

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

Timur N. Karimov, Kazan Federal University, Naberezhnye Chelny Institute

Shamil A. Khamadeev, Kazan Federal University, Naberezhnye Chelny Institute, e-mail: shamil.hamadeev@mail.ru, тел. +79272447671

Аннотация

В данной работе рассмотрена проблема управления знаниями по различным несоответствиям в ходе реализации технологических процессов производства полипропиленовых трубок. Такие несоответствия могут оказывать серьезное воздействие на эффективность производственного предприятия по причинам вынужденного останова оборудования, резкого снижения качества конечного изделия, срывов сроков отгрузки продукции потребителю и т.д. Для устранения несоответствий требуется оперативно определить основные факторы, оказывающие негативное влияние на качество производственного процесса. Причем для каждого типа изделия может быть свое множество факторов. Основным способом для работы с большим объемом информации относительно проблем, их факторов и способов устранения является опыт специалистов. Такой способ в силу человеческого фактора не является надежным и не может считаться эффективным решением рассматриваемой проблемы. В статье в качестве решения предлагается структура базы знаний на основе прецедентов (use cases). Прецедент представляет собой информационный блок, включающий в себя базовую ситуацию и соответствующее ей решение. Предложенная структура основана на иерархии диаграмме Исикавы, одного из популярных инструментов контроля качества, и номенклатуре изделий. Для наполнения базы прецедентами предложено использовать алгоритм кластеризации данных CLOPE. Результатами работы являются трехуровневая структура базы знаний, модель прецедента, модели процессов добавления нового прецедента и поиска прецедента в базе знаний, алгоритм кластеризации прецедентов. Было выявлено, что предварительная кластеризация позволяет значительно сократить

время поиска. Данный подход может быть использован на этапе технологической подготовки производства.

Ключевые слова: База знаний, прецедентный подход, управление качеством, технологическая подготовка производства, кластеризация, диаграмма Исикавы

Введение

Эффективная работа современного промышленного предприятия невозможна без автоматизированных информационных систем. При наличии единого информационного пространства подобные системы, охватывающие все функции предприятия, осуществляют информационную поддержку изделий на всех этапах жизненного цикла, что позволяет сокращать затраты, уменьшать сроки технологической подготовки производства, оперативно устранять проблемы, возникающие в процессе производства, снижать себестоимость выпускаемой продукции [1].

В рамках этапа подготовки производства при включении в производственный план нового изделия инженерам требуется прогнозировать все возможные проблемы, которые могут возникнуть в процессе его производства. Успешность данного прогнозирования напрямую влияет на успешность функционирования всего производственного процесса в целом. Данная задача является нетривиальной, так как требует анализа большого числа потенциальных причин, влияющих на качество изготавливаемого изделия, что является труднореализуемым при ручном режиме обработки исходных данных.

Целью данного исследования является разработка методики поддержки принятия решений на основе прецедентов по устранению несоответствий, выявленных в ходе производства полиамидных труб и нанесения полиамидных покрытий, а также формовки полиамидных труб, изготовления спиральных трубопроводов и спиральных полиэтиленовых лент.

Методы

С целью повышения эффективности устранения производственных проблем предлагается поддержка принятия решений на основе базы знаний по прецедентам. Под прецедентом (Use Case) в данной работе понимается кластер,

содержащий в себе правила для принятия решения по устранению схожих производственных проблем у схожих изделий [2]. Таким образом, прецедент будет содержать в себе информацию о нескольких проблемах, и обо всех изделиях, с которыми вероятнее всего могут произойти эти проблемы. Пример структурного представления прецедента изображен на рисунке 1. В основе представления лежит иерархическая структура диаграммы Исикавы [3], входящей в состав «семи базовых инструментов управления качеством».

