag{14}$$

between the fuzzy cluster $A^l_{(\alpha)}$ and the crisp set $\underline{A}^l_{(\alpha)}$ which is defined by formula (12).

Indexes (10) and (13) measure the degree of fuzziness of fuzzy clusters which are elements of the allotment $R^*_c(X)$. Obviously, $I_L(A^l_{(\alpha)}) = I_Q(A^l_{(\alpha)}) = 0$ for a crisp set $A^l_{(\alpha)} \in R^*_c(X)$. Otherwise, if $\mu_{li} = 0.5$, $\forall x_i \in A^l_\alpha$, then the fuzzy cluster $A^l_{(\alpha)} \in R^*_c(X)$ is a maximally fuzzy set and $I_L(A^l_{(\alpha)}) = I_Q(A^l_{(\alpha)}) = 1$.

The density of a fuzzy cluster was defined as follows [12]:

$$D(A^l_{(\alpha)}) = \frac{1}{n_l} \sum_{x_i \in A^l_\alpha} \mu_{li} \tag{15}$$

where $n_l = \mathrm{card}(A^l_\alpha)$, $A^l_{(\alpha)} \in R^*_c(X)$ and the membership degree μ_{li} is defined by formula (1). It is obvious that the condition

$$0 < D(A^l_{(\alpha)}) \le 1 \tag{16}$$

is met for each fuzzy cluster $A^l_{(\alpha)}$ in $R^*_c(X)$. Moreover, $D(A^l_{(\alpha)}) = 1$ for any crisp set $A^l_{(\alpha)} \in R^*_c(X)$ and any tolerance threshold α, $\alpha \in (0,1]$. The density of a fuzzy cluster measures the average membership degree of the elements of a fuzzy cluster.

2.3. Notes on interval-valued data preprocessing

Interval uncertainty in the initial data is a basic type of uncertainty in clustering problems. This fact was shown by Kreinovich and Kosheleva in [6]. Hence, initially interval-valued data preprocessing methods must be considered.

Let $X = \{x_1, \ldots, x_n\}$ be a set of n objects in an m_1-dimensional feature space with coordinate axis labels $(x^1, \ldots, x^{t_1}, \ldots, x^{m_1}\}$. Each object x_i is represented as a vector of intervals $x_i = (\hat{x}_i^1, \ldots, \hat{x}_i^{t_1}, \ldots, \hat{x}_i^{m_1})$, where $\hat{x}_i^{t_1} = [\hat{x}_i^{t_1(\min)},\ \hat{x}_i^{t_1(\max)}]$. Hence, the table of interval-valued data $\hat{X}_{n\times m_1} = [\hat{x}_i^{t_1}]$ is made up of *n* rows representing *n* objects to be clustered, and m_1 columns representing m_1 interval variables. In other words, each cell of this table contains an interval $\hat{x}_i^{t_1} = [\hat{x}_i^{t_1(\min)}, \hat{x}_i^{t_1(\max)}]$, $i \in \{1, \ldots, n\}$, $t_1 \in \{1, \ldots, m_1\}$.

The initial data matrix can be represented as a set of two matrices $\hat{X}_{n\times m_1}^{t_2} = [\hat{x}_i^{t_1(t_2)}]$, $i = 1, \ldots, n$, $t_2 \in \{\min, \max\}$ and the "plausible" number *c* of fuzzy clusters can be different for each matrix $\hat{X}_{n\times m_1}^{t_2} = [\hat{x}_i^{t_1(t_2)}]$, $t_2 \in \{\min, \max\}$. The clustering structure of a data set depends on the type of the initial data. Three types of such structures are defined in [12]. Firstly, if the number of clusters *c* is some constant for each matrix $\hat{X}_{n\times m_1}^{t_2} = [\hat{x}_i^{t_1(t_2)}]$, $t_2 \in \{\min, \max\}$ and the coordinates of the prototypes $\{\overline{\tau}^1, \ldots, \overline{\tau}^c\}$ of the clusters $\{A^1, \ldots, A^c\}$ are constant, then the clustering structure is called stable. Secondly, if the actual number of clusters *c* is some constant for each matrix $\hat{X}_{n\times m_1}^{t_2} = [\hat{x}_i^{t_1(t_2)}]$, $t_2 \in \{\min, \max\}$ and the coordinates of the prototypes of the clusters are not constant, then the clustering structure is called quasi-stable. Thirdly, if the number of clusters *c* is different for each matrix $\hat{X}_{n\times m_1}^{t_2} = [\hat{x}_i^{t_1(t_2)}]$, $t_2 \in \{\min, \max\}$, then the corresponding clustering structure is called unstable. The purpose of clustering is to classify the set $X = \{x_1, ..., x_n\}$ into c fuzzy clusters and the number of clusters *c* can be unknown because it depends on the situation. In other words, the clustering structure of the set of objects $X = \{x_1, ..., x_n\}$ must be stable in each situation. A simple and effective technique for constructing a stationary clustering structure for an interval-valued data set is described in [12].

