

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования

«Южно-Уральский государственный университет
(национальный исследовательский университет)»
Высшая школа электроники и компьютерных наук
Кафедра «Информационно-измерительная техника»

ДОПУСТИТЬ К ЗАЩИТЕ
Заведующий кафедрой,
д.т.н., профессор
_____ А.Л. Шестаков
_____ 2020 г.

РАЗРАБОТКА ПРИЛОЖЕНИЯ КЛАСТЕРНОГО АНАЛИЗА РЕЗУЛЬТАТОВ ЭКСПЕРТНОГО ОЦЕНИВАНИЯ

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
ЮУрГУ–12.03.01.2020.272. ВКР

Руководитель проекта,
ст. преподаватель кафедры ИнИТ
_____ / С.В. Пашнин
_____ / 2020 г.

Автор проекта
студент группы КЭ-414
_____ / Д.В. Александров
_____ / 2020 г.

Нормоконтролер,
доцент кафедры ИнИТ
_____ / А.А. Лысова
_____ / 2020 г.

Челябинск 2020

АННОТАЦИЯ

Александров, Д. В. Разработка приложения кластерного анализа результатов экспертного оценивания – Челябинск: ЮУрГУ, ВШЭКН, 2020, 83 с. библиогр. список – 25 наим., 5 прил., 31 ил., 14 табл.

В рамках выпускной квалификационной работы разработано приложение кластерного анализа результатов экспертного оценивания.

В данной работе разработан алгоритм работы приложения кластерного анализа. Данный алгоритм реализован в программном пакете Microsoft Visual Studio на языке программирования C# при помощи встроенной среды Console Application.

Также разработан диалоговый интерфейс при помощи интерфейса программирования приложений (API) Windows Forms, входящий в состав среды разработки MS Visual Studio.

Проведено тестирование приложения на примере массива точек в двумерном пространстве, а также на примере объектов в многомерном пространстве. Также проведен кластерный анализ на основе результатов анкетирования среди студентов-выпускников бакалавриата посредством разработанного приложения. Результаты кластеризации были подвергнуты анализу.

					ЮУрГУ – 12.03.01.2020.272 ВКР			
<i>Изм</i>	<i>Лист</i>	<i>№ докум.</i>	<i>Подп.</i>	<i>Дата</i>	Разработка приложения кластерного анализа результатов экспертного оценивания	<i>Лит</i>	<i>Лист</i>	<i>Листов</i>
<i>Разраб.</i>		Александров Д.В.					5	83
<i>Проверил</i>		Пашнин С.В.				ЮУрГУ Кафедра ИНИТ		
<i>Т. контр.</i>								
<i>Н. контр.</i>		Лысова А.А.						
<i>Утверд.</i>		Шестаков А.Л.						

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	7
1 ОБЗОРНЫЙ РАЗДЕЛ	9
1.1 Анализ технического задания.....	9
1.2 Определение кластерного анализа и обзор его методов.....	10
1.3 Обзор мер расстояния между кластерами	13
1.4 Обзор методов связи кластеров.....	15
2 ПРОЕКТНЫЙ РАЗДЕЛ	20
2.1 Кластеризация полным перебором	20
2.2 Алгоритм иерархической последовательной кластеризации.....	21
2.3 Разработка алгоритма работы приложения кластерного анализа.....	25
2.4 Визуализация результатов кластерного анализа	28
3 РАЗДЕЛ РАЗРАБОТКИ ПРИЛОЖЕНИЯ	32
3.1 Разработка листинга программы	32
3.2 Разработка графического интерфейса приложения	40
4 РАЗДЕЛ ТЕСТИРОВАНИЯ	45
4.1 Тестирование приложения на примере множества точек в двумерном пространстве	45
4.2 Тестирование приложения в многомерном пространстве.....	48
4.3 Проведение кластерного анализа результатов анкетирования студентов посредством приложения.....	50
4.4 Анализ результатов кластерного анализа.....	52
ЗАКЛЮЧЕНИЕ	62
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	64
ПРИЛОЖЕНИЯ	
ПРИЛОЖЕНИЕ А Алгоритм работы приложения кластерного анализа	67
ПРИЛОЖЕНИЕ Б Листинг программы приложения кластерного анализа....	68
ПРИЛОЖЕНИЕ В Исходные данные для тестирования	78
ПРИЛОЖЕНИЕ Г Перечень вопросов анкетирования студентов.....	79
ПРИЛОЖЕНИЕ Д Выборка для проведения кластерного анализа	83

ВВЕДЕНИЕ

С развитием цифровых технологий возникает все большее количество данных, способность порождения которых значительно превышает способность их хранения и обработки.

