

МЕТОДИКА АДАПТИВНОЙ КЛАСТЕРИЗАЦИИ ФАКТОГРАФИЧЕСКИХ ДАННЫХ

Цель работы:

Разработка адаптивной методики кластеризации фактографических данных на основе дивизимных и итерационных методов.

Назначение работы:

Предназначена для бизнес - аналитиков, специалистов по анализу данных, разработчиков систем класса Data Mining.

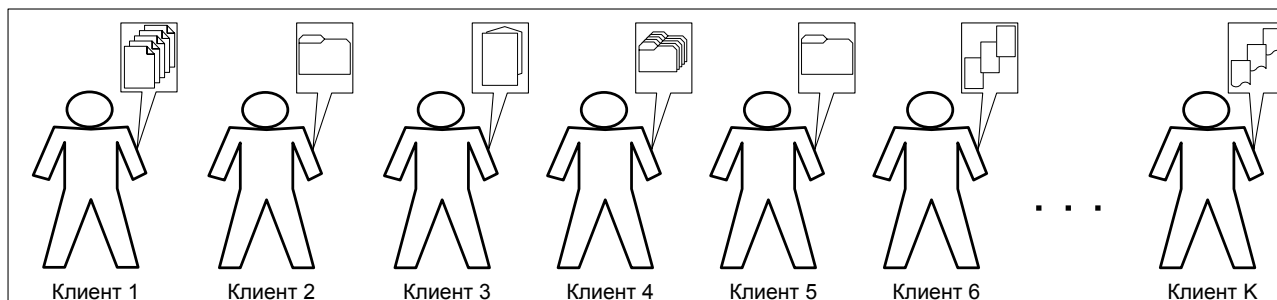
Задачи, подлежащие решению:

1. Исследование методов и подходов интеллектуального анализа данных, используемых для кластеризации фактографических данных.
2. Разработка методики выбора существующих алгоритмов кластеризации.
3. Разработка методики адаптивной кластеризации фактографических данных.
4. Разработка методов докластеризации.
5. Разработка программного комплекса для автоматизации предложенной методики кластеризации.
6. Оценка эффективности предложенной методики с помощью экспериментальных исследований.

ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ

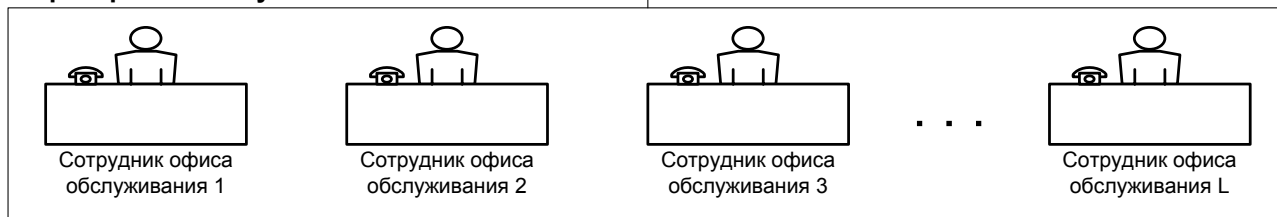
2

Клиенты



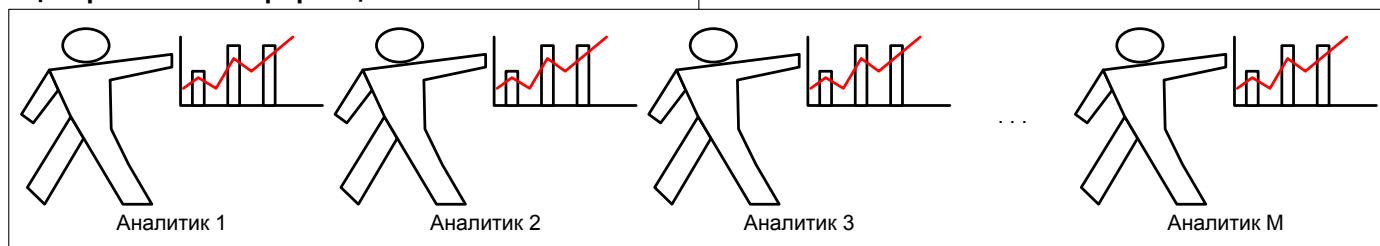
$K \in [500; 50000]$, $Клиент_i = [k_1, k_2, \dots, k_j]$, $j \in [70; 150]$