Relational heuristic algorithms for possibilistic clustering can be applied directly to the data given as a matrix with some fuzzy tolerance $T = [\mu_T(x_i, x_j)]$, $i, j = 1, \ldots, n.$ This means that it can be used with objects by attributes data by choosing a suitable metric to measure similarity. However, the initial data should be normalized. A method of normalizing interval-valued data was defined in [12] as follows:

$$x_i^{t_1(t_2)} = \frac{\hat{x}_i^{t_1(t_2)} - \min\limits_{i,t_2} \hat{x}_i^{t_1(t_2)}}{\max\limits_{i,t_2} \hat{x}_i^{t_1(t_2)} - \min\limits_{i,t_2} \hat{x}_i^{t_1(t_2)}}, \tag{17}$$

where $t_1 = 1, \ldots, m_1$, $t_2 \in \{\min, \max\}$. Hence, each object x_i, $i = 1, \ldots, n$ can be considered as an interval-valued fuzzy set and $\mu_{x_i}(x^{t_1}) = [\mu_{x_i}(x^{t_1(\min)}),\ \mu_{x_i}(x^{t_1(\max)})]$, $i = 1, \ldots, n$, $t = 1, \ldots, m$ is its membership function, where $\mu_{x_i}(x^{t_1(\min)}) \in [0,1]$, $\mu_{x_i}(x^{t_1(\max)}) \in [0,1]$.

Various distance and similarity measures for interval-valued fuzzy sets have been proposed in the literature. Firstly, some methods for measuring distances between interval-valued fuzzy sets were proposed by Burillo and Bustince in [2]. For example, the normalized Euclidean distance was defined as follows:

$$d_I(x_i, x_j) = \sqrt{\frac{1}{2m_1} \sum_{t_1=1}^{m_1} \left\{ \left(\mu_{x_i}(x^{t_1(\min)}) - \mu_{x_j}(x^{t_1(\min)}) \right)^2 + \left(\mu_{x_i}(x^{t_1(\max)}) - \mu_{x_j}(x^{t_1(\max)}) \right)^2 \right\}} \tag{18}$$

for all $i, j = 1, \ldots, n.$

Secondly, the normalized Euclidean distance between interval-valued fuzzy sets based on the Hausdorff metric was defined by Grzegorzewski [3]:

$$e_I(x_i, x_j) = \sqrt{\frac{1}{m_1} \sum_{t_1=1}^{m_1} \max\left\{ \left(\mu_{x_i}(x^{t_1(\min)}) - \mu_{x_j}(x^{t_1(\min)}) \right)^2, \left(\mu_{x_i}(x^{t_1(\max)}) - \mu_{x_j}(x^{t_1(\max)}) \right)^2 \right\}} \tag{19}$$

for all $i, j = 1, \ldots, n.$

Thirdly, a similarity measure between interval-valued fuzzy sets was defined by Ju and Yuan in [4] as follows:

$$s_I(x_i, x_j) = 1 - \frac{1}{\sqrt[\lambda]{m_1}} \sqrt[\lambda]{\sum_{t_1=1}^{m_1} \left| \frac{\mu_{x_i}(x^{t_1(\min)}) + \mu_{x_i}(x^{t_1(\max)})}{2} - \frac{\mu_{x_j}(x^{t_1(\min)}) + \mu_{x_j}(x^{t_1(\max)})}{2} \right|^{\lambda}} \tag{20}$$

for all $i, j = 1, \ldots, n$ and $1 \le \lambda < \infty$.