Кластерным анализом (или кластеризацией) называют процесс объединения близких (схожих) объектов, получившие название кластеры. Процедура кластеризации позволяет получать определенные классификации объектов по совокупности признаков.

Кластеризация используется во многих сферах, например как статистика, сегментация изображений, анализ больших данных (Big Data), распознавание объектов, маркетинг, прогнозирование.

Методов кластеризации на сегодняшний день возникло множество – несколько десятков алгоритмов, а также их модификации.

Вследствие того, что измерение нефизических величин, производящееся, по большей части, посредством субъективного оценивания, является одним из приоритетных направлений кафедры «Информационно-измерительная техника» Южно-Уральского государственного университета, было предложено разработать приложение, осуществляющее процедуру кластерного анализа.

Целью данной работы является разработка приложения кластерного анализа результатов экспертного оценивания для образования классифицированной экспертной системы по совокупности критериев.

Для достижения этой цели необходимо выполнить следующие задачи:

- выполнить обзор методов кластерного анализа, мер сходства и методов объединения (связи) кластеров;
- разработать алгоритм процедуры кластерного анализа, реализуемый в приложении;
- разработать листинг программы и диалоговый интерфейс приложения;
- провести тестирование работы приложения и процедуру кластеризации на основе результатов анкетирования с последующим их анализом.

Изм	Лист	№ докум.	Подпись	Дата

В ходе выполнения выпускной квалификационной работы используются приложения Google Forms, MS Visual Studio и КОМПАС-3D.

					<i>ЮУрГУ – 12.03.01.2020.272 ВКР</i>	<i>Лист</i>
<i>Изм</i>	<i>Лист</i>	<i>№ докум.</i>	<i>Подпись</i>	<i>Дата</i>		8

1 ОБЗОРНЫЙ РАЗДЕЛ

1.1 Анализ технического задания

Согласно техническому заданию на выпускную квалификационную работу необходимо разработать приложение кластерного анализа экспертного оценивания. Приложение предназначено для проведения кластерного анализа и с его помощью полученных результатов выделить из множества объектов группы похожих по совокупности критериев.

Вследствие того, что количество кластеров не задано, используются иерархические методы кластерного анализа, которые будут рассмотрены далее. Ограниченное число исходных объектов в задании (до 100) также предполагает применение иерархических методов кластерного анализа.

В условиях того, что изначально неизвестна размерность выборки, то есть количество объектов и признаков, следует использовать динамические массивы, которые не имеют фиксированного размера.

Разработка диалогового интерфейса осуществляется в среде разработки Windows Form, входящей в состав пакета MS Visual Studio на языке программирования C#.

Диалоговый интерфейс должен быть разработан в соответствии с требованиями [1], а именно

- функциональность;
- информативность;
- соответствие ожиданиям;
- простота обучения пользователя;
- устойчивость к сбоям;
- удовлетворение предпочтениям пользователя [1].

В рамках тестирования приложения кластерного анализа будут использованы простые примеры, которые должны показать работоспособность приложения, а также качественное разбиение объектов на кластеры. Такими примерами будут массив точек в двумерном пространстве и массив объектов в многомерном

пространстве, некоторые группы кластеров из которых специальным образом будут разделены от других кластеров ограниченным диапазоном возможных значений их признаков. Тем самым определим способность кластерного анализа в первую очередь распознавать наиболее схожие объекты.

Также необходимо провести процедуру кластерного анализа с реальными данными. Наиболее простым способом образования выборки из реальных данных является анкетирование. Результаты анкетирования будут использованы в качестве исходных данных для процедуры кластерного анализа. А его результаты используются для анализа по дополнительным данным анкетирования.

1.2 Определение кластерного анализа и обзор его методов

Термин «кластерный анализ» предложен К. Трионом в 1939 г. в работе [2]. Синонимами к данному термину выступают различные выражения, например, автоматическая классификация, таксономия, распознавание без обучения, самообучение [3].