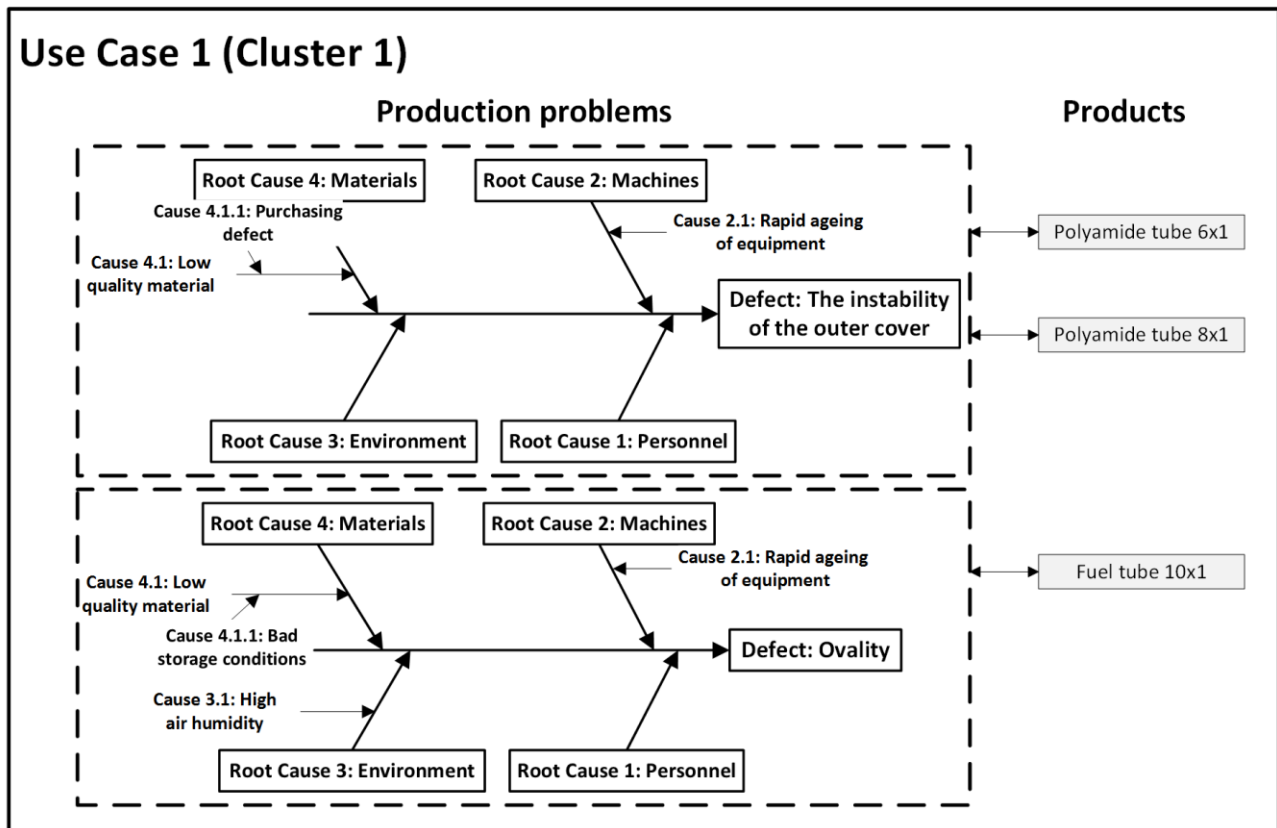


Fig. 1. An example of the structure of precedent

Из данного примера можно сделать несколько выводов относительно производственных процессов на предприятии:

- 1) У проблем овальности изделий и нестабильности наружного покрытия есть 2 общие причины: быстрый износ оборудования и низкое качества материала на стадии заготовки. Устранение этих двух причин возможно устранит обе проблемы.
- 2) Проблему нестабильность наружного покрытия может вызывать так же и повышенная влажность воздуха.

3) У топливной трубки размерности 10x1 может возникнуть проблема нестабильности наружного покрытия, а у полиамидных трубок размерностью 6x1 и 8x1 может возникнуть проблема овальности.

На основе анализа предметной области была разработана общая структура формирования базы знаний по прецедентам, представленная на рисунке 2 [1].

