

THE PRODUCTION QUALITY MANAGEMENT BASED ON THE PRECEDENT APPROACH AND THE CLUSTERING OF USE CASES

Timur N. Karimov¹Shamil A. Khamadeev²

Abstract: In this work the problem of various discrepancies knowledge management is considered during realization of technological production processes of polypropylene tubes. Such discrepancies can make serious impact on manufacturing efficiency enterprise for the reasons of the compelled equipment stand still, sharp decline in final product quality, failures to meet time constraints of production shipment for the consumer, etc. In order to remove discrepancies it is required to define quickly the major factors exerting negative impact on production quality. And for each type of a product there can be the set of factors. The main way for work with the large volume of information concerning problems, their factors and ways of elimination is experience of experts. Such way owing

to a human factor is not reliable and cannot be considered as the effective solution of the considered problem. In article as the decision the structure of the knowledge base on the basis of precedents (use cases) is offered. The precedent represents the information block including a basic situation and the decision corresponding to it. The offered structure is founded on hierarchy to Isikava's chart, one of popular instruments of quality control, and listed products. For filling of base precedents it is offered to use an algorithm of a clustering of data CLOPE. Results of work are the three-level structure of the knowledge base, model of a precedent, model of processes of addition of a new precedent and search of a precedent in the knowledge base, an algorithm of a clustering of precedents. It was revealed

¹ Kazan Federal University, Naberezhnye Chelny Institute. e-mail: shamil.hamadeev@mail.ru. +79272447671.

² Kazan Federal University, Naberezhnye Chelny Institute. e-mail: shamil.hamadeev@mail.ru. +79272447671

that the preliminary clustering allows reducing search time considerably. This approach can be used at a stage of technological preparation of production.

Keywords: Knowledge base, case approach, quality management, technological preparation of production, clustering, Isikava's chart

Introduction

Effective work of the modern industrial enterprise is impossible without the automated information systems. In the presence of a common information space the similar systems covering everything functions of the enterprise carry out information support of products at all stages of life cycle that allows to reduce expenses, to reduce terms of technological preparation of production, to quickly fix the problems arising in the course of production, to reduce prime cost of products [1].

Within a stage of preparation of production at inclusion in the production plan of a new product engineers need to predict all possible problems which can arise in the course of its production. Success of this forecasting directly influences success of functioning of all

production in general. This task is uncommon as demands the analysis of a large number of the potential reasons influencing quality of the made product that is poor-selling at the manual mode of processing of basic data.

Objective of this research is development of a technique of support of decision-making on the basis of precedents on elimination of the discrepancies revealed during production of polyamide pipes and drawing polyamide coverings, and also molding of polyamide pipes, production of spiral pipelines and spiral polyethylene tapes.

Methods

For the purpose of increase in efficiency of elimination of production problems support of decision-making on the basis of the knowledge base on precedents is offered. The precedent (Use Case) in this work is understood as the cluster comprising rules for decision-making on elimination of similar production problems at similar products [2]. Thus, the precedent will comprise information on several problems, and on all products to most of which likely there

can be these problems. The example of structural representation of a precedent is represented in the figure 1. The hierarchical structure of the chart of

Isikava [3] which is a part of "seven basic instruments of quality management" is the cornerstone of representation.