Брокерское обслуживание



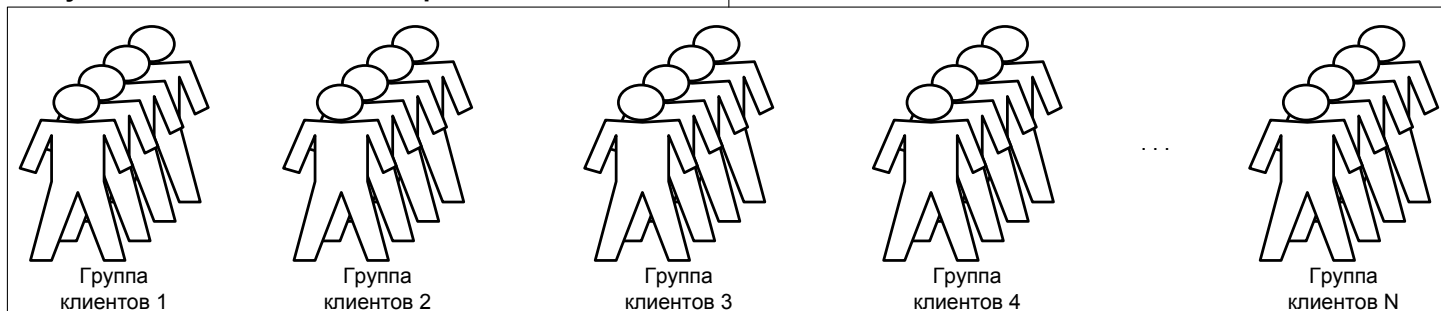
$L \in [20; 100]$

Центр анализа информации



$M \in [5; 20]$

Результаты аналитической обработки



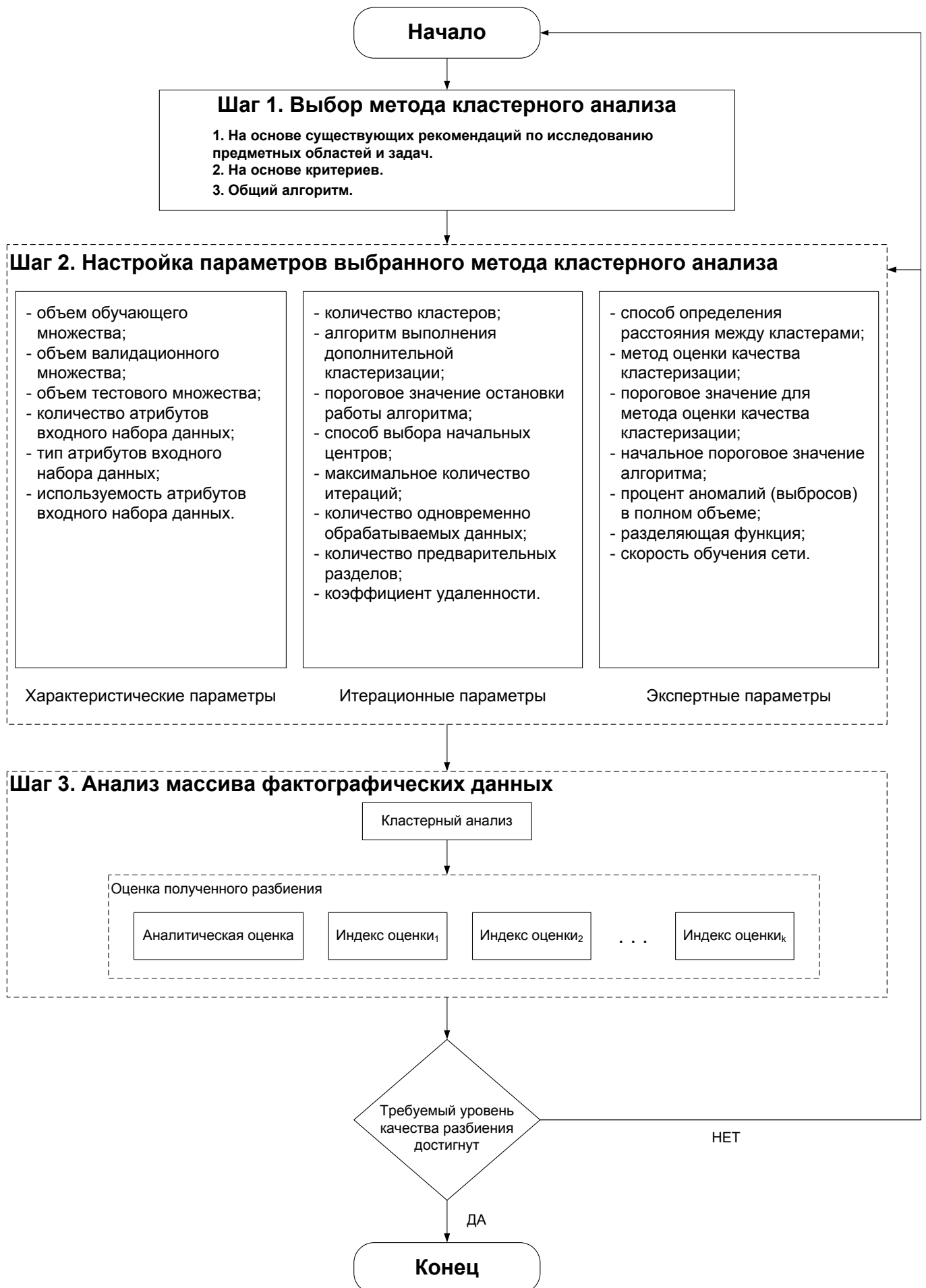
$N \in [5; 30]$

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ВЫБРАННОГО КЛАССА ЗАДАЧ

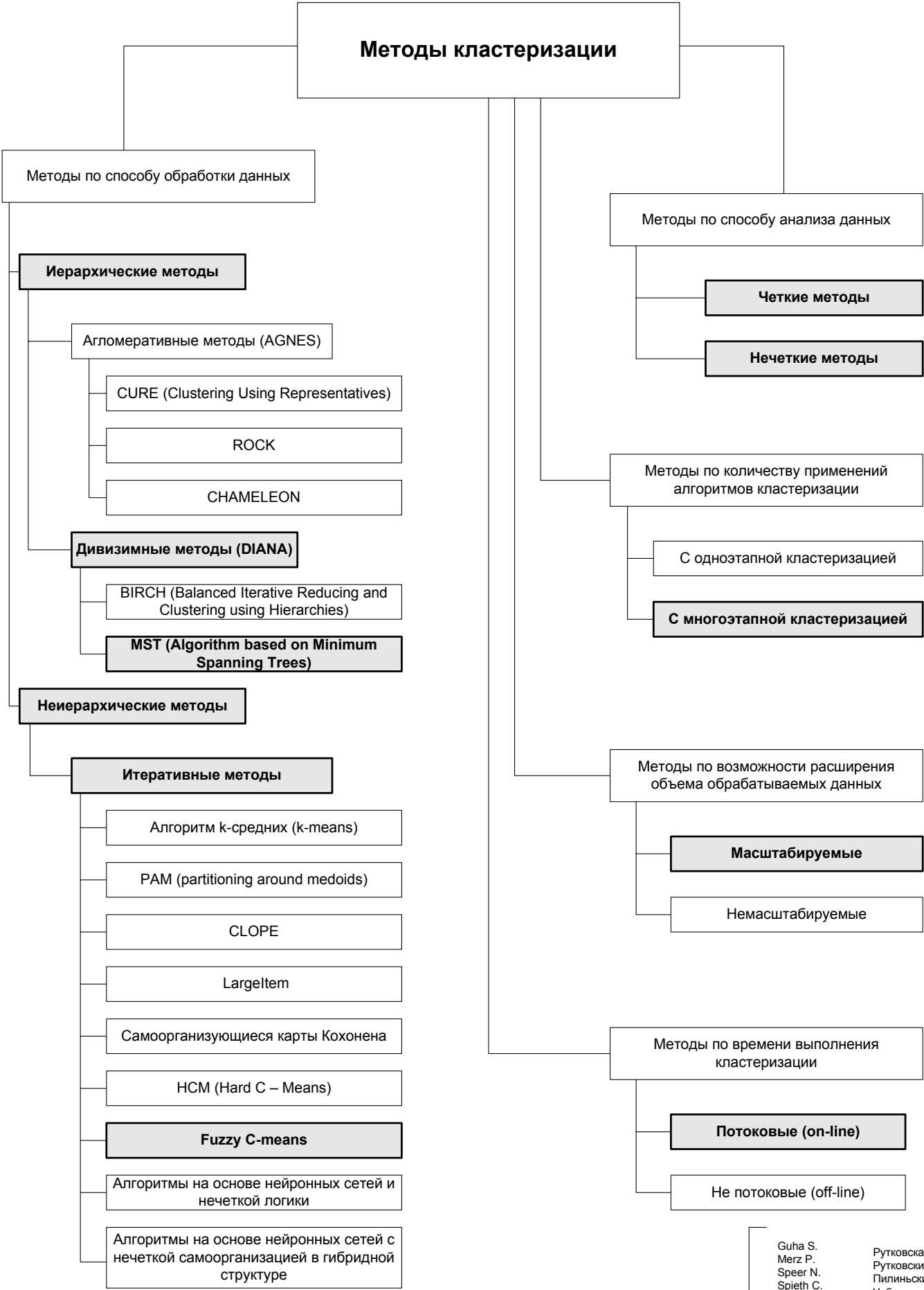
Показатель	Количественная характеристика	Качественная характеристика
Количество исходных объектов	[500 ; 50 000] шт.	-
Количество значимых характеристик объектов	[70 ; 150] шт.	-
Типы характеристик	два вида	числовые, лингвистические
Форма получаемых кластеров	-	сложная, с пересечениями
Количество получаемых кластеров	[5 ; 30] шт.	результат исследования

ФАКТОГРАФИЧЕСКИХ ДАННЫХ





ВЫБОР МЕТОДА КЛАСТЕРИЗАЦИИ НА ОСНОВЕ РЕКОМЕНДАЦИЙ



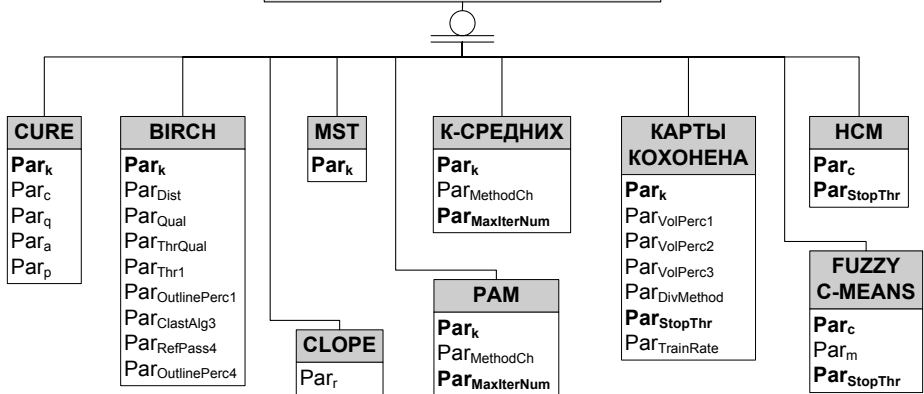
Guha S. Merz P. Speer N. Spieth C. Zadeh L. Zell A.	Рутковская Д. Рутковский Л. Пилиньский М. Чубукова И.А. Штовба С. Д.
--	--