Кластерный анализ является многомерным классификационным анализом, главным назначением которого является разбиение множества объектов и их признаков на схожие по совокупности этих признаков группы. Пример группировки объектов на кластеры представлен на рисунке 1.

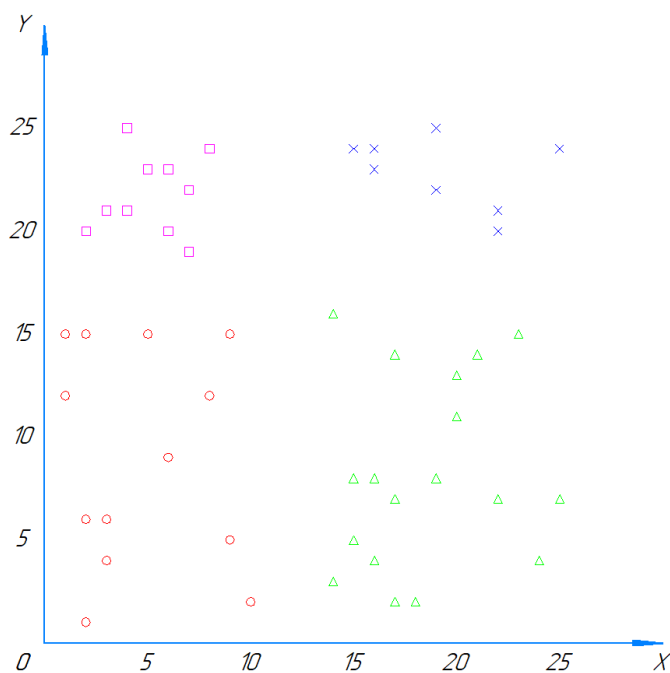


Рисунок 1 – Разбитые на кластеры точки в двумерном пространстве

Изм	Лист	№ докум.	Подпись	Дата

Существуют две основные группы методов кластерного анализа: иерархические и неиерархические.

Иерархические методы кластерного анализа заключаются в последовательном объединении кластеров небольшого размера в крупные кластеры и, наоборот.

Важным достоинством данной группы методов является возможность визуализации, удобной для исследователя. Однако, основным недостатком таких методов является сложность использования при наличии большого количества объектов.

Существует множество вариантов иерархического кластерного анализа из-за различий критериев построения иерархии, которые используются при объединении объектов, близких по этим критериям, в кластеры или, наоборот, разделения на группы.

В зависимости от того, объединяются кластеры или происходит их разделение, различают иерархические агломеративные и дивизимные методы кластерного анализа.

Агломеративные методы характеризуются последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров.

В начале процесса таких методов все объекты являются обособленными. На первом шаге наиболее «схожие» объекты объединяются в кластер. На последующих шагах объединение продолжается до того момента, когда все объекты образуют один кластер, как это показано на рисунке 2.

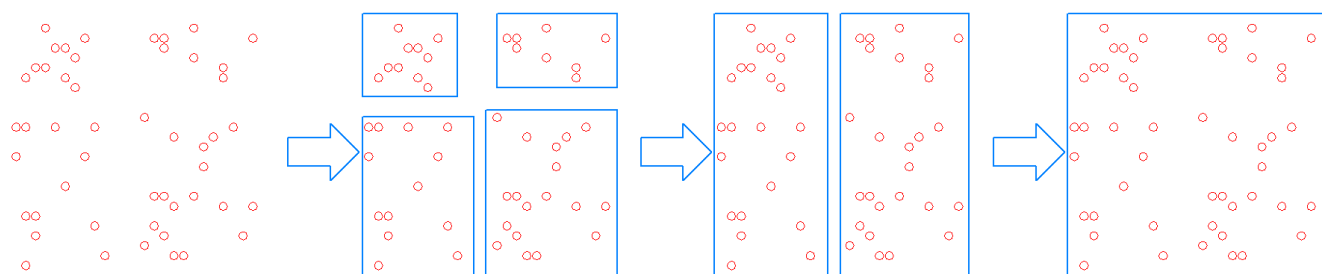


Рисунок 2 – Процесс агломеративного метода

Дивизимные методы, в отличие от агломеративных, осуществляют обратную процедуру.

В начале процесса этих методов объекты составляют один большой кластер, который на последующих шагах делится более мелкие кластеры. В результате получаются обособленные исходные объекты.