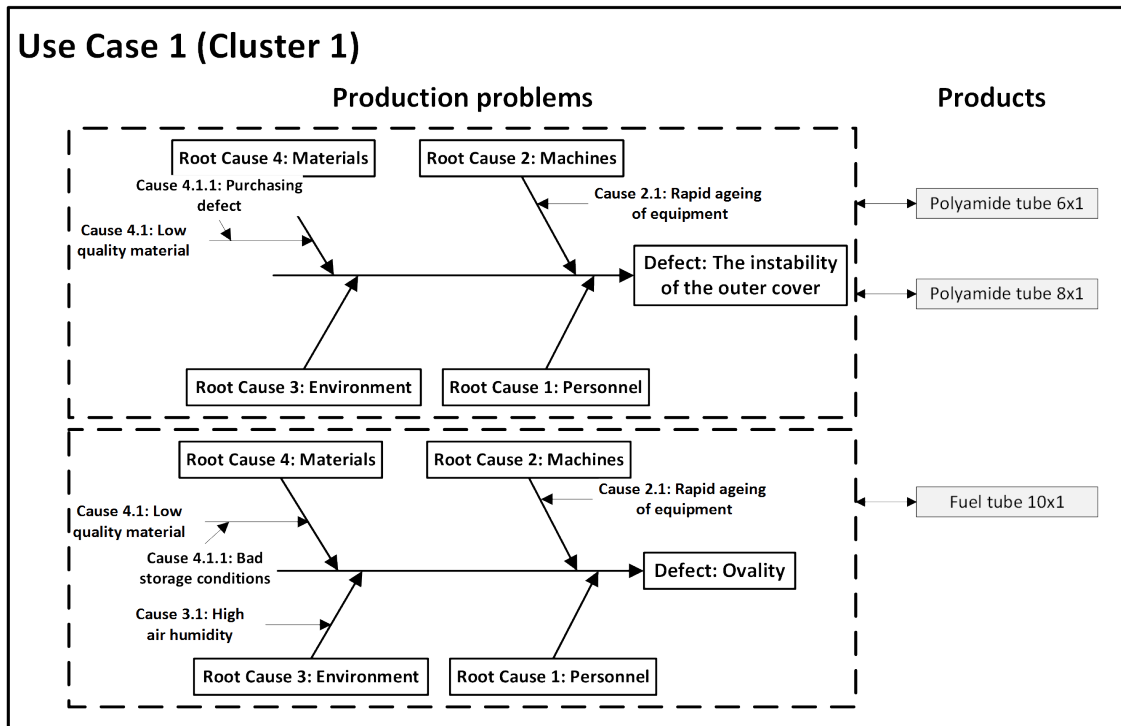


Fig. 1. An example of the structure of precedent

From this example it is possible to draw several conclusions concerning productions at the enterprise:

- 1) Problems of ovality of products and instability of an external covering have 2 general reasons: swift wear of the equipment and low qualities of material at a preparation stage. Elimination of these two reasons will perhaps fix both problems.
- 2) Instability of an external covering can also cause a problem and the increased humidity of air.
- 3) The fuel tube of dimension 10x1 can have a problem of instability of an external covering, and polyamide tubes dimension 6x1 and 8x1 can have an ovality problem.

On the basis of the analysis of subject domain the general structure of

formation of the knowledge base on precedents presented in the figure 2 [1] was developed.

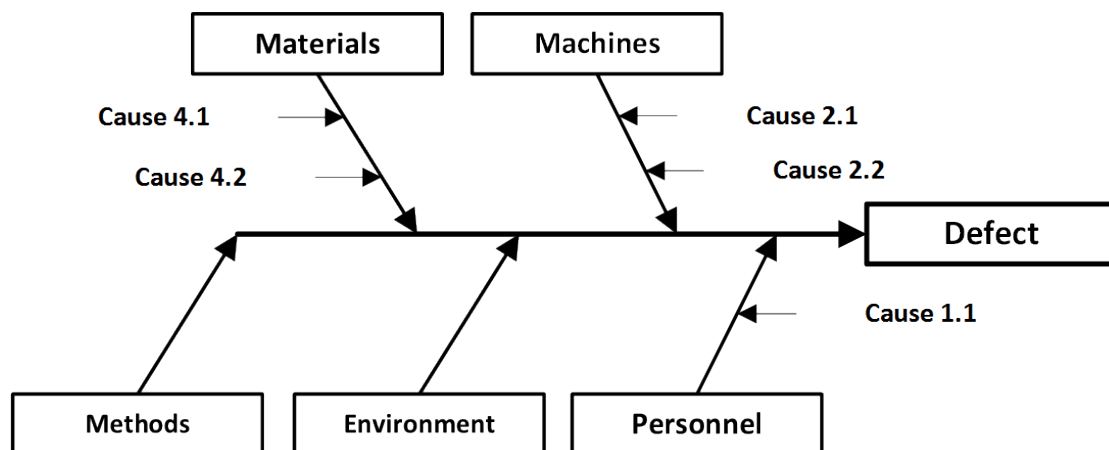


Fig. 2. An example of the structure of Ishikawa diagrams

Thus, the structure of the description of problems can be presented the following levels of hierarchy:

- the production problem (PP), each problem can cause a number of the general reasons;
- the factor (RC) of emergence of production problems, each factor comprises a variety of reasons;
- the reason (C), can characterize emergence of a problem at the lowest level, or consist of a number of more concrete characteristics (subreasons).