ВЫБОР МЕТОДА КЛАСТЕРИЗАЦИИ
НА ОСНОВЕ КРИТЕРИЕВ

Критерий Метод	Критерий 1 (Cr ₁) Объем информации	Критерий 2 (Cr ₂) Размерность информации	Критерий 3 (Cr ₃) Тип атрибутов	Критерий 4 (Cr ₄) Чувствительность к выбросам	Критерий 5 (Cr ₅) Форма кластеров	Критерий 6 (Cr ₆) Количество кластеров	Критерий 7 (Cr ₇) Перекрываемость кластеров	Критерий 8 (Cr ₈) Качество кластеризации
CURE	линейная зависимость, большой объем данных	невысокая размерность	числовые	нечувствителен	сложная форма	входной параметр	без пересечений	высокое
BIRCH	линейная зависимость, большой объем данных	невысокая размерность	числовые	чувствителен	сферическая форма	входной параметр	без пересечений	среднее
MST	линейно-логарифмическая зависимость, большой объем данных	средняя размерность	числовые	чувствителен	сложная форма, в т.ч. выпуклой и вогнутой форм	входной параметр	без пересечений	высокое
k-средних	нелинейная зависимость, небольшой объем данных	низкая размерность	числовые	очень чувствителен	сферическая форма	входной параметр	без пересечений	низкое
PAM	нелинейная зависимость, небольшой объем данных	низкая размерность	числовые	чувствителен	сложная форма	входной параметр	без пересечений	среднее
CLOPE	линейная зависимость, огромный объем данных	средняя размерность	числовые	чувствителен	сложная форма	вычисляемая величина	без пересечений	высокое
Самоорганизующиеся карты Кохонена	нелинейная зависимость, большой объем данных	высокая размерность	числовые	низкая чувствительность	сложная форма	входной параметр	без пересечений	среднее
HCM	нелинейная зависимость, большой объем данных	средняя размерность	числовые	чувствителен	сложная форма	входной параметр	без пересечений	среднее
Fuzzy C-means	нелинейная зависимость, большой объем данных	высокая размерность	числовые	чувствителен	сложная форма	входной параметр	с пересечениями	среднее

Guha S., Rastogi R., Shim K.; He H., Singh A.; Jain A.K., Dubes R.C.; Kosko B.; Zhang T., Ramakrishnan R., Livny M.; Паклин Н.; Хайкин С.; Штовба С. Д.

МЕТОДЫ КЛАСТЕРНОГО АНАЛИЗА



$$Par_k^{Cure} \equiv Par_k^{Birch} \equiv Par_k^{MST} \equiv Par_k^{k-средних} \equiv Par_k^{PAM} \equiv Par_k^{Kohonen} \equiv Par_c^{HCM} \equiv Par_c^{Fuzzy\ C-means} \equiv Par_c$$

$$Par_{VolPerc1}^{Kohonen} + Par_{VolPerc2}^{Kohonen} + Par_{VolPerc3}^{Kohonen} = 100\%; Par_{MethodCh}^{k-средних} \equiv Par_{MethodCh}^{PAM} \equiv Par_{MethodCh}^{k-средних} \equiv Par_{MaxIterNum}^{PAM} \equiv Par_{MaxIterNum}^{k-средних}$$

$$Par_{StopThr}^{Карты\ Кохонена} \equiv Par_{StopThr}^{HCM} \equiv Par_{StopThr}^{Fuzzy\ C-Means} \equiv Par_{StopThr}$$

ВЫБОР МЕТОДА КЛАСТЕРИЗАЦИИ НА ОСНОВЕ ОБЩИХ ПРЕДСТАВЛЕНИЙ О ХАРАКТЕРЕ РАЗБИЕНИЯ

ЕСЛИ Критерий 6 = «Неизвестно» и Критерий 8 <> «Высокое»	
ТО	Метод кластеризации отсутствует
ИЛИ	
ЕСЛИ Критерий 8 = «Низкое» и Критерий 7 = «Без пересечений» и Критерий 5 = «Сферическая форма» и Критерий 4 = «Данные равномерны» и Критерий 3 = «Числовые атрибуты» и Критерий 2 = «Низкая размерность»	
ТО	МЕТОД := k – средние
ИЛИ	
ЕСЛИ Критерий 8 = «Среднее»	
ТО	ЕСЛИ Критерий 7 = «С пересечениями» и Критерий 5 = «Сложная форма» и Критерий 4 = «Данные равномерны» и Критерий 3 = «Числовые атрибуты» и Критерий 2 <= «Высокая размерность»
	ТО МЕТОД := Fuzzy C-means
	ИЛИ
	ЕСЛИ Критерий 7 = «Без пересечений» и Критерий 5 = «Сферическая форма» и Критерий 4 = «Данные равномерны» и Критерий 3 = «Числовые атрибуты» и Критерий 2 = «Невысокая размерность»
	ТО МЕТОД := BIRCH
	ИЛИ
	ЕСЛИ Критерий 7 = «Без пересечений» и Критерий 5 = «Сложная форма» и Критерий 4 = «Данные неравномерны» и Критерий 3 = «Числовые атрибуты» и Критерий 2 <= «Высокая размерность»
	ТО МЕТОД := Самоорганизующиеся карты Кохонена
	ИЛИ
	ЕСЛИ Критерий 7 = «Без пересечений» и Критерий 5 = «Сложная форма» и Критерий 4 = «Данные равномерны» и Критерий 3 = «Числовые атрибуты» и Критерий 2 <= «Средняя размерность»
	ТО МЕТОД := HCM
	ИЛИ
	ЕСЛИ Критерий 7 = «Без пересечений» и Критерий 5 = «Сложная форма» и Критерий 4 = «Данные равномерны» и Критерий 3 = «Числовые атрибуты» и Критерий 2 = «Низкая размерность» и Критерий 1 = «Небольшой объем данных»
	ТО МЕТОД := PAM
	ИЛИ
	Метод кластеризации отсутствует
ИЛИ	
ЕСЛИ Критерий 8 = «Высокое»	
ТО	ЕСЛИ Критерий 7 = «Без пересечений» и Критерий 5 = «Сложная форма» и Критерий 4 = «Данные неравномерны» и Критерий 3 = «Числовые атрибуты» и Критерий 2 = «Низкая размерность»
	ТО МЕТОД := CURE
	ИЛИ
	ЕСЛИ Критерий 7 = «Без пересечений» и Критерий 5 = «Сложная форма» и Критерий 4 = «Данные равномерны» и Критерий 3 = «Числовые атрибуты» и Критерий 2 <= «Средняя размерность» и Критерий 1 <= «Большой объем данных»
	ТО МЕТОД := MST
	ИЛИ
	ЕСЛИ Критерий 7 = «Без пересечений» и Критерий 5 = «Сложная форма» и Критерий 4 = «Данные равномерны» и Критерий 3 = «Числовые атрибуты» и Критерий 2 <= «Средняя размерность» и Критерий 1 <= «Огромный объем данных»
	ТО МЕТОД := CLOPE
	ИЛИ
	Метод кластеризации отсутствует
ИЛИ	
Метод кластеризации отсутствует	

МЕТОДИКА АДАПТИВНОЙ КЛАСТЕРИЗАЦИИ ФАКТОГРАФИЧЕСКИХ ДАННЫХ (Укрупненная схема)

8



Входные данные алгоритма:

$D = \{u_1, u_2, \dots, u_m\}$, где u_i – объекты кластеризации, m – количество объектов кластеризации, $i = \overline{1, m}$;
 $u_i = \{(Value_{i1}, t_1), (Value_{i2}, t_2), \dots, (Value_{in}, t_n)\}$, где $Value_{ij}$ – значение j^{zo} атрибута i^{zo} объекта кластеризации, t_j – тип атрибута объекта кластеризации, n – количество атрибутов объекта кластеризации, $j = \overline{1, n}$;
 $t_j = \{ValueType_j, FieldType_j\}$, где $ValueType_j$ – тип значения атрибута, $ValueType_j \in ValueTypes$, $FieldType_j$ – вид значения атрибута, $FieldType_j \in FieldTypes$; $FieldTypes = \{Входное, Идентифицирующее, Информационное\}$;
 $ValueTypes = \{\text{Целочисленный тип}, \text{Денежный тип}, \text{Лингвистический тип}\}$, где $\text{Целочисленный тип} \subset \mathbb{Z}$,
 $\text{Денежный тип} \subset \mathbb{R}$, $\text{Лингвистический тип} \subset \text{Словарная система}$;
 $\text{Словарная система} = \{\text{Лингв.тип1}, \text{Лингв.тип2}, \dots, \text{Лингв.типS}\}$, где Лингв.тип_i – объект словарной системы, характеризующий оценочные/качественные показатели объектов кластеризации;
 q – максимальное количество кластеров, $q \leq m$; p – размазанность кластеров, $p \in (0; 10]$;
 $K = \{K_1, K_2, \dots, K_n\}$, где K_i – весовой коэффициент влияния атрибута объекта, $K_i \in [0; 1]$;
 w – степень удаленности элементов, $w \in (0; 1]$;
 $Metric$ – способ определения расстояния между объектами, $Metric \in Metrics$;
 $Metrics = \left\{ \begin{array}{l} \text{Евклидово расстояние, Квадрат Евклидова} \\ \text{расстояния, расстояние Чебышева} \end{array} \right\}$;
 $OstTreeMethod$ – способ построения минимального остовного дерева, $OstTreeMethod \in OstTreeMethods$;
 $OstTreeMethods = \left\{ \begin{array}{l} \text{Алгоритм Борушки, Алгоритм} \\ \text{Крускала, Алгоритм Прима} \end{array} \right\}$;
 $NormMethod$ – способ проведения нормализации значений числовых атрибутов, $NormMethod \in NormMethods$;
 $NormMethods = \{\text{Линейная нормализация, Статистическая нормализация}\}$

Выходные данные алгоритма: $C = \{C_1, C_2, \dots, C_c \mid O^c \rightarrow \max, c \leq q, C_1 \cup C_2 \cup \dots \cup C_c = D\}$, $\mu_{ij} \in [0; 1]$, $u_i \in C_j$, $i = \overline{1, m}$, $j = \overline{1, c}$

Описание алгоритма:

Этап 1. Нормализация значений числовых атрибутов.

В случае линейной нормализации выполняется следующее:

$$Value_{ij} := \left\{ \frac{Value_{ij}}{\max_j(Value_{ij})} \mid \max_j(Value_{ij}) \neq 0, \right. \\ \left. t_j \in \{\text{Целочисленный тип}, \text{Денежный тип}\} \right\}$$

В случае статистической нормализации выполняется

$$\text{следующее: } Value_{ij} := \left\{ \frac{Value_{ij} - \frac{\sum_{i=1}^m Value_{ij}}{m}}{\sqrt{\frac{\sum_{i=1}^m (Value_{ij})^2}{m} - \left(\frac{\sum_{i=1}^m Value_{ij}}{m}\right)^2}} \mid \max_j(Value_{ij}) \neq 0, \right. \\ \left. t_j \in \{\text{Целочисленный тип}, \text{Денежный тип}\} \right\}$$

Этап 2. Вычисление матрицы взаимных расстояний между объектами.

$Dist_{ij} = \|u_i - u_j\| = Metric(u_i, u_j)$, где $Metric$ – способ определения расстояния между объектами.

Если $Metric = \text{Евклидово расстояние}$, то $Dist_{ij} = \sqrt{\sum_w ([Value_{iw} - Value_{jw}] * K_w)^2}$,

Если $Metric = \text{Квадрат Евклидова расстояния}$, то $Dist_{ij} = \sum_w ([Value_{iw} - Value_{jw}] * K_w)^2$,

Если $Metric = \text{Расстояние Чебышева}$, то $Dist_{ij} = \max_w [Value_{iw} - Value_{jw}] * K_w$,

где $i, j \in [1, m]$, $w = \overline{1, n}$ при условии $FieldType[w] = \text{"Входное"}$

Этап 3. Построение минимального остовного дерева по выбранному способу построения дерева с использованием матрицы взаимных расстояний между объектами $Dist$.

Этап 4. Разделение объектов на кластеры и построение матрицы нечеткого разбиения F .

Матрица нечеткого разбиения: $F = [\mu_{ij}]$, $\mu_{ij} \in [0, 1]$, $i \leq q$, $j = \overline{1, m}$, где μ_{ij} – степень принадлежности i^{zo} объекта к j^{ny} кластеру. Матрица разбиения обладает следующими свойствами: $\sum_{i=1}^k \mu_{ij} = 1$, $j = \overline{1, m}$, $0 < \sum_{j=1}^m \mu_{ij} \leq m$, $i = \overline{1, k}$.

Шаг 1. Определение количества кластеров итерации: $k := q$.

Шаг 2. $Dist_{ij}^k := \left\{ 0 \mid Dist_{ij}^k = Max \right\}$.

Шаг 3. Расчет центров выделенных кластеров V_i^k .

$V_i^k = Avg \left(\left\{ u_j \mid u_j \in C_i^k \right\} \right)$, где Avg – оператор вычисления среднего значения показателей объектов, входящих в кластер k , $i = \overline{1, k}$, $j = \overline{1, m}$.

Для числовых типов оператор Avg определяется выражением:

$$Avg[r] = \frac{\sum_{u_j \in V_i^k} \{ Value_{jr} \mid FieldType[w] = "Входное" \}}{|V_i^k|}, \quad j = \overline{1, m}, \quad r = \overline{1, n}.$$

Для лингвистических типов оператор Avg определяется выражением:

$$Avg[r] = \left\{ \frac{Value_{jr}}{\sum_{\substack{u_j \in V_i^k, u_l \in V_l^k \\ \|Value_{jr} - Value_{lr}\| = \min \text{ или } \varphi = \max}} \mid FieldType[w] = "Входное" \right\}, \quad r = \overline{1, n}, \quad j = \overline{1, m}, \quad l = \overline{1, m}, \text{ где } \varphi - \text{частота значения}$$

атрибута в пределах кластера V_i^k .

Шаг 4. Расчет матрицы расстояний от объектов до центров кластеров V_i^k : $Dist_{ij}^k = \|V_i^k - u_j\| = Metric(V_i^k, u_j)$, $i = \overline{1, k}$, $j = \overline{1, m}$, где $Metric$ – способ определения расстояния между объектами.

Шаг 5. Нормализация матрицы расстояний от объектов до центров кластеров V_i^k :

$$Dist_{ij}^{k'} = \begin{cases} \frac{Dist_{ij}^k}{Max(Dist_{ij}^k)}, & Max(Dist_{ij}^k) \neq 0 \\ 1, & Max(Dist_{ij}^k) = 0 \end{cases}, \quad i = \overline{1, k}, \quad j = \overline{1, m}.$$

Шаг 6. Соотнесение объектов к кластерам в соответствии со степенью удаленности элементов кластера (w):

$$u_j \in V_i^{k'} \mid Dist_{ij}^{k'} \leq w \text{ или } Dist_{ij}^{k'} = Min(Dist_{ij}^{k'}), \quad i = \overline{1, k}, \quad j = \overline{1, m}.$$

Шаг 7. Расчет степени принадлежности кластеру: $\mu_{ij} = (1 - Dist_{ij}^{k'})^2$, $i = \overline{1, k}$, $j = \overline{1, m}$.