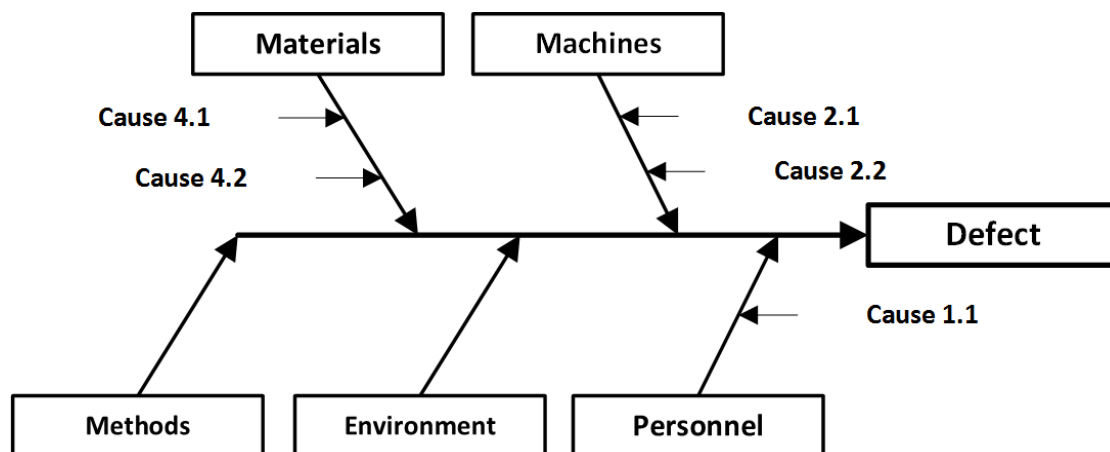


Fig. 2. An example of the structure of Ishikawa diagrams

Таким образом, структуру описания проблем можно представить следующими уровнями иерархии:

- производственная проблема (PP), каждую проблему может вызвать ряд общих причин;
- фактор (RC) возникновения производственных проблем, каждый фактор содержит в себе ряд причин;
- причина (C), может характеризовать появление проблемы на низшем уровне, или же состоять из ряда более конкретных характеристик (подпричин).

Проблемы могут возникнуть у ряда изделий (Pr) или услуг, которые в свою очередь описываются набором характеристик (SPC).

За основу было принято структурное разделение производственной проблемы (PP) на иерархические уровни (RC, C, Pr, SPC). Таким образом, каждый иерархический уровень разбивается на соответствующие классы, и модель PP представляет собой древовидную иерархическую структуру.

Классификацию прецедентов в базе прецедентов (PB) предлагается организовать аналогичным способом — разделять общие причины на более конкретные составляющие в соответствии со структурой PP.

Производственные проблемы являются пересекающимися множествами, так как конкретные PP могут иметь одинаковые вызвавшие их причины, которые и определяют их пересечение.

В свою очередь, каждый RC представляет собой множество C. Подпричины также являются пересекающимися подмножествами. Аналогично происходит разбиение всей PB на всех уровнях PP. Таким образом, множество всех значений, хранящихся в PB, можно организовать подобным образом, где каждый конкретный вектор значений представляет собой элемент подмножества. Исходные данные представлены в таблице 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

Параметры и значения структурных элементов таблицы исходных данных для каждого предприятия являются уникальными в зависимости от характера деятельности.

В соответствии с подходом Case-based Reasoning (CBR, рассуждения на основе прецедентов [4,5]) прецедент включает в себя следующую информацию:

- заголовок;
- таблицу условий;
- таблицу решений.

Заголовок представляет собой имя прецедента. Таблица условий содержит описание входных параметров (таблица исходных данных) и содержит следующие поля: наименование условия и их значения. Значения условий могут представляться в виде конкретных значений или их интервалов. Таблица решений представляет данные по выходным параметрам прецедента — ранжированный список причин производственной проблемы [6].

Можно выделить два основных режима работы базы прецедентов:

- наполнение базы прецедентов;
- использование базы прецедентов.

Наполнение базы прецедентов выполняется при отсутствии схожего прецедента при поиске. Процесс наполнения представлен в виде диаграммы BPMN2 на рисунке 3.