Problems can arise at a number of products (Pr) or services which are in turn described by a set of characteristics (SPC).

Structural division of a production problem (PP) into hierarchical levels was assumed as a basis (RC, C, Pr, SPC). Thus, each hierarchical level breaks into the corresponding classes, and the PP model represents treelike hierarchical structure.

Classification of precedents in base of precedents (PB) is offered to be organized in the similar way - to divide the general reasons into more concrete components according to structure of PP.

Production problems are the crossed sets as concrete PP can have the identical reasons which caused them which define their crossing.

In turn, each RC represents a great number of C. The subreasons are also the crossed subsets. Similarly there is a splitting all PB at all PP levels. Thus, a set of all values which are kept in PB it

is possible to organize in this way where each concrete vector of values represents a subset element. Basic data are presented in table 1.

Table 1. The source data table

PP	RC	C	Pr	SPC
Name of production problem	Factors (Root causes)	Causes	Products that can have relation with problems	Product specifications

Parameters and values of structural elements of the table of basic data for each enterprise are unique depending on the nature of activity.

According to Case-based Reasoning approach (CBR, reasoning on the basis of precedents [4,5]) the precedent includes the following information:

- heading;
- table of conditions;
- table of decisions.

The heading represents a precedent name. The table of conditions contains the description of input parameters (the table of basic data) and contains the following fields: name of a

condition and their value. Values of conditions can be presented in the form of concrete values or their intervals. The table of decisions represents given in output parameters of a precedent - the ranged list of the reasons of a production problem [6].

It is possible to allocate two main working hours of base of precedents:

- filling of base of precedents;
- use of base of precedents.

Filling of base of precedents is carried out in the absence of a similar precedent by search. Process of filling is presented in the chart BPMN2 form in the figure 3.

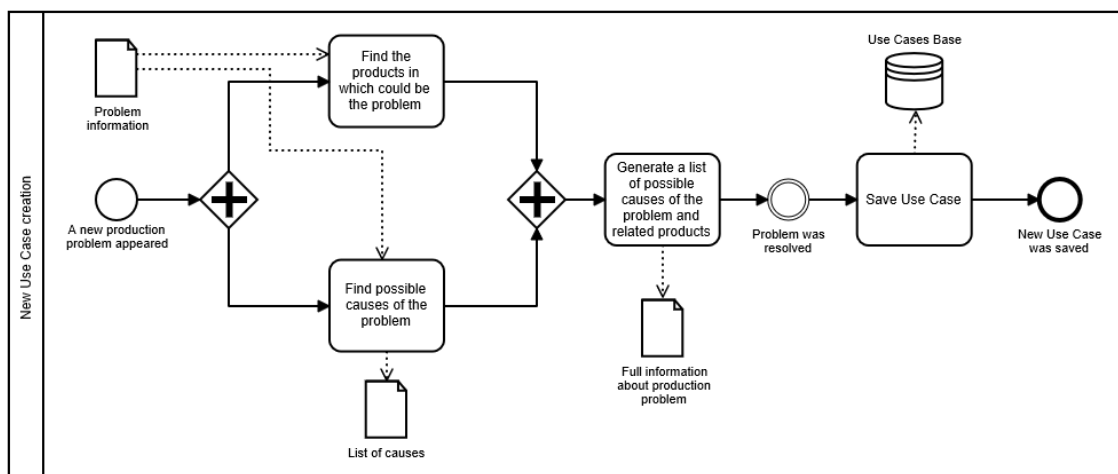


Fig. 3. The process of creating a new use case

Process is started by an event "A new production problem". Is followed by the description of a problem, presented in the form of a matrix of basic data. On the basis of basic data and according to similar already known problems the list of the possible reasons of a problem is created. The second action is definition of products which have an emergence of this problem most

likely. Final information on a problem is formed on the basis of the found most probable causes of a problem and observations of experts. Full information on a production problem is result. The received decision remains as a precedent in base of precedents. In the figure 4 process of use of base of precedents is presented.