Moreover, the coefficients of dissimilarity between objects can be constructed on the basis of generalizing distances between fuzzy sets [12]. In particular, a generalization of the normalized Euclidean distance for interval-valued fuzzy sets can be described by the expression

$$\varepsilon_I(x_i, x_j) = \sqrt{\frac{1}{m_1}\sum_{t_1=1}^{m_1}\left(\frac{1}{2^2}\sum_{\substack{t_2\in\{\min,\\ \max\}}}\left(\mu_{x_i}(x^{t_1(t_2)}) - \mu_{x_j}(x^{t_1(t_2)})\right)^2\right)} \tag{21}$$

for all $i, j = 1, \ldots, n.$

The matrix of fuzzy intolerance $I = [\mu_I(x_i,\ x_j)],\ i, j = 1,\ \ldots,\ n$ is the result of applying formulae (18), (19), (21) to the family of interval-valued fuzzy sets $\{x_1, ..., x_n\}$. The matrix of fuzzy tolerance $T = [\mu_T(x_i, x_j)],\ i, j = 1, \ldots, n$ can be obtained after applying the complementary operation [5]

$$\mu_T(x_i, x_j) = 1 - \mu_I(x_i, x_j), \qquad \forall i, j = 1, \ldots, n \tag{22}$$

to the matrix of fuzzy intolerance $I = [\mu_I(x_i, x_j)]$.

2.4. A heuristic technique for selecting paths in a minimized project-network

Let graph $G = (X^{t_1}, E)$ be a minimal subgraph of the original graph $G^{(*)} = (X^{(*)t_1}, E^{(*)})$. We assume that the subgraph $G = (X^{t_1}, E)$ is constructed according to the approach proposed in [9] and [10]. Let $X = \{x_1,\ \ldots, x_n\}$ denote the set of all paths in the graph $G = (X^{t_1}, E)$ from vertex $x^1 \in X^{t_1}$ to vertex $x^{m_1} \in X^{t_1}$ and $\text{card}(X) = n$ be the cardinality of the set *X*. In other words, the value *n* is the number of all paths from start-vertex x^1 to end-vertex x^{m_1} in the graph $G = (X^{t_1}, E)$. Hence, all the paths can be numbered, $x_i \in X$, $i \in \{1, \ldots, n\}$.

Each path can be considered as an object x_i, $i \in \{1, \ldots, n\}$, and the vertices $\hat{x}^{t_1}$, $t_1 \in \{1, \ldots, m_1\}$ are attributes of the object. This is why the problem of constructing a minimal subset of paths can be considered as a task of classifying an interval-valued data set. Hence, a technique for selecting the most appropriate paths in a minimized project-network can be summarized as follows:

1. The initial interval-valued data are contained in the matrix of attributes $\hat{X}_{n\times m_1} = [\hat{x}_i^{t_1(t_2)}],\ i = 1,\ \ldots,\ n,\ t_1 = 1,\ \ldots,\ m_1,\ t_2 = \{\min, \max\}$ and a technique for constructing a stationary clustering structure $R_c^*(X)$ describing interval-valued data should be applied to the data set.

2. Calculate the value of the density $D(A^l_{(\alpha)})$ of each fuzzy cluster in the allotment constructed, $A^l_{(\alpha)} \in R^*_c(X)$, and construct a subset $\{A^l_{(\alpha)}\} \subset R^*_c(X)$ of fuzzy clusters with the maximal values of the density.

3. Typical points τ^l_e of the fuzzy clusters which are elements of the subset constructed $\{A^l_{(\alpha)}\} \subset R^*_c(X)$ should be selected as elements of the minimal subset of paths.

The modified linear index of fuzziness (10) and the modified quadratic index of fuzziness (13) can also be used instead of the density of a fuzzy cluster (15) in the technique presented and a subset of fuzzy clusters with the minimal value of index (10) or (13) must be constructed in step 2.