Процесс дивизимного кластерного анализа представлен на рисунке 3.

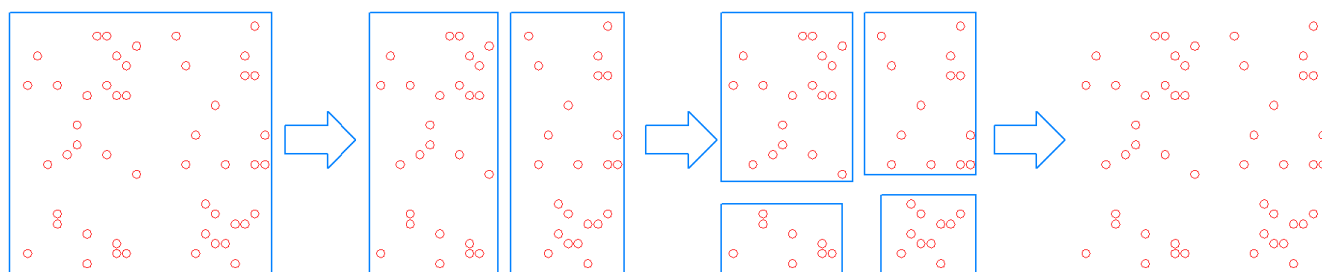


Рисунок 3 – Процесс дивизимного метода

Неиерархические методы кластерного анализа отличаются от иерархических главным образом тем, что при их использовании необходимо знать количество кластеров, на которые будут разбиты объекты, а также могут использоваться при большом количестве объектов. Вследствие отсутствия знаний и том, на какое число кластеров необходимо разбить выборку, данные методы использоваться не будут. Но рассмотрим один из самых распространенных для понимания работы таких методов.

Самым распространенным неиерархическим методом является метод k -средних (k -means), описанный в работе [4].

Метод k -средних – это один из неиерархических методов кластерного анализа, целью которого является разбиение m объектов на множестве пространства R^n на k кластеров, при этом каждый объект относится к тому кластеру, к центру которого оно ближе всего.

В качестве меры близости используется евклидово расстояние, которое будет рассмотрено в подразделе 1.3.

Алгоритм представляет собой итерационную процедуру, в которой выполняются следующие шаги [5].

1. Заранее задается число кластеров k . Из исходного множества данных случайным образом выбираются k объектов, которые будут служить начальными центрами кластеров.

2. Для каждого объекта исходного множества определяется ближайший к ней центр кластера.

3. Вычисляются центры тяжести кластеров, называемые центроидами. Центр кластера смещается в его центроид, который затем становится центром нового кластера.

4. 3-й и 4-й шаги итеративно повторяются. В итоге минимизируется расстояние между элементами внутри кластеров, и увеличиваются межкластерные расстояния.

Алгоритм останавливается тогда, когда границы кластеров и расположения центроидов перестанут изменяться на каждом последующем шаге итерации

Достоинством данного алгоритма являются скорость и простота реализации.

Недостатками является неопределенность выбора начальных центров кластеров, а также задание изначально числа кластеров, что может потребовать некоторой априорной информации об исходных данных.

В выпускной квалификационной работе будут использоваться иерархические агломеративные методы.

1.3 Обзор мер расстояния между кластерами

Меры расстояния являются самыми распространенными мерами сходства. Меры расстояния отталкиваются от понятия метрики – правила определения того или иного расстояния между двумя точками (элементами) данного пространства (множества) [6].

Свойства, которым должны удовлетворять метрики:

— симметрия (расстояние от точки x до точки y должно быть таким же, как от y до x), при этом расстояние может быть либо равно нулю, либо положительным;

— неравенство треугольника (расстояние от x до y всегда меньше расстоянию от x до какой-то другой точки z плюс расстояние от y до той же точки);

Изм	Лист	№ докум.	Подпись	Дата

— различимость тождественных объектов (если объекты не одинаковы, то расстояние между ними не равно нулю);

— неразличимость идентичных объектов (если объекты одинаковые, то расстояние должно быть равно нулю).

Различают следующие виды мер расстояния.