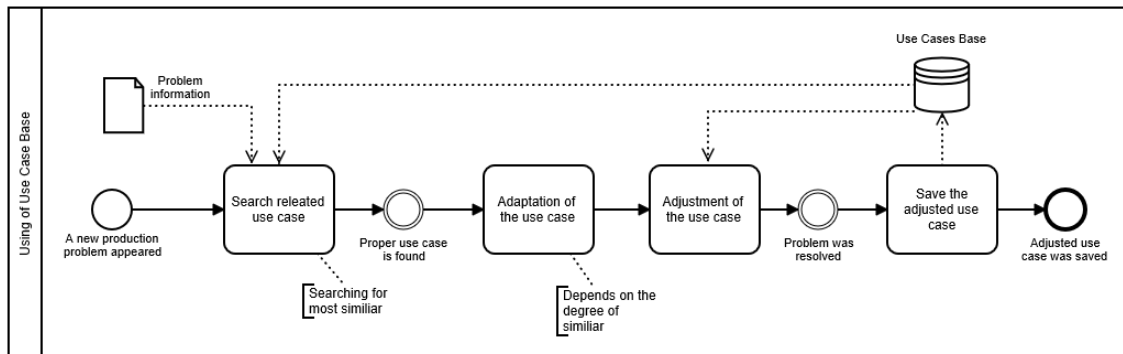


Fig. 4. The process of using Use case database

On value of the table of basic data of a new production problem search of a similar precedent in base is carried out. In case of lack of full coincidence the closest precedent gets out of base. Depending on degree of proximity of a precedent and the description of a new problem adaptation of a precedent, i.e. creation of a new precedent on the basis of existing is made. Further the adapted precedent is compared to data of base of precedents, and on the basis of this comparison, the new precedent is corrected. The corrected precedent remains in base.

For increase in efficiency of search of precedents in base their cluster splitting was offered. After the analysis of methods of a clustering [7-9] as an algorithm the algorithm CLOPE was chosen. It best of all is suitable for a clustering of large volumes of category

data. Cluster methods are the most suitable for the analysis of difficult multidimensional objects, do not demand preliminary structuration of information, that is it is possible to classify objects, without having data on their preliminary splitting (the analysis of unstructured set of precedents) [10, 11].

Let there is a base of precedents D consisting of a set of precedents $\{p_1, \dots, p_n\}$. Each precedent is characterized by a set of unique indicators $\{i_1, \dots, i_m\}$ which values are presented in table 2. The set of clusters $\{C_1, \dots, C_k\}$ are splitting a set $\{p_1, \dots, p_n\}$, it, $C_1 \cup \dots \cup C_k = \{p_1, \dots, p_n\}$ as $C_1 \supset \emptyset$ and $C_i \cap C_j = \emptyset \forall i \geq 1, k \geq j$.

Each element C_i is called a cluster, and n, m, k - the number of

precedents, quantity of indicators in base of precedents and number of clusters respectively.

Table 2. Values of product attributes

Attributes	Values of attributes					
	1	2	3	4	5	...
Product	a_1	a_2	a_3	a_4	a_5	...
Size type	b_1	b_2	b_3	b_4	b_5	...
Defect	c_1	c_2	c_3	c_4	c_5	...
Root Cause	d_1	d_2	d_3	d_4	d_5	...
Cause p_1	e_1	e_2	e_3	e_4	e_5	...
Cause p_2	f_1	f_2	f_3	f_4	f_5	...
Cause p_3	g_1	g_2	g_3	g_4	g_5	...
Cause p_4	h_1	h_2	h_3	h_4	h_5	...
Cause p_5	i_1	i_2	i_3	i_4	i_5	...