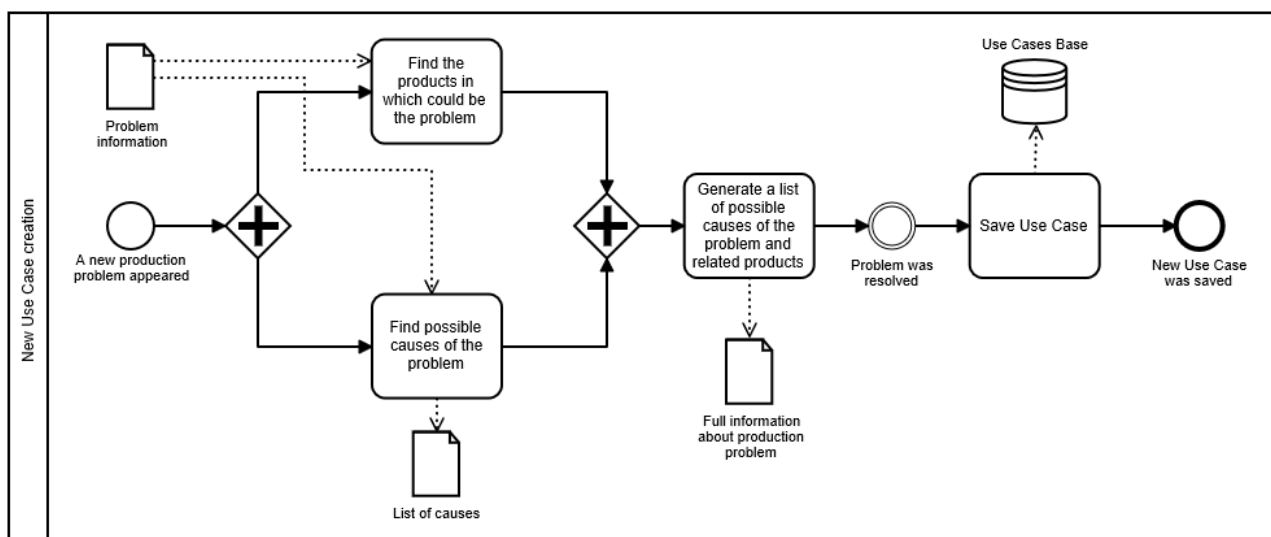


Fig. 3. The process of creating a new use case

Процесс запускается событием «Новая производственная проблема». Сопровождается описанием проблемы, представленное в виде матрицы исходных данных. На основании исходных данных и в соответствии с похожими уже известными проблемами создается список возможных причин проблемы. Вторым действием является определение изделий, у которых возникновение данной проблемы вероятнее всего. Итоговая информация о проблеме формируется на основе найденных наиболее вероятных причинах проблемы и наблюдений специалистов. Результатом является полная информация о производственной проблеме. Полученное решение сохраняется как прецедент в базе прецедентов. На рисунке 4 представлен процесс использования базы прецедентов.

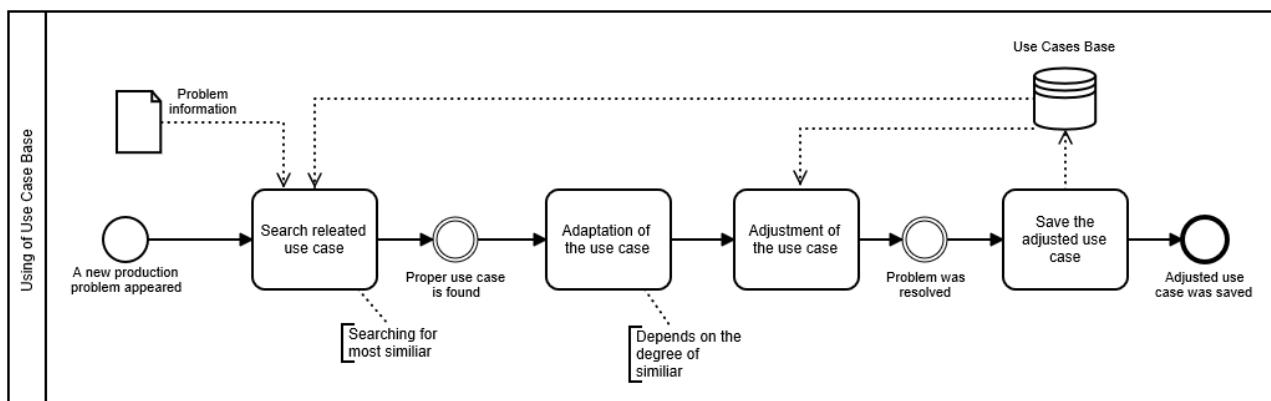


Fig. 4. The process of using Use case database

По значению таблицы исходных данных новой производственной проблемы осуществляется поиск похожего прецедента в базе. В случае отсутствия полного совпадения из базы выбирается максимально близкий прецедент. В зависимости от степени близости прецедента и описания новой проблемы производится адаптация прецедента, т.е. создание нового прецедента на основе существующего. Далее адаптированный прецедент сравнивается с данными базы прецедентов, и на основе этого сравнения, новый прецедент корректируется. Скорректированный прецедент сохраняется в базе.