Note at this point that an allotment among fuzzy clusters, where one object is a unique element of some fuzzy cluster and its typical point, can be obtained as a result of applying the proposed technique to the data. On the other hand, an allotment among fuzzy clusters, where each object is a unique element of the corresponding fuzzy cluster and its typical point, can also be obtained. This is why either a unique path or all the paths from a minimized project-network will be extreme cases of solutions to the classification task. The proposed technique should be illustrated by a simple example.

3. An illustrative example

Let us consider an application of the proposed technique based on the network $G^{(*)} = (X^{(*)t_1}, E^{(*)})$ describing the project considered in [9]. The corresponding initial network (Fig. 1) is presented in a form wherein activities are represented by vertices of the graph.

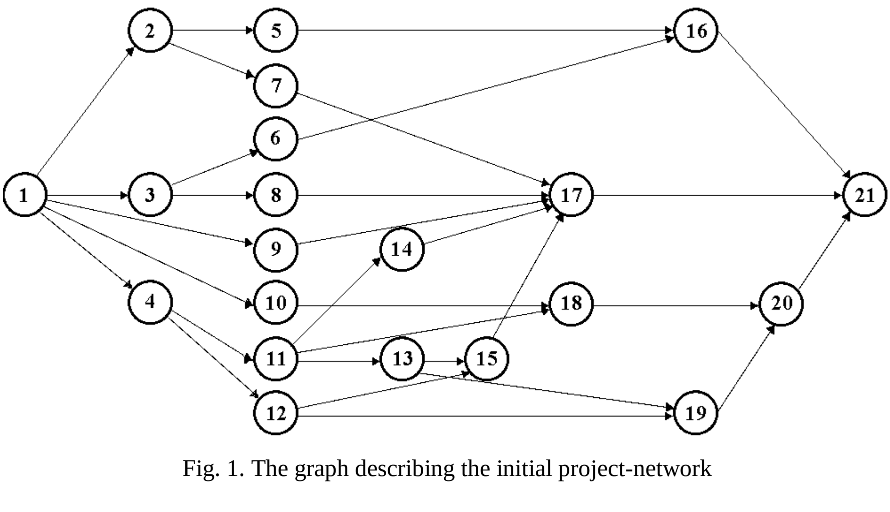

Fig. 1. The graph describing the initial project-network

Thus the set $X^{(*)} = \{x^{(*)1}, \ldots, x^{(*)21}\}$ is the set of vertices of the initial graph $G^{(*)}$ and $\mathrm{card}(X^*) = 12$. The numerical values of the durations of the activities in the initial network $G^{(*)} = (X^{(*)t_1}, E^{(*)})$ describing the project are given in Table 1.

Table 1. Durations of the activities in the initial network of the project

Duration of activity	Vertex number																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Minimal	0	11	2	1	12	16	15	16	15	3	1	14	2	18	8	8	9	22	14	8	0
Maximal	0	14	4	3	14	19	19	22	22	4	2	18	8	24	14	10	13	25	16	12	0

An approach to minimizing the initial project-network [9] was applied to the original graph $G^{(*)} = (X^{(*)t_1}, E^{(*)})$. As a result, the graph $G = (X^{t_1}, E)$ describing the minimized project-network was obtained. The result of applying this approach to minimizing the project-network $G^{(*)} = (X^{(*)t_1}, E^{(*)})$ is presented in Fig. 2.