Each cluster C has the following characteristics:

$D(C)$ - set of unique indicators;
 $Occ(i, C)$ - the number of occurrences (frequency) of an indicator i into a cluster C ;

$$S(C) = \sum_{i \in D(C)} Occ(i, C) = \sum_{t_i} |p_i| \quad (1)$$

$$W(C) = |D(C)| \quad H(C) = S(C) / W(C) \quad (2)$$

The histogram of a cluster C represents the graphic representation of its settlement characteristics: on an axis of OX cluster objects in decreasing order of size $Occ(i, C)$, and size $Occ(i, C)$ - on OY axis are postponed $Occ(i, C)$. The histogram is submitted in the figure 5.

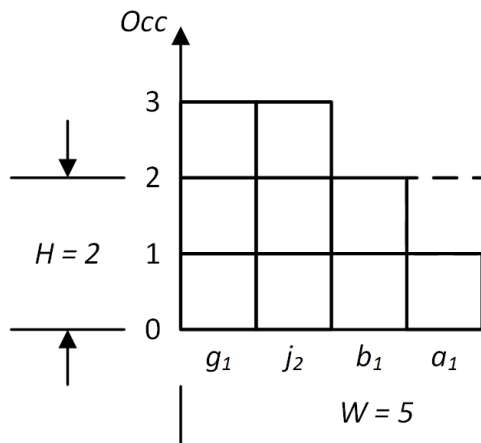


Fig. 5. Cluster histogram

In the figure 5 $S(C)$, equal 10, corresponds to the area of the rectangle limited to axes of coordinates and a dashed line. The more value H , the bigger similarity between two products. Therefore the algorithm chooses such splitting which maximizes H .

Formula of calculation of global criterion - function of cost:

$$Profit(C) = \frac{\sum_{i=1}^k G(C_i) \cdot |C_i|}{\sum_{i=1}^k |C_i|} = \frac{\sum_{i=1}^k \frac{S(C_i)}{W(C_i)^r} \cdot |C_i|}{\sum_{i=1}^k |C_i|},$$

(3)

where $|C_i|$ - quantity of objects in i - m a cluster, k - quantity of clusters, r - pushing away coefficient.

The coefficient r allows regulating the level of similarity of precedents in a cluster, and, as a result, final quantity of clusters. This coefficient is selected the expert. The more r , the level of similarity is lower and the more clusters will be generated.

Formal problem definition of a clustering an algorithm CLOPE looks as follows: for set D and r to find splitting $C : Profit [11](C, r) \rightarrow \max$.

Results and Discussion

Let's review an example from 5 transactions presented in table 3.

Table 3. The Example data for clusterization

Nº	Case	Values of attributes
----	------	----------------------

		Product name	Size type	Numbers of ply	Defect
1	A	Polyamide tube	10x1	5	External diameter instability
2	B	Polyamide tube	6x1	3	Internal diameter instability
3	C	Fuel tube	-	5	Low strength
4	D	Spiral tape	22x18x19	-	Low strength
5	E	Fuel tube	-	3	Internal diameter instability

Let's assign for each unique value of attribute a code and we will change the table with data:

Table 4. The Modified Example data for clusterization

№	Case	Values of attributes			
		Product name	Size type	Numbers of ply	Defect
1	A	a1	b1	c1	d1
2	B	a1	b2	c2	d2
3	C	a2	-	c1	d3
4	D	a3	b3	-	d3
5	E	a2	-	c2	d2

It is required to compare among themselves two splittings into clusters:

Table 5. 2 options of cluster division

Option	Cluster 1	Cluster 2
1	A(a1,b1,c1,d1);C(a2,c1,d3);D(a3,b3,d3)	B(a1,b2,c2,d2);E(a2,c2,d2)

2	A(a1,b1,c1,d1);B(a1,b2,c2,d2)	C(a2,c1,d3);D(a3,b3, d3);E(a2, c2,d2)
---	-------------------------------	---------------------------------------

For the first and second options of splitting in each cluster it is required to calculate the number of entries into it each indicator of a product, then to calculate height H and width W of a cluster. Histograms of clusters of the first splitting are submitted in the figure 6, and the second in the figure 7.