Самым распространенным является евклидово расстояние, вычисляемое по формуле

$$d_{ij} = \sqrt{\sum_{k=1}^n (i_k - j_k)^2}.$$

Также используют квадрат стандартного евклидова расстояния, вычисляемого по формуле

$$d_{ij}^2 = \sum_{k=1}^n (i_k - j_k)^2.$$

Одним из распространенных мер расстояния является расстояние городских кварталов. Данное расстояние рассчитывается как среднее разностей по координатам двух кластеров.

В основном такая мера расстояния приводит к результатам, аналогичным евклидову расстоянию. Стоит отметить, что для этой меры влияние отдельных выбросов меньше, чем при использовании евклидова расстояния, поскольку здесь координаты не возводятся в квадрат [7].

Расстояние городских кварталов рассчитывается по формуле

$$d(i, j) = \sum_{k=1}^n |i_k - j_k|.$$

Расстояние Чебышева используется для определения «различных» объектов, отличающихся по какому-то одному измерению [7].

Вычисляется расстояние Чебышева по формуле

$$l_{\infty}(i, j) = \max_{k=1, \dots, n} |i_k - j_k|.$$

Процент несогласия вычисляется, если данные являются категориальными, то есть, данные, которые качественно характеризуют процесс или объект и не имеют количественного выражения. Вычисляется процент несогласия по формуле

$$d(i, j) = \frac{\text{Count}(i_k \neq j_k)}{n},$$

где $k = 1, \dots, n$.

В рамках выпускной квалификационной работы мерой расстояния является квадрат евклидова расстояния.

1.4 Обзор методов связи кластеров

Существуют различные правила определения расстояния между кластерами, которые называются методами связи (объединения) кластеров. Рассмотрим следующие методы, используемые в кластерном анализе, которые основаны на евклидовой метрике:

- метод одиночной связи (ближнего соседа) [8];
- метод полной связи (дальнего соседа) [9], [10];
- метод двухгруппового среднего [11], [12];
- центроидный метод [11], [12];
- метод Уорда [13].

Рассмотрим каждый из них подробнее.

Метод ближнего соседа или одиночная связь заключается в следующем. Расстояние кластеров определяется расстоянием между наиболее близкими друг к другу объектами этих кластеров [7]. Расстояние вычисляется по формуле

$$D_{IJ} = \min d(X_i, Y_j).$$

На примере двумерного пространства это выглядит следующим образом (Рисунок 4).

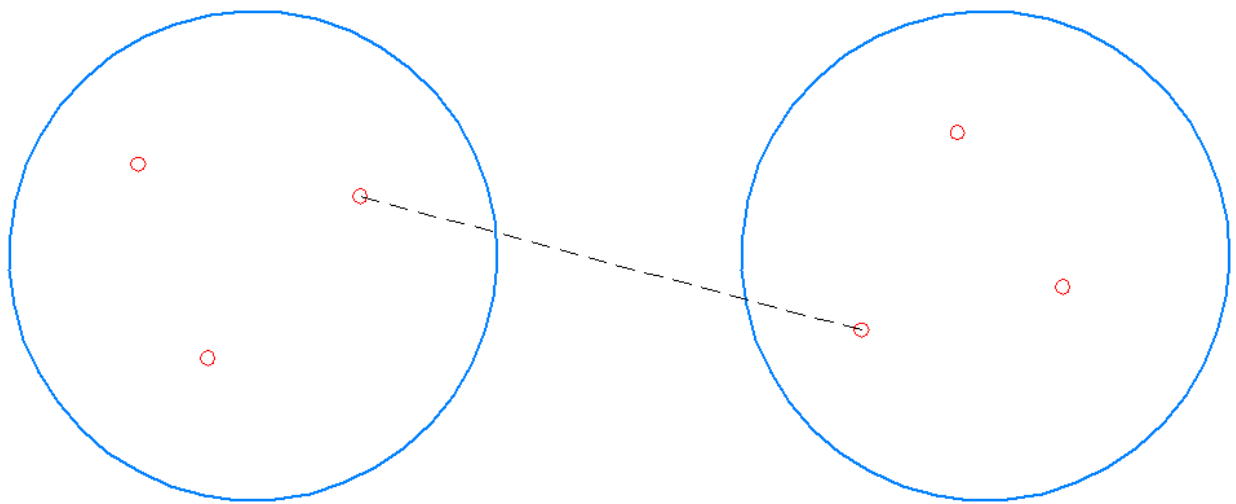


Рисунок 4 – Метод ближнего соседа (одиночной связи)

Этот метод позволяет выделять кластеры сколь угодно сложной формы при условии, что различные части таких кластеров соединены цепочками близких друг к другу элементов [7].