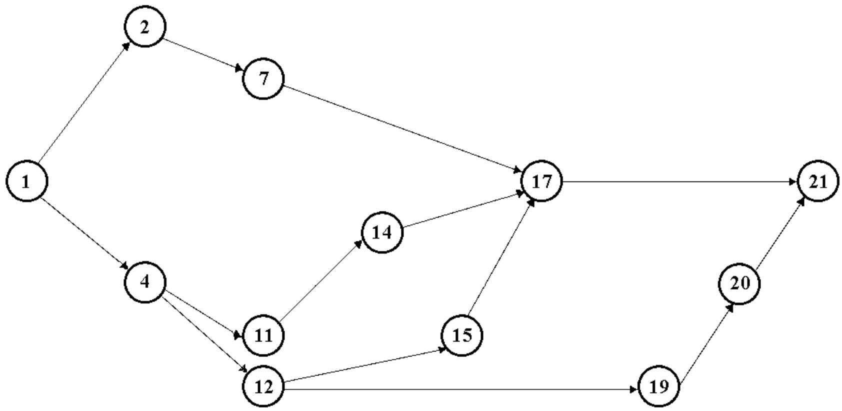

Fig. 2. The graph describing the minimized project-network

The set $X = \{x_1, \ldots, x_4\}$ of these paths is the dominant network for the initial graph $G^{(*)} = (X^{(*)t_1}, E^{(*)})$. In other words, all the paths in the digraph constructed, $G = (X^{l_1}, E)$, are critical for some admissible set of durations of the activities for the original project-network.

The project-network presented in Fig. 2 can be described by a matrix of attributes, where paths correspond to objects and vertices to attributes of the objects classified, where the values of attributes are represented by intervals. The matrix of interval-valued data for classification is given in Table 2, where the durations of activities corresponding to the start-vertex x^1 and the end-vertex x^{21} are omitted.

Table 2. Characteristics of paths in the minimized network of the project

Path number, i	Vertex number, t_1									
	2	4	7	11	12	14	15	17	19	20
1	[11, 14]	[0, 0]	[15, 19]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[9, 13]	[0, 0]	[0, 0]
2	[0, 0]	[1, 3]	[0, 0]	[1, 2]	[0, 0]	[18, 24]	[0, 0]	[9, 13]	[0, 0]	[0, 0]
3	[0, 0]	[1, 3]	[0, 0]	[0, 0]	[14, 18]	[0, 0]	[8, 14]	[9, 13]	[0, 0]	[0, 0]
4	[0, 0]	[1, 3]	[0, 0]	[0, 0]	[14, 18]	[0, 0]	[0, 0]	[0, 0]	[14, 16]	[8, 12]

The proposed technique for selecting the most appropriate paths in the minimized project-network G was applied to these interval-valued data. Let us consider the results of classification obtained after applying the first step of the technique.

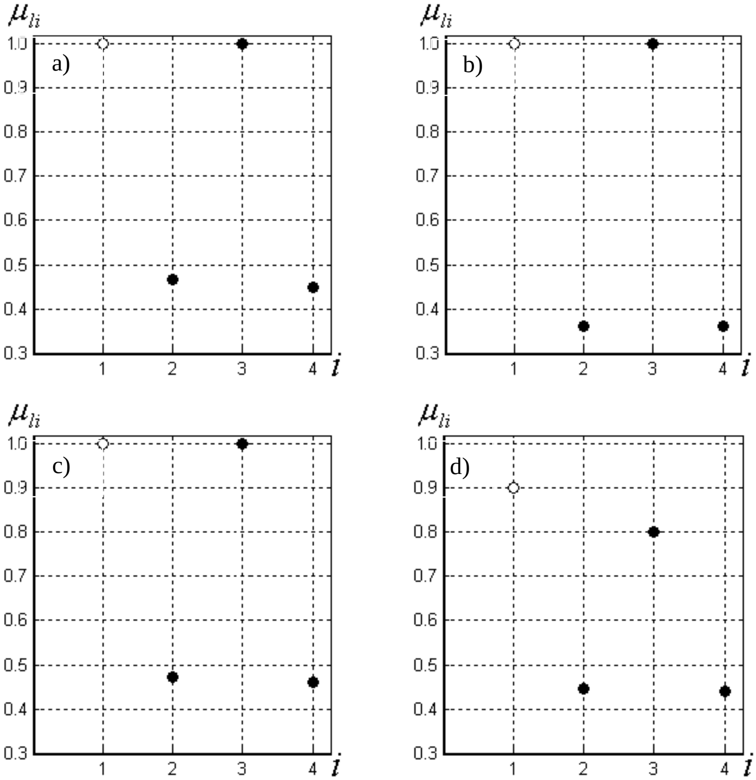

Fig. 3. Membership functions for two fuzzy clusters obtained after applying the first step of the proposed technique: a) using the distance (18), b) using the distance (19), c) using the similarity measure (20), d) using the dissimilarity measure (21)