Шаг 8. Нормализация матрицы нечеткого разбиения: $\mu_{ij} = \frac{\mu_{ij}}{\sum_{i=1}^k \mu_{ij}}$, $j = \overline{1, m}$.

Шаг 9. Вычисление центров полученных кластеров с использованием матрицы нечеткого разбиения:

$$V_i^{k'} = \frac{\sum_{j=1}^m \mu_{ij}^p * u_j}{\sum_{\substack{j=1 \\ u_j \in V_i^{k'}}}^m \mu_{ij}^p}, \quad i = \overline{1, k}.$$

Для лингвистических атрибутов центра кластера вычисление производится с

использованием выражения: $V_i^{k'}[r] = Value_{jr} \mid \mu_{ij} = Max(\mu_{ij})$.

$$\sum_{i=1, k} \frac{|V_i^{k'}| * \sum_{j=1}^m \mu_{ij}^p * \|V_i^{k'} - u_j\|}{Min(\|V_i^{k'} - u_j\|) * Max(\|V_i^{k'} - u_j\|) * \sum_{\substack{j=1 \\ u_j \in V_i^{k'}}}^m \|V_i^{k'} - u_j\| * k}$$

Шаг 10. Оценка качества разбиения: $O^k = \frac{\sum_{i=1, k} |V_i^{k'}| * \sum_{j=1}^m \mu_{ij}^p * \|V_i^{k'} - u_j\|}{m * k^2}$, где

$|V_i^{k'}|$ – количество элементов в кластере i ;

$\|V_i^{k'} - u_j\| = Metric(V_i^{k'}, u_j)$ – расстояние от центра кластера i до элемента u_j ;

$u_j \in V_i^{k'}$ – отражение условия о принадлежности элемента кластеру.

Шаг 11. $k := k - 1$.

Шаг 12. Если $k > 0$, то переход на шаг 2.

Этап 5. Выбор наилучшего разбиения: $O_{Onm} = MAX_{i=1, q}(O^i)$.

СРАВНЕНИЕ ЭФФЕКТИВНОСТИ МЕТОДИК

- Исследование 1: выделение секторов инвестирования на основе анализа показателей финансовых инструментов;
- Исследование 2: выделение групп клиентов на основе данных об оборотах за период, частоты проведения операций, количестве финансовых инструментов, группы используемых финансовых инструментов;
- Исследование 3: выявление категорий финансовых инструментов для оценки эффективности операций;
- Исследование 4: выделение классов автомобилей на основе данных о максимальной скорости, цвете кузова, воздушном сопротивлении, массе.

Метод 1 – карта Кохонена, Метод 2 – алгоритм к – средних, Метод 3 – ADAKL

Общая оценка разбиений

Показатель Метод		Средневзвешенная оценка разбиения	Средневзвешенная оценка разбиения с заданным количеством кластеров (без учета лингвистических атрибутов)	Средневзвешенная оценка разбиения с заданным количеством кластеров (с учетом лингвистических атрибутов)	Средняя оценка разбиения	Средняя оценка разбиения с заданным количеством кластеров (без учета лингвистических атрибутов)	Средняя оценка разбиения с заданным количеством кластеров (с учетом лингвистических атрибутов)
Исследование 1	Метод 1	0.9133	-	0.9627	0.9142	-	0.9615
	Метод 2	-	-	-	-	-	-
	Метод 3	0.9825	-	0.9997	0.9822	-	0.9999
Исследование 2	Метод 1	0.8720	0.9414	0.9616	0.8807	0.9539	0.9660
	Метод 2	-	0.7630	-	-	0.7584	-
	Метод 3	0.9825	0.9969	0.9975	0.9802	0.9969	0.9981
Исследование 3	Метод 1	0.7874	0.8857	0.9030	0.8206	0.9077	0.9214
	Метод 2	-	0.7309	-	-	0.7259	-
	Метод 3	0.9770	0.9984	0.9989	0.9807	0.9984	0.9992
Исследование 4	Метод 1	0.5925	0.9178	0.8675	0.6821	0.9495	0.8985
	Метод 2	-	0.9757	-	-	0.9764	-
	Метод 3	0.9628	0.9991	0.9999	0.9603	0.9987	0.9995

Средняя оценка разбиений

Показатель Оценка		Средневзвешенная оценка разбиения	Средневзвешенная оценка разбиения с заданным количеством кластеров (без учета лингвистических атрибутов)	Средневзвешенная оценка разбиения с заданным количеством кластеров (с учетом лингвистических атрибутов)	Итоговая оценка
Метод 1		0.7913	0.9150	0.9237	0.8767
Метод 2		-	0.8232	-	0.8232
Метод 3		0.9762	0.9981	0.9990	0.9911

Оценка разбиения выполняется на основе показателей выполненной кластеризации: $O = \frac{r}{n} \times \begin{cases} \frac{q}{k}, & \text{если } q \leq k \\ \frac{k}{q}, & \text{если } q > k \end{cases}$,
где

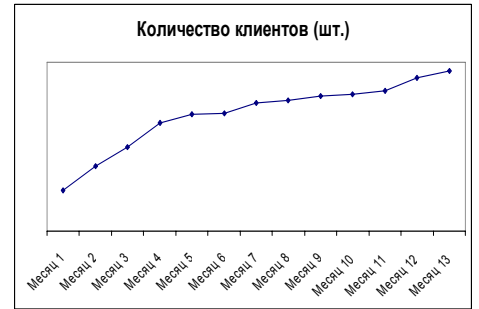
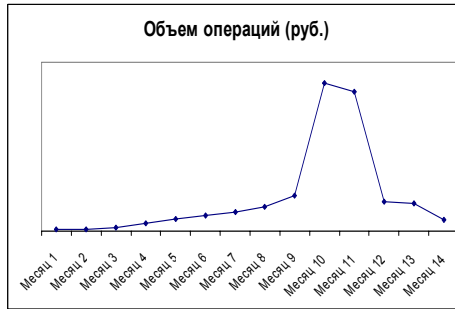
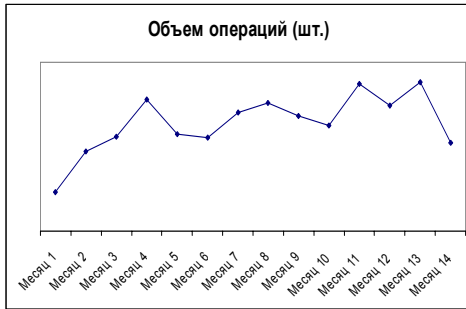
q – количество кластеров по итогам кластеризации;

r – количество элементов, правильно распределенных по соответствующим кластерам;

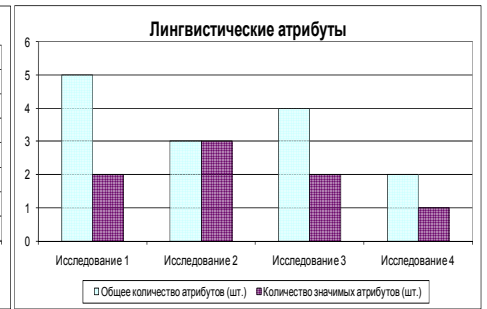
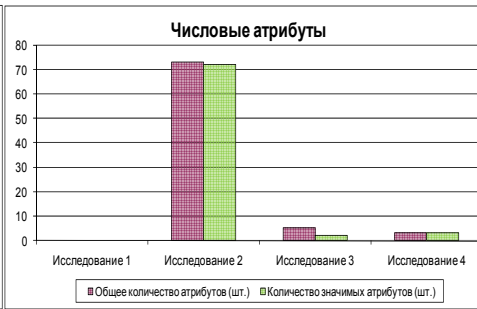
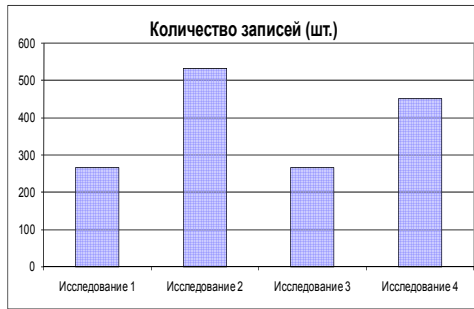
k – исходное количество кластеров;

n – количество объектов кластеризации.