The histogram of a cluster { A, C,D } contains 8 various elements ($W=8$) and has the area equal 10 ($S=10$), and a cluster { C, D,E } - 7 various elements ($W=7$) with an area of 9 ($S=9$); that is indicators of clusters differ insignificantly.

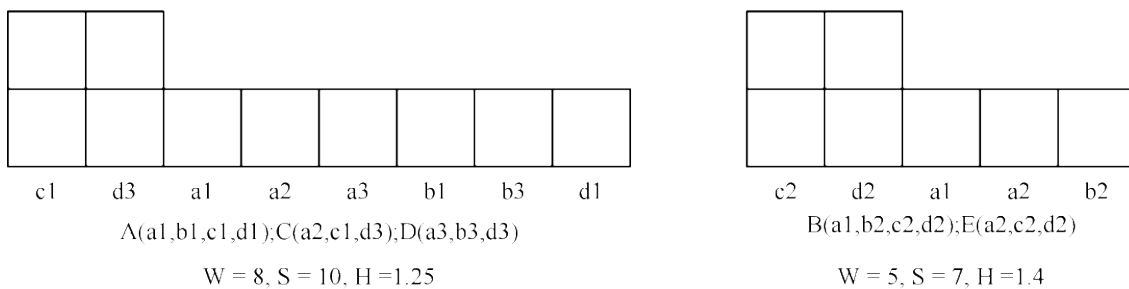


Fig. 6. First option histogram

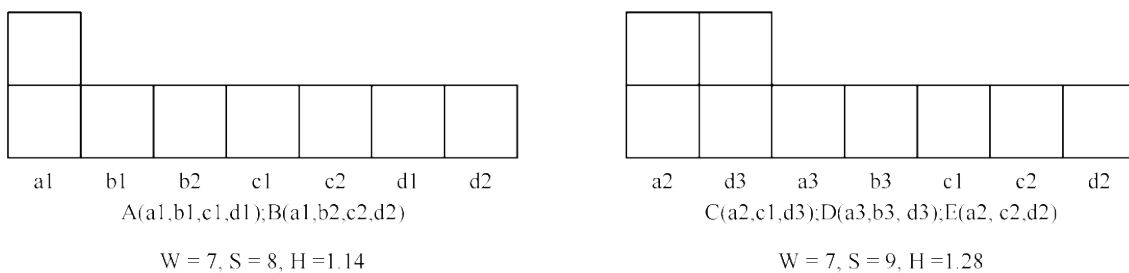


Fig. 7. Second option histogram

But when comparing two other clusters the difference of two splittings is visible. The area of a cluster { B, E } is equal 7 ($S=7$) at 5 various elements, and the area of a cluster { C, D,E } is equal 9 ($S=9$) at 7 various elements ($W=7$). It is

obvious that splitting 1 is better as provides bigger imposing of transactions at each other (respectively a parameter H higher there). Realization of an algorithm of a clustering is presented in the flowchart form in the figure 8.

The following designations are presented in the flowchart: n - the number of precedents in base, r - pushing away coefficient, P_m - m -th precedent, k - quantity of clusters $C[i]$, $C[j]$ - i -th and j -th clusters, C_{temp} - a temporary cluster, $Profit(C, r)$ - cluster cost on a formula (3), h_m - characterizes accessory m precedent to i cluster.

The algorithm consists of two main phases: initialization and iteration. For work of an algorithm it is necessary to enter two values: n and r . The first precedent is brought in the first cluster at once ($C1 = bypl$), the counter of clusters is equated to unit ($k = 1$), the counter of precedents - to two ($m = 2$). Then the first cycle in which all precedents of base get over $m \leq n$ (so far) follows $m \leq n$. It

is entered m -th precedent, added in i -th cluster, calculation of value of function of the cost ($Profit$) for a cluster is made i . In case the cost of a cluster increases, then the precedent is added to a cluster, and the algorithm passes to the next precedent ($m = m + 1$). Otherwise, the algorithm checks change of the following cluster at addition of a precedent in it m . If any cluster does not increase value of the cost at addition of a precedent, then the new cluster is created.