The clustering structure of the interval-valued data set presented is the quasi-stable clustering structure. The initial interval-valued data were normalized according to formula (17), and the allotment $R_c^*(X)$ among two fully separated fuzzy clusters was constructed using formulae (18)–(21). We assume that $\lambda = 2$ in formula (20). The first class is formed by one object and the second class is formed by three objects in all the experiments. The membership functions of the two fuzzy clusters obtained after applying the first step of the proposed procedure using formulae (18)–(21) are presented in Fig. 3. The membership values for the first fuzzy cluster are represented by ○ and the membership values for the second fuzzy cluster are represented by black dots.

The first path $x_1 \in X$ is the unique typical point τ^1 of the first fuzzy cluster and the third path $x_3 \in X$ is the unique typical point τ^2 of the second fuzzy cluster from the allotment obtained. It should be noted that both fuzzy clusters are subnormal fuzzy sets in the case of using dissimilarity measure (21), because this dissimilarity measure does not satisfy the antireflexivity condition. Calculating the values of the indexes (10), (13), and (15), as well as constructing the subsets of fuzzy clusters with extreme values of these indexes, are the subject of the second step of the proposed technique. The main characteristic features of the classification obtained by applying the second step of the proposed technique are summarized in Table 3.

Table 3. Characteristic features of the results obtained using the proposed technique

Characteristics of fuzzy clusters in the allotment obtained	Values of the characteristic features obtained using							
	distance (19)		distance (20)		similarity measure (21)		dissimilarity measure (22)	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Index (10)	0.0000	0.6104	0.0000	0.4900	0.0000	0.6252	0.1923	0.7103
Index (13)	0.0000	0.7476	0.0000	0.6001	0.0000	0.7658	0.1923	0.7449
Index (15)	1.0000	0.6385	1.0000	0.5783	1.0000	0.6459	0.9038	0.5575

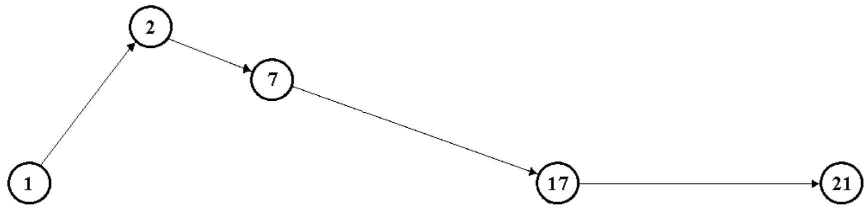

Fig. 4. The unique selected path in the minimized project-network

The values of the density of a fuzzy cluster, based on equation (15), are maximal and the values of the modified indexes of fuzziness, based on Eequations (10) and (13), are minimal for the first fuzzy cluster in all the experiments. Hence, the first fuzzy cluster should be selected as the unique element of the subset $\{A^l_{(\alpha)}\} \subset R_c^*(X)$ of the allotment obtained, $R_c^*(X)$, among two fuzzy clusters and its typical point τ^1 is the unique element of the minimal subset of paths. Choosing the typical point is the subject of the third step of the proposed heuristic technique. Thus, the path selected, x_1, is shown in Fig. 4.

Moreover, the first path can be considered as an outlier [11] in the clustering structure of the set $X = \{x_1, \ldots, x_4\}$ of paths in the minimized project-network, because this path is the unique element of the first fuzzy cluster.

Hence, the results obtained for the numerical example seem to be satisfactory for distances (18), (19), similarity measure (20) and dissimilarity measure (21). It should be noted that the unique selected path includes five vertices, while the other paths include six vertices.

4. Conclusions

A technique has been described for categorizing paths in a minimized project-network with given bounds on the durations of activities, where lower and upper bounds on the possible durations of activities are given at the stage of project planning. The proposed technique can be summarized as a three-step procedure where a minimized project-network is represented as a matrix of interval-valued data and the minimal subset of paths is derived from the subset of dominant paths using heuristic possibilistic clustering. Hence, either a unique path or all the paths from a minimized project-network will be extreme cases of solutions to the classification task.