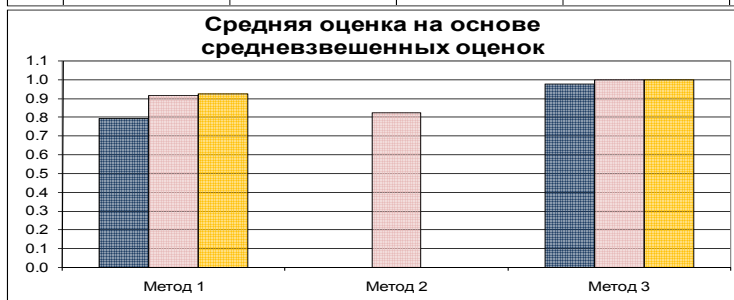
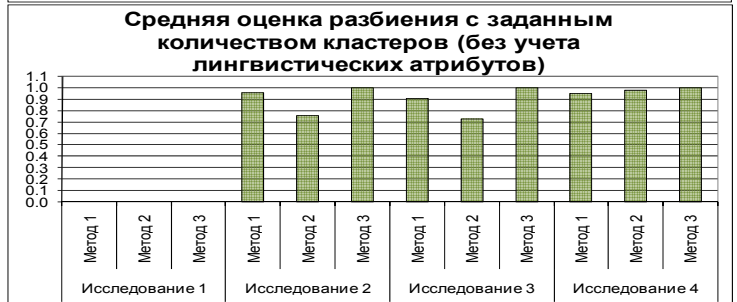
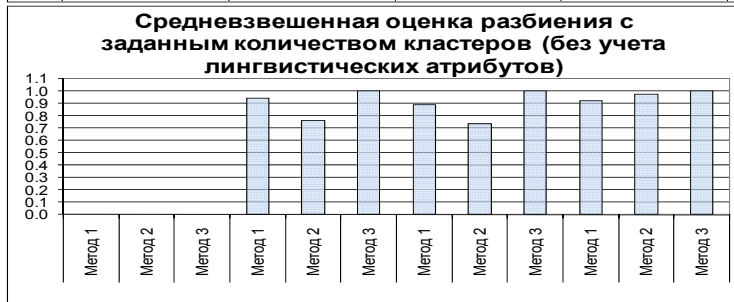
Показатели операций брокерского обслуживания



Характеристики исследуемых данных

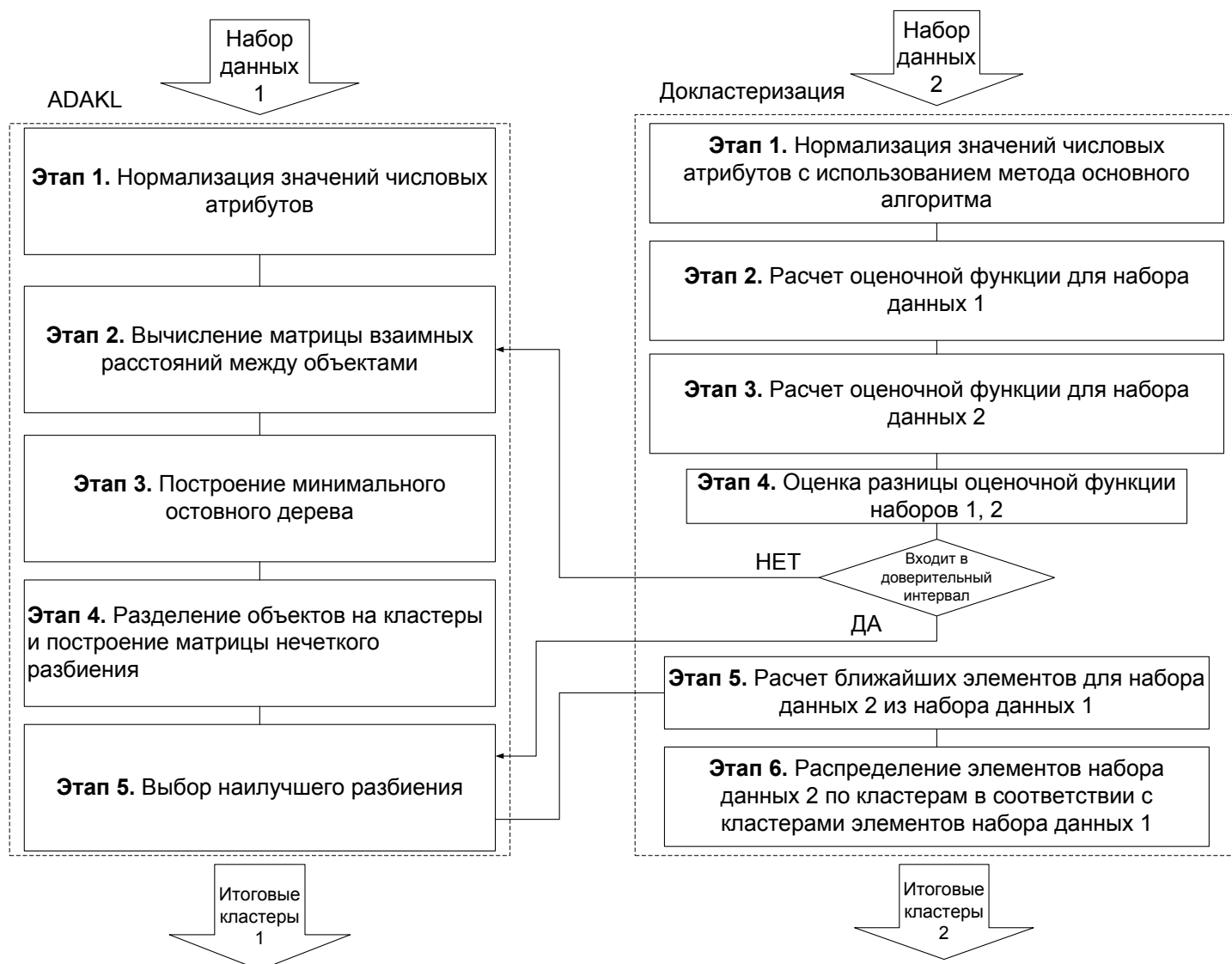


ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДИК



ДОКЛАСТЕРИЗАЦИЯ ФАКТОГРАФИЧЕСКИХ ДАННЫХ

Укрупненная схема



Оценка исходных наборов данных

$$D_1 = \sqrt{\frac{\sum_{i=1}^r \|A_i - \text{Avg}[A]\|^2}{r}} \quad D_2 = \sqrt{\frac{\sum_{i=1}^o \|B_i - \text{Avg}[B]\|^2}{o}} \quad , \text{ где } D_1, D_2 - \text{оценочная функция исходных наборов данных 1 и 2 соответственно}$$

Оценка разницы оценочных функций: $|D_1 - D_2| \leq \delta$, где δ – доверительный интервал

Распределение элементов расширяющего множества по кластерам

$$\mu_{ij} = \frac{1}{1 - \text{Dist}_k^{\text{Norm}}} * \mu_{kj} = \frac{1}{1 - \text{Dist}_k^j / \text{Max}(\text{Dist}_k)} * \mu_{kj} \quad , \text{ где}$$