Достоинством такого метода является его простота.

Недостатками являются необходимость постоянного хранения матрицы сходства и невозможность определения по результатам кластеризации, какое число кластеров можно образовать.

Суть метода дальнего соседа или полной связи заключается в том, что два объекта из одной и той же группы имеют коэффициент сходства, который меньше некоторого порогового значения.

Другими словами, евклидово расстояние d не должно превышать данный установленный порог. В этом случае r определяет максимально допустимый диаметр подмножества, образующего кластер [14]. Расстояние вычисляется по формуле

$$D_{IJ} = \max d(X_i, Y_j).$$

Графически данный метод представлен на рисунке 5.

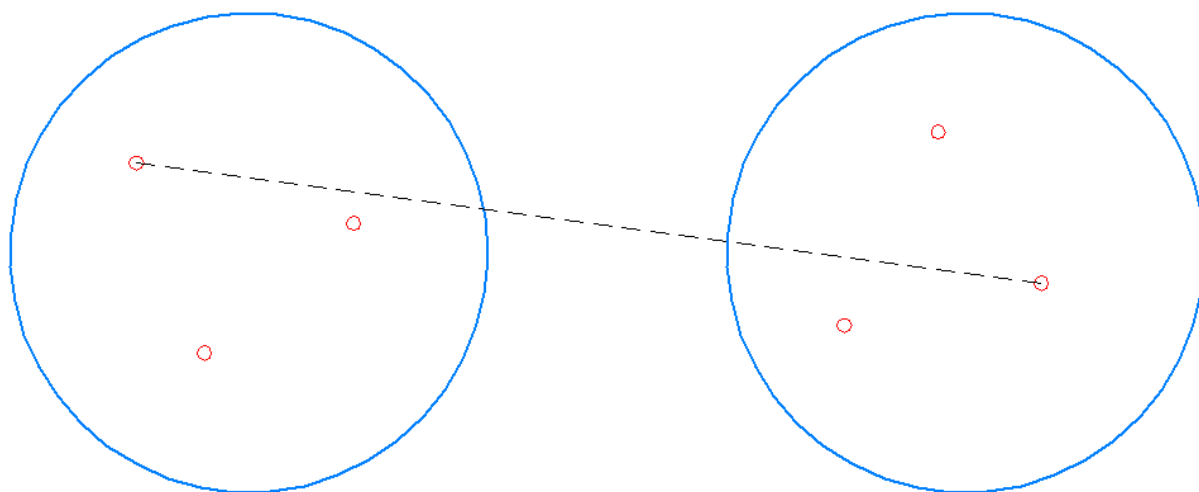


Рисунок 5 – Метод дальнего соседа

Метод двухгруппового среднего опирается на связь между объектом i и кластером I . Эта связь выражается в виде среднего коэффициента сходства между объектом i и всеми объектами, входящими в кластер I [14]. Применяется данный метод при предположении о неравном количестве признаков кластеров. Расстояние вычисляется по формуле

$$D_{il} = \frac{1}{n_l} \sum_{j=1}^{n_l} (X_j - Y)^T \cdot (X_j - Y),$$

Графически данный метод представлен на рисунке 6.