Для повышения эффективности поиска прецедентов в базе было предложено их кластерное разбиение. После анализа методов кластеризации [7-9] в качестве алгоритма был выбран алгоритм CLOPE. Он лучше всего подходит для кластеризации больших объемов категориальных данных. Кластерные методы являются наиболее подходящими для анализа сложных многомерных объектов, не требуют предварительной структуризации информации, то есть можно классифицировать объекты, не имея данных о предварительном их разбиении (анализ неструктурированной совокупности прецедентов) [10, 11].

Пусть имеется база прецедентов D , состоящая из множества прецедентов $\{p_1, \dots, p_n\}$. Каждый прецедент характеризуется набором уникальных показателей $\{i_1, \dots, i_m\}$, значения которых представлены в таблице 2. Множество кластеров $\{C_1, \dots, C_k\}$ есть разбиение множества $\{p_1, \dots, p_n\}$, такое, что $C_1 \cup \dots \cup C_k = \{p_1, \dots, p_n\}$ и $C_i \cap C_j = \emptyset \quad \forall i \geq 1, k \geq j$. Каждый элемент C_i

называется кластером, а n , m , k — количество прецедентов, количество показателей в базе прецедентов и число кластеров соответственно.

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	...

Каждый кластер C имеет следующие характеристики:

$D(C)$ — множество уникальных показателей;

$Occ(i, C)$ — количество вхождений (частота) показателя i в кластер C ;

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

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

Гистограмма кластера C представляет собой графическое изображение его расчетных характеристик: по оси ОХ откладываются объекты кластера в порядке убывания величины $Occ(i, C)$, а сама величина $Occ(i, C)$ — по оси ОУ.

Гистограмма представлена на рисунке 5.

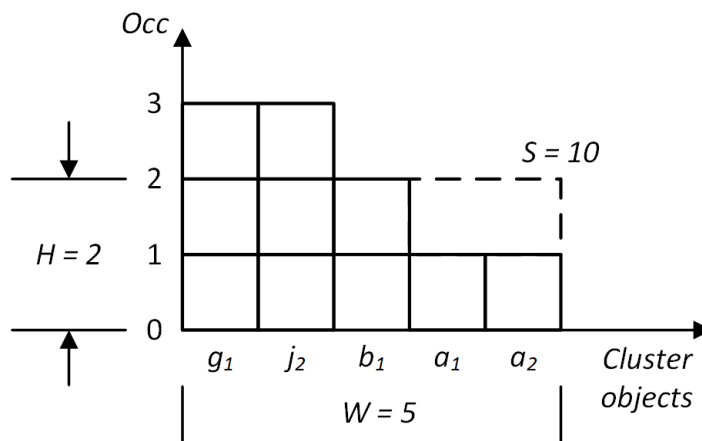


Fig. 5. Cluster histogram

На рисунке 5 $S(C)$, равное 10, соответствует площади прямоугольника, ограниченного осями координат и пунктирной линией. Чем больше значение H , тем большее сходство между двумя изделиями. Поэтому алгоритм выбирает такое разбиение, которое максимизирует H .

Формула вычисления глобального критерия — функции стоимости:

$$Profit(C) = \frac{\sum_{i=1}^k G(C_i) \cdot |C_i|}{\sum_{i=1}^k |C_i|} = \frac{\sum_{i=1}^k \frac{S(C_i)}{W(C_i)^r} \cdot |C_i|}{\sum_{i=1}^k |C_i|}, \quad (3)$$

где $|C_i|$ — количество объектов в i -м кластере, k — количество кластеров, r — коэффициент отталкивания.

Коэффициент r позволяет регулировать уровень сходства прецедентов внутри кластера, и, как следствие, финальное количество кластеров. Этот коэффициент подбирается экспертом. Чем больше r , тем ниже уровень сходства и тем больше кластеров будет сгенерировано.

Формальная постановка задачи кластеризации алгоритмом CLOPE выглядит следующим образом: для заданных D и r найти разбиение $C : Profit(C, r) \rightarrow \max$ [11].

Результаты и обсуждение

Рассмотрим пример из 5 транзакций, представленных в таблице 3.

Table 3. The Example data for clusterization

№	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

Присвоим для каждого уникального значения атрибута код и изменим таблицу с данными:

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

Требуется сравнить между собой два разбиения на кластеры:

Table 5. 2 variants of cluster division

Variant	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)

Для первого и второго вариантов разбиения в каждом кластере требуется рассчитать количество вхождений в него каждого показателя изделия, затем вычислить высоту H и ширину W кластера. Гистограммы кластеров первого разбиения представлены на рисунке 6, а второго на рисунке 7.