For each precedent m record h_m in which accessory m precedent to a new cluster is specified is created h_m . The phase of initialization is characterized by distribution of precedents on clusters.

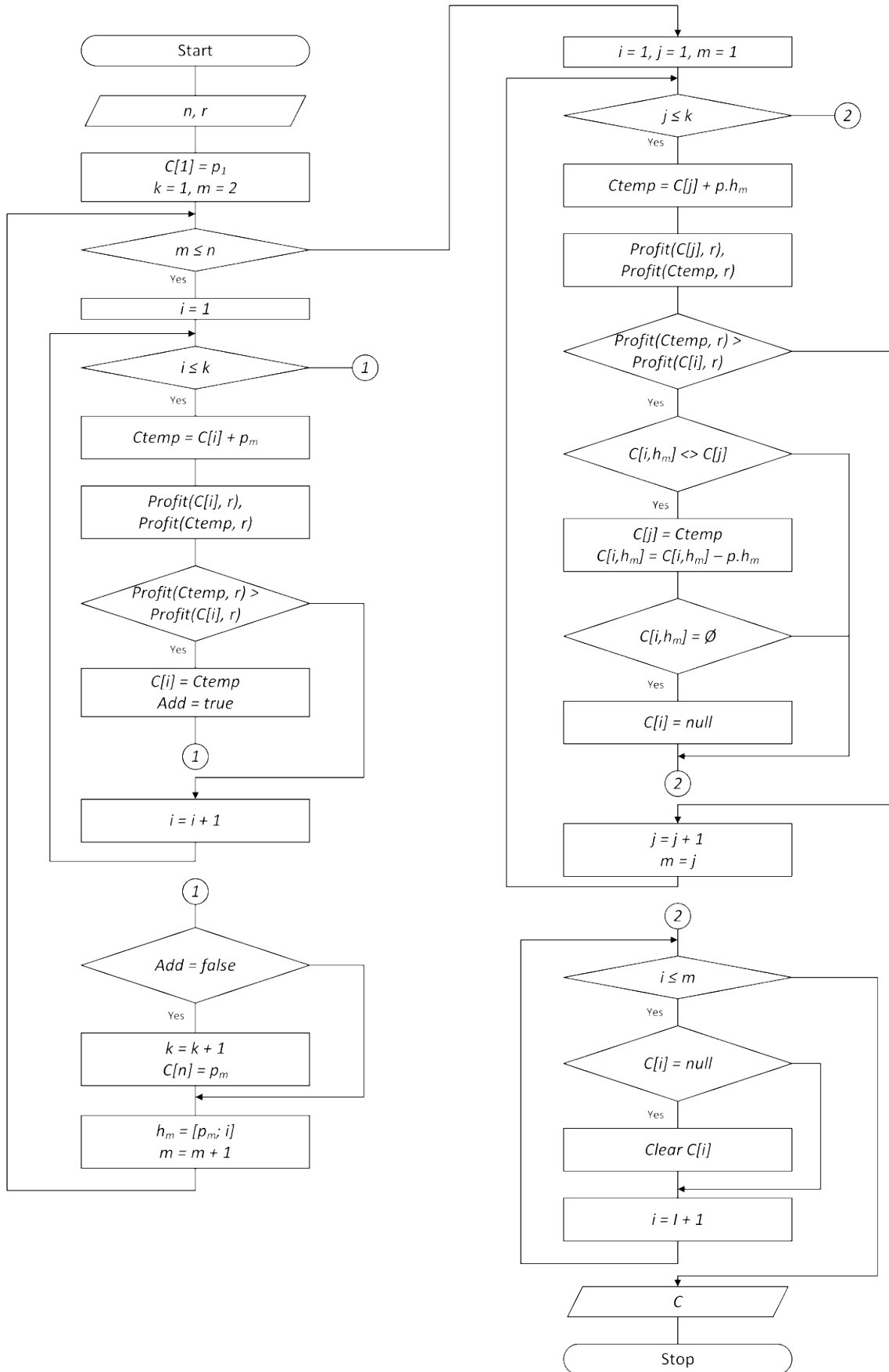


Fig. 8. Clustering Algorithm

For improvement of quality of a clustering it is necessary to carry out several additional scanning's - phases of iterations. Precedents are verified with other clusters and, in case of increase in cost of a cluster of a new design the precedent is postponed to a new cluster, and from old is removed. Empty clusters ($C [i] = null$) in conclusion of an algorithm are removed.