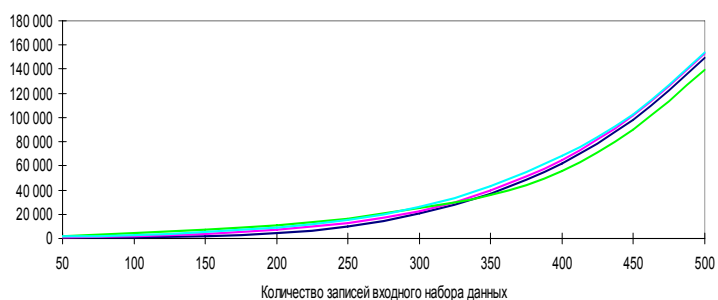
$\text{Dist}_k^j = \|B_i - A_j\|$ – расстояние между ближайшими элементами из множеств A и B $i = \overline{1, o}$ – порядковый номер элемента из множества B

$\text{Max}(\text{Dist}_k)$ – максимальное расстояние от элемента из множества B до элемента из множества A $j \in [1, r]$ – порядковый номер ближайшего элемента из множества A для соответствующего элемента из множества B

μ_{kj} – степень принадлежности элемента из множества A к кластеру k

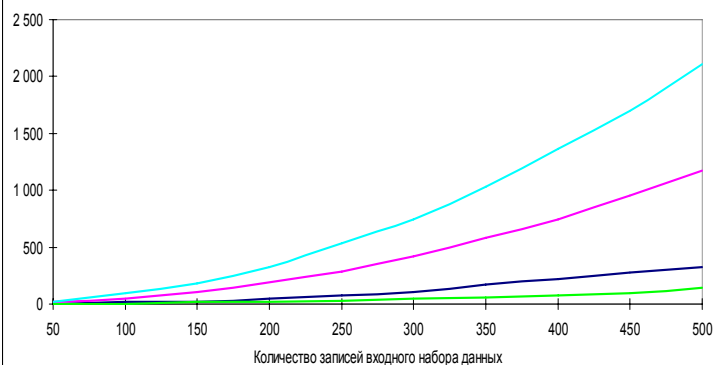
k – порядковый номер ближайшего элемента

Общее время выполнения анализа (миллисек.)



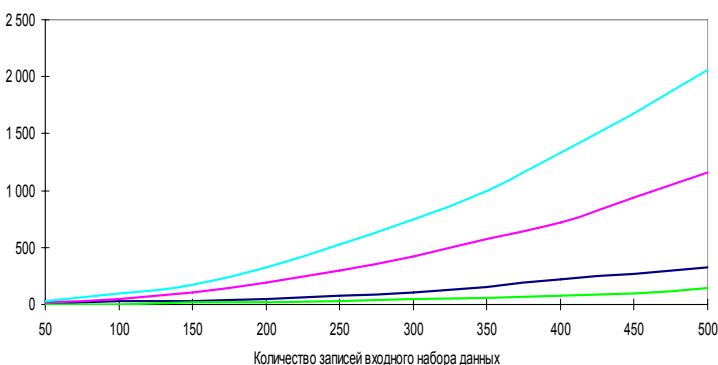
Серия экспериментов 1 Серия экспериментов 2 Серия экспериментов 3 Серия экспериментов 4

Время выполнения 2-го этапа анализа (миллисек.)



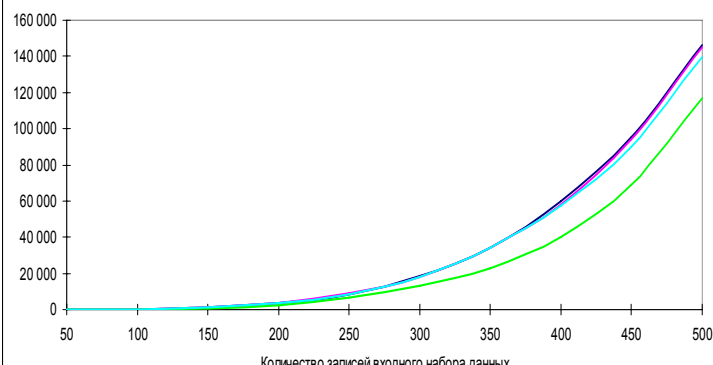
Серия экспериментов 1 Серия экспериментов 2 Серия экспериментов 3 Серия экспериментов 4

Общее время выполнения доп. анализа (миллисек.)



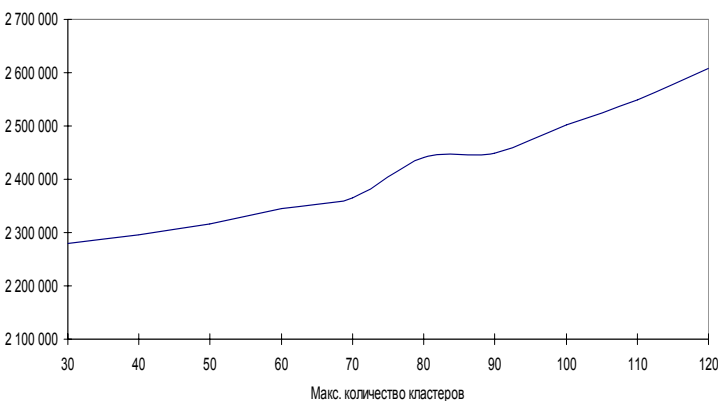
Серия экспериментов 1 Серия экспериментов 2 Серия экспериментов 3 Серия экспериментов 4

Время выполнения 3-го этапа анализа (миллисек.)



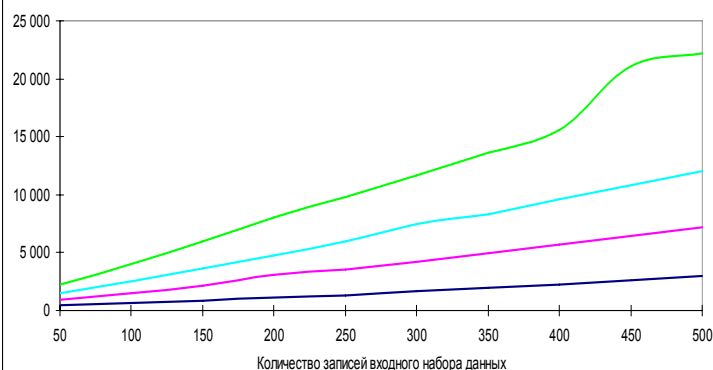
Серия экспериментов 1 Серия экспериментов 2 Серия экспериментов 3 Серия экспериментов 4

Общее время выполнения анализа (миллисек.)



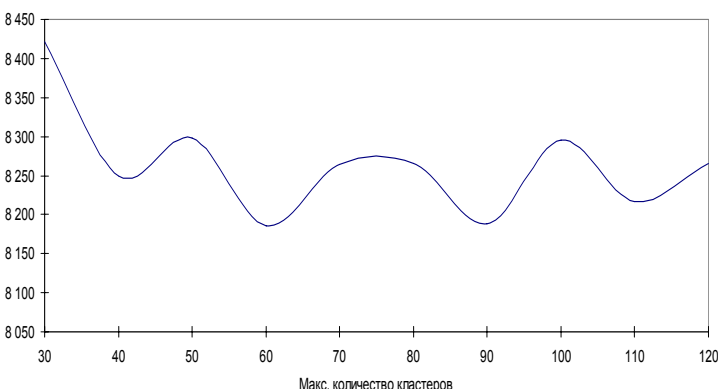
Серия экспериментов 1 Серия экспериментов 2 Серия экспериментов 3 Серия экспериментов 4

Время выполнения 4-го этапа анализа (миллисек.)