Гистограмма кластера $\{A,C,D\}$ содержит 8 различных элементов ($W=8$) и имеет площадь равную 10 ($S=10$), а кластер $\{C,D,E\}$ — 7 различных элементов ($W=7$) с площадью 9 ($S=9$); то есть показатели кластеров отличаются несущественно.

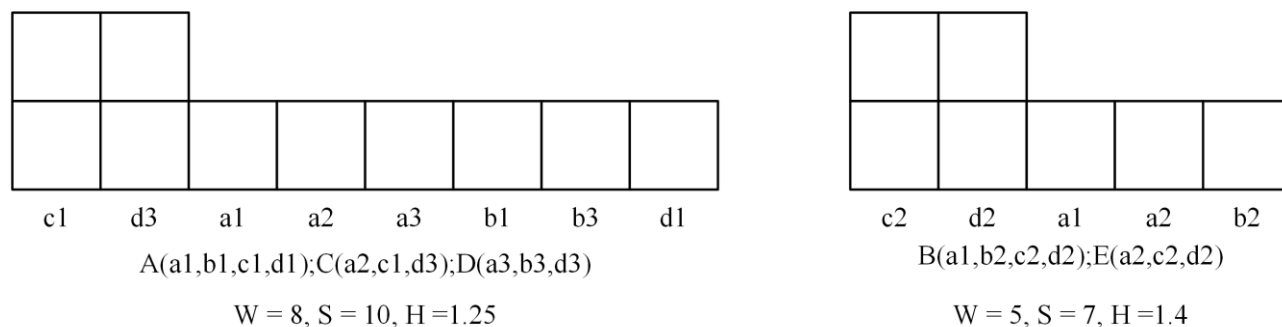


Fig. 6. First variant histogram

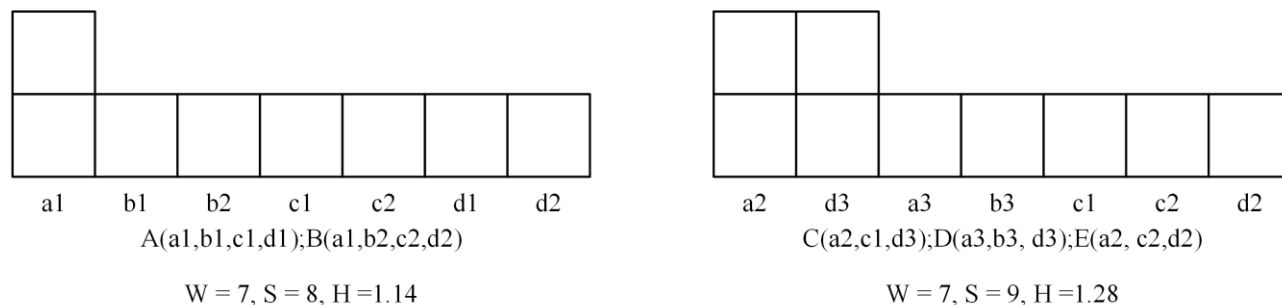


Fig. 7. Second variant histogram

Но при сравнении двух других кластеров видна разница двух разбиений. Площадь кластера {B,E} равна 7 ($S=7$) при 5 различных элементах, а площадь кластера {C,D,E} равна 9 ($S=9$) при 7 различных элементах ($W=7$). Очевидно, что разбиение 1 лучше, поскольку обеспечивает большее наложение транзакций друг на друга (соответственно параметр N там выше). Реализация алгоритма кластеризации представлена в виде блок-схемы на рисунке 8.

В блок-схеме представлены следующие обозначения: n — количество прецедентов в базе, r — коэффициент отталкивания, p_m — m -й прецедент, k — количество кластеров, $C[i]$, $C[j]$ — i -й и j -й кластеры, $Ctemp$ — временный кластер, $Profit(C,r)$ — стоимость кластера по формуле (3), h_m — характеризует принадлежность m -го прецедента к i -му кластеру.