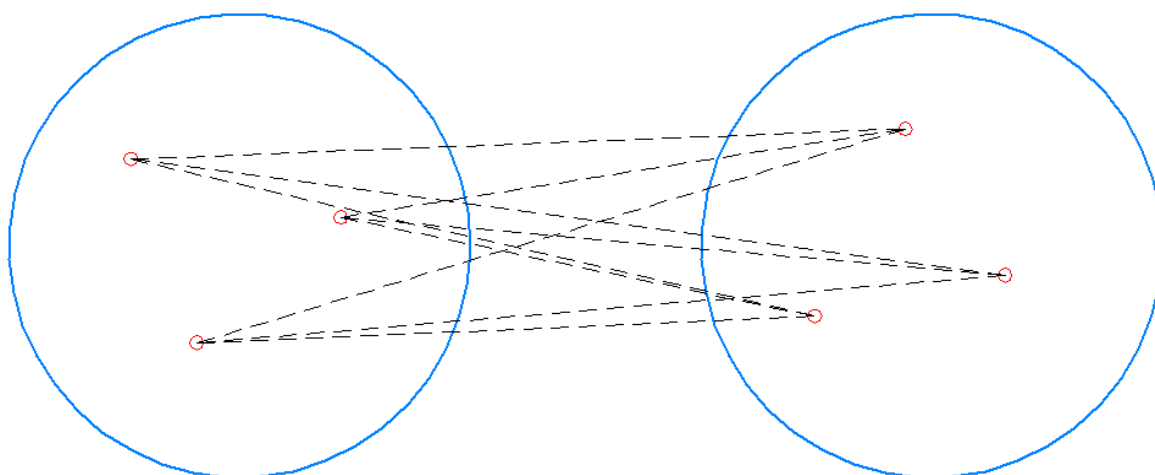


Рисунок 6 – Метод двухгруппового среднего

Центроидный метод заключается в следующем. Расстояние между двумя кластерами I и J в этом методе определяется как евклидово расстояние между центрами этих кластеров [14]. Расстояние вычисляется по формуле

$$d_{IJ}^2 = (\bar{X} - \bar{Y})^T \cdot (\bar{X} - \bar{Y}),$$

где X, Y - векторы средних по кластерам I и J соответственно.

Кластеризация осуществляется поэтапно: на каждом из $n-1$. шагов объединяют два кластера I и J , имеющие минимальное значение d_{IJ}^2 . Если n_1 много больше n_2 , то центры $I \cup J$ близки друг к другу и характеристики J при объединении кластеров исключаются. [14].

Графически данный метод представлен на рисунке 7.

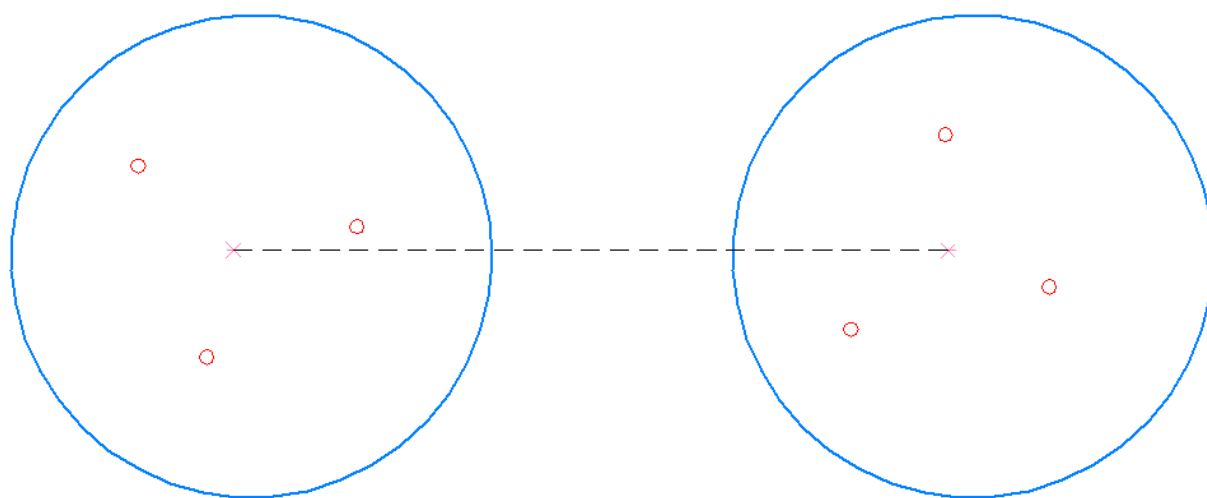


Рисунок 7 – Центроидный метод

Метод Уорда заключается в следующем. В качестве целевой функции применяется внутригрупповая сумма квадратов (ВСК) отклонений, обозначающая сумму квадратов расстояний между каждым объектом и средней по кластеру, содержащему этот объект.

На каждом шаге объединяются такие два кластера, которые приводят к минимальному увеличению целевой функции, которой является ВСК [14]. При объединении кластеров I и J это увеличение равно

$$D_{IJ} = \frac{n_1 \cdot n_2}{n_1 + n_2} (\bar{X} - \bar{Y})^T \cdot (\bar{X} - \bar