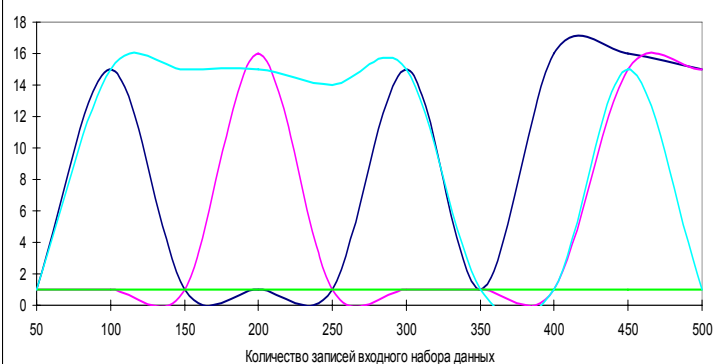
Серия экспериментов 1 Серия экспериментов 2 Серия экспериментов 3 Серия экспериментов 4

Общее время выполнения доп. анализа (миллисек.)



Серия экспериментов 1 Серия экспериментов 2 Серия экспериментов 3 Серия экспериментов 4

Время выполнения 5-го этапа анализа (миллисек.)



Серия экспериментов 1 Серия экспериментов 2 Серия экспериментов 3 Серия экспериментов 4

УКРУПНЕННАЯ СХЕМА АРХИТЕКТУРЫ ПРОГРАММНОГО КОМПЛЕКСА

16



ЭКРАННЫЕ ФОРМЫ ПРОГРАММНОГО КОМПЛЕКСА

fmDataQuery

select rtrim(Brief) SecBrief, case when SecType = 0 then 'Акции' else 'Облигации' end SecType, Branch SecBr

SecBrief

46014

46018

48001

DCL ASSETS Int. Ltd.

IMPEX07

RusFed05

RusFed18

RusFed30

SISFIN08

ГМК Норникель АО в 5

Газэнергопромбанк АОИ

Отменить Очистить текст Выполнить запрос Импортировать данные

fmConfigDataSet

PortfolioID

DaysCount

BuyCount

SellCount

BckRepoCount

FrwRepoCount

BuyCountOnTS

SellCountOnTS

BuyCountOnVNB

SellCountOnVNB

BuyCountWShr

SellCountWShr

BckRepoCountWShr

FrwRepoCountWShr

BuyCountWShrOnTS

Вид значения атрибута: Входной

Тип значения атрибута: Дробный

Наименование атрибута: PortfolioID

Весовой коэффициент атрибута: 1

Макс. количество кластеров: 533

Разноразность кластеров: 1.0000

Степень удаленности элементов: 1.0000

Способ определения расстояния: Евклидово расстояние

Способ построения мин. остоного дерева: Алгоритм Борувки

Очистить виды/типы значений Закрыть

fmResults											
Исходные данные	Обработанные данные		Матрица расстояний		Матрица связности		Матрица нечеткого разбиения		Оценка разбиений		
N	1	2	3	4	5	6	7	8	9	10	11
1	0.000000	2.351880	1.452179	1.378286	1.415287	1.706679	1.378950	1.405524	1.390798	1.379021	1.604517
2	2.351880	0.000000	1.397319	1.710077	1.560639	1.157141	1.721131	1.599861	1.640260	1.995235	1.093817
3	1.452179	1.397319	0.000000	0.340590	0.225283	0.702109	0.388000	0.261687	0.272932	0.674564	0.326773
4	1.378286	1.710077	0.340590	0.000000	0.153178	0.838709	0.119818	0.117207	0.034306	0.000110	0.629761
5	1.415287	1.560639	0.225283	0.153178	0.000000	0.717207	0.197978	0.147978	0.043005	0.000110	0.486703
6	1.706679	1.157141	0.702109	0.838709	0.717207	0.000000	0.784366	0.755816	0.794204	1.175497	0.668681
7	1.378950	1.721131	0.388000	0.119818	0.197978	0.784366	0.000000	0.158302	0.153434	0.521858	0.687618
8	1.405524	1.599861	0.261687	0.117207	0.043005	0.755816	0.158302	0.000000	0.064715	0.607118	0.529572
9	1.390798	1.640260	0.272932	0.074377	0.090110	0.794204	0.153434	0.064715	0.000000	0.555840	0.559935
10	1.379021	1.995235	0.674564	0.525782	0.629761	1.175497	0.521858	0.607118	0.555840	0.000000	0.990377
11	1.604517	1.093817	0.326773	0.631276	0.486503	0.668681	0.687820	0.529572	0.559635	0.990377	0.000000
12	1.425326	1.449742	0.211072	0.301091	0.179705	0.642615	0.322509	0.206101	0.241446	0.773516	0.398414
13	1.445397	2.095987	0.869575	0.786958	0.673640	1.359553	0.784318	0.857124	0.810433	0.293083	1.164202
14	1.893646	0.709980	0.778562	1.097652	0.954913	0.873180	1.401766	0.996306	1.026050	1.411270	0.473523
15	1.372492	1.745955	0.393929	0.092395	0.200738	0.826546	0.409443	0.159399	0.139973	0.509561	0.680618
16	1.440223	1.469073	0.215610	0.276733	0.151324	0.578251	0.262430	0.171995	0.222147	0.680319	0.424040
17	1.671778	0.956938	0.477264	0.762191	0.614054	0.433434	0.765025	0.650903	0.634781	1.083674	0.236414
18	1.694120	1.629745	0.320320	0.622061	0.492034	0.637232	0.675910	0.234140	0.763325	0.901301	0.604517
19	1.599120	1.923711	0.371607	0.633649	0.407133	0.400335	0.630937	0.522257	0.509547	0.970060	0.220101
20	1.479542	1.379953	0.289952	0.394912	0.273057	0.455976	0.359781	0.292327	0.345273	0.765577	0.489935
21	1.384452	2.049034	0.778220	0.673760	0.767307	1.277169	0.670193	0.740453	0.700356	0.210676	1.005517
22	2.203615	0.794917	1.330380	1.551403	1.417559	0.744893	1.507670	1.443958	1.497066	1.825021	1.116212

fmResults

Исходные данные	Обработанные данные	Матрица расстояний	Матрица связности	Матрица нечеткого разбиения	Оценка разбиений
N	Кластер 1	Кластер 2	Кластер 3	Кластер 4	Кластер 5
1	0.000587	0.002354	0.000141	0.000181	0.000056
2	0.007476	0.002153	0.000209	0.000184	0.000055
3	0.000050	0.001319	0.000000	0.000000	0.000000
4	0.002345	0.010523	0.000196	0.000035	0.000174
5	0.002096	0.005899	0.000462	0.000430	0.000135
6	0.001515	0.001619	0.000829	0.000829	0.000894
7	0.000193	0.000163	0.016120	0.007972	0.007444
8	0.000197	0.000420	0.007896	0.016230	0.006373
9	0.000000	0.000000	0.000196	0.005189	0.016919
10	0.000587	0.002354	0.000141	0.000181	0.000056
11	0.007476	0.002153	0.000209	0.000184	0.000055
12	0.000050	0.001319	0.000000	0.000000	0.000000
13	0.002345	0.010523	0.000196	0.000035	0.000174
14	0.002096	0.005899	0.000462	0.000430	0.000135
15	0.001515	0.001619	0.000829	0.000829	0.000894
16	0.000193	0.000163	0.016120	0.007972	0.007444
17	0.000197	0.000420	0.007896	0.016230	0.006373
18	0.000000	0.000000	0.000196	0.005189	0.016919
19	0.000587	0.002354	0.000141	0.000181	0.000056
20	0.007476	0.002153	0.000209	0.000184	0.000055
21	0.000050	0.001319	0.000000	0.000000	0.000000
22	0.002345	0.010523	0.000196	0.000035	0.000174