Summary

Temporary complexity of one iteration is equal $N \cdot K \cdot A$ where N - total number of precedents, K - the maximum possible number of clusters, A - the average length of a precedent. Speed of work of an algorithm grows linearly with increase in the size of massifs of basic data that characterizes it as swift and effective on large volumes.

During realization of an algorithm CLOPE in the Java language the following advantages were sun-dried:

- minimum quantity of addressing the database;
- use of limited volume of random access memory;

- low computing complexity;
- scalability.

It is possible to refer complexity with selection of coefficient of pushing away to insignificant shortcomings. But for elimination of this complexity the variation of an algorithm Fuzzy CLOPE [12] is used.

Conclusion

The clustering technique on the basis of an algorithm CLOPE allows to increase efficiency of work with base of precedents and, unlike other methods of a clustering, allows working with large volumes of category data. Use of the integrated approach based on accumulation of knowledge of precedents will allow increasing efficiency of process of technological preparation of production [2, 6].

ACKNOWLEDGEMENTS

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References

Valiev R.A., Khairullin A.Kh., Shibakov V.G. Automated Design Systems for Manufacturing Processes//Russian Engineering Research, 2015. - No. 35(9). - p. 662-665.

Hamadeev Sh. And, Bukharov of S. I. Baz of solutions knowledge of elimination of the discrepancies revealed during work of LLC Tissan// "The Kama readings. Collection of materials of the 2nd interregional scientific and practical conference of students, graduate students and young scientists". - Chelny Emb.: Publishing house Kama State Ing.-Ec. Acad.-2010. - page 123 - 126.

Lysenko of E. The automated synthesis of models of technological processes on the basis of case approach//Scientific sheets of the Belgorod state university. Series: Economy. Informatics. 2014. T. 29. No. 1-1 (172). Page 121-129.

Ishikawa, Kaoru (1990); (Translator: J. H. Loftus); Introduction to Quality Control; 448 p.

Yeremeyev, A.P., Warsaw, P. R. Search of the decision on the basis of structural analogy for intellectual systems of

support of decision-making//News of RAS. Theory and control systems. 2005. No. 1. Page 97-109.

A. Aamodt, E. Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications. IOS Press, 1994, Vol. 7: 1.

Hamadeev Sh. A., Simonova L. A., Ilyukhin of A. K. Baz of precedents of technological routes of forming production within MES systems//Forge and forming production. Processing of materials pressure. 2009. No. 8. Page 29-35.

Jain A., Murty M., Flynn P. Data Clustering: A Review//ACM Computing Surveys, 1999. Vol. 31, No. 3.

Yang J, Leskovec J. Defining and evaluating network communities based on ground-truth. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. Beijing, China; 2012. p. 1-10.

Rui Xu, D. Wunsch. Survey of clustering algorithms. IEEE Transactions on Neural Networks, 2005. Vol. 16, No. 3. p. 645-678

Paklin, N. Klasterization of category data: scalable algorithm CLOPE.//BaseGroup Labs - Technologies of the analysis of data. - 2004. URL: <https://basegroup.ru/community/article/s/clope>, free. It is checked 14.04.2017.

Yang, Y., Guan, H., You. J. CLOPE: A fast and Effective Clustering Algorithm for Transactional Data In Proc. of SIGKDD '02, July 23-26, 2002, Edmonton, Alberta, Canada.

Li Jie, Gao Xinbo, Jiao Licheng. A fuzzy CLOPE algorithm and its optimal parameter choice//Journal of Electronics (China), 2006. Vol. 23, No. 3. p. 384-388