Алгоритм состоит из двух основных фаз: инициализация и итерация. Для работы алгоритма необходимо ввести два значения: n и r . Первый прецедент сразу заносится в первый кластер ($C_1 = p_1$), счетчик кластеров приравнивается единице ($k = 1$), счетчик прецедентов — двум ($m = 2$). Затем следует первый цикл, в котором перебираются все прецеденты базы (пока $m \leq n$). Вводится m -й прецедент, добавляется в i -й кластер, производится расчет значения функции стоимости ($Profit$) для кластера i . В случае, если стоимость кластера увеличивается, то прецедент добавляется в кластер, и алгоритм переходит к следующему прецеденту ($m = m + 1$). Иначе, алгоритм проверяет изменение следующего кластера при добавлении в него прецедента m . Если ни один кластер не увеличивает значение своей стоимости при добавлении прецедента, то создается новый кластер.

Для каждого прецедента m создается запись h_m , в которой указывается принадлежность m -го прецедента к новому кластеру. Фаза инициализации характеризуется распределением прецедентов по кластерам.

Для повышения качества кластеризации следует провести несколько дополнительных сканирования — фазы итераций. Прецеденты сверяются с другими кластерами и, в случае увеличения стоимости кластера новой конструкции прецедент переносится в новый кластер, а из старого удаляется. Пустые кластеры ($C[i] = null$) в завершение алгоритма удаляются.

Выводы

Временная сложность одной итерации равна $N \cdot K \cdot A$, где N — общее число прецедентов, K — максимальное возможное число кластеров, A — средняя длина прецедента. Скорость работы алгоритма растет линейно с увеличением размера массивов исходных данных, что характеризует его как быстрый и эффективный на больших объемах.

В ходе реализации алгоритма CLOPE на языке Java были выявлены следующие преимущества:

- минимальное количество обращений к базе данных;
- использование ограниченного объема оперативной памяти;
- низкая вычислительная сложность;
- масштабируемость.

К несущественным недостаткам можно отнести сложность с подбором коэффициента отталкивания. Но для устранения этой сложности используется вариация алгоритма Fuzzy CLOPE [12].

Методика кластеризации на основе алгоритма CLOPE позволяет повысить эффективность работы с базой прецедентов и, в отличие от других методов кластеризации, позволяет работать с большими объемами категориальных данных. Использование комплексного подхода, основанного на накоплении знаний о прецедентах, позволит повысить эффективность процесса технологической подготовки производства [2, 6].

ACKNOWLEDGEMENTS

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

Литература

1. Valiev R.A., Khairullin A.Kh., Shibakov V.G.. Automated Design Systems for Manufacturing Processes // Russian Engineering Research, 2015. – №35(9). – p. 662–665.
2. Хамадеев Ш.А, Бухаров С.И. База знаний-решений по устранению несоответствий, выявленных в ходе работы ООО «Тиссан» // «Камские чтения. Сборник материалов 2-й межрегиональной научно-практической конференции студентов, аспирантов и молодых ученых». — Наб. Челны: Изд-во Камской госуд. инж.-экон. акад. —2010. — с. 123 –126.
3. Лысенко Д.Э. Автоматизированный синтез моделей технологических процессов на основе прецедентного подхода // Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика. 2014. Т. 29. № 1-1 (172). С. 121–129.
4. Ishikawa, Kaoru (1990); (Translator: J. H. Loftus); Introduction to Quality Control; 448 p.
5. Еремеев, А.П., Варшавский, П.Р. Поиск решения на основе структурной аналогии для интеллектуальных систем поддержки принятия решений // Известия РАН. Теория и системы управления. 2005. №1. С.97–109.
6. A. Aamodt, E. Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications. IOS Press, 1994, Vol. 7: 1.
7. Хамадеев Ш.А., Симонова Л.А., Илюхин А.К. База прецедентов технологических маршрутов штамповочного производства в рамках MES-систем // Кузнечно-штамповочное производство. Обработка материалов давлением. 2009. №8. С. 29–35.
8. Jain A., Murty M., Flynn P. Data Clustering: A Review. // ACM Computing Surveys, 1999. Vol. 31, №. 3.
9. 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.

10. Rui Xu, D. Wunsch. Survey of clustering algorithms. IEEE Transactions on Neural Networks, 2005. Vol. 16, № 3. p. 645–678
11. Паклин, Н. Кластеризация категориальных данных: масштабируемый алгоритм CLOPE. // BaseGroup Labs — Технологии анализа данных. — 2004. URL: <https://basegroup.ru/community/articles/clope>, свободный. Проверено 14.04.2017.
12. 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.
13. Li Jie, Gao Xinbo, Jiao Licheng. A fuzzy CLOPE algorithm and its optimal parameter choice // Journal of Electronics (China), 2006. Vol. 23, № 3. p. 384–388.