Summarizing, we should note that the derived minimal subset of dominant paths contains all the necessary information about the critical paths in the initial network $G^{(*)}$ and may be used at the stage of project control.

Some other ways of deriving the minimal number of paths in a minimized project-network with given bounds on the durations of activities can be investigated. Firstly, a methodology for outlier detection in an interval-valued data set [11] can be applied to the set of paths in the minimized project-network and this set of outliers can be interpreted as a solution of the problem. Secondly, the set of all paths $X = \{x_1, \ldots, x_n\}$ in the graph $G = (X^{t_1}, E)$ from vertex $x^1 \in X^{t_1}$ to vertex $x^{m_1} \in X^{t_1}$ can be considered as a set of alternatives. If a weak fuzzy preference relation is given on *X*, then the fuzzy set of non-dominated alternatives can be constructed and an alternative can be selected according to the method proposed by Orlovsky [8]. Hence, a method of constructing an appropriate weak fuzzy preference relation on the set of paths in a minimized project-network with given bounds on the durations of activities should be developed.

These perspectives for future investigations are of great interest from both a theoretical and practical point of view.

Acknowledgements

I would like to thank Prof. Yuri N. Sotskov for his help in formulating the problem and fruitful discussions during the preparation of the paper. I am grateful to Prof. Janusz Kacprzyk and Prof. Jan W. Owsinski for their interest in my investigations and support. I also thank the anonymous referees for their valuable comments.

References

[1] BRUCKER F., DREXL A., MÖHRING R., NEUMANN K., PESCH E., *Resource-constrained project scheduling: notation, classification, models, and methods*, European Journal of Operation Research, 1999, 112 (1), 3.

[2] BURILLO P., BUSTINCE H., *Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets,* Fuzzy Sets and Systems, 1996, 78 (3), 305.

[3] GRZEGORZEWSKI P., *Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on Hausdorff metric*, Fuzzy Sets and Systems, 2004, 148 (2), 319.

[4] JU H., YUAN X.H., *Similarity measures on interval-valued fuzzy sets and application to pattern recognition*, [in:] D.Y. Cao (Ed.), *Fuzzy Information and Engineering*, Springer-Verlag, Heidelberg 2007, 384.

[5] KAUFMANN A., *Introduction to the Theory of Fuzzy Subsets*, Academic Press, New York 1975.

[6] KREINOVICH V., KOSHELEVA O., *Towards dynamical systems approach to fuzzy clustering*, [in:] D.A. Viattchenin (Ed.), *Developments in Fuzzy Clustering*, VEVER Publishing House, Minsk 2009, 10.

[7] LAI T.-C., SOTSKOV Y.N., *Sequencing with uncertain numerical data for makespan minimization*, Journal of the Operations Research Society, 1999, 50 (3), 230.

[8] ORLOVSKY S.A., *Problems of Decision Making for Fuzzy Initial Information*, Nauka, Moscow 1981 (in Russian).

[9] SOTSKOV Y.N., SHILAK A.N., *Minimization of project-network with given bounds on permissible values of activity durations*, Technical Report No. 2, Institute of Engineering Cybernetics of the National Academy of Sciences of Belarus, Minsk 2000 (in Russian).

[10] SOTSKOV Y.N., SHILAK A.N., *Minimization of project-network with given bounds on activity durations*, [in:] B. Fleischmann, R. Lasch, U. Dengs, W. Domschke (Eds.), Operations Research Proceedings 2000, Springer-Verlag, Heidelberg 2001, 384.

[11] VIATTCHENIN D.A., *Detecting outliers in interval-valued data using heuristic possibilistic clustering*, Journal of Computer Science and Control Systems, 2012, 5 (2), 39.

[12] VIATTCHENIN D.A., *A Heuristic Approach to Possibilistic Clustering. Algorithms and Applications*, Springer-Verlag, Heidelberg 2013.

Received 23 March 2015
Accepted 30 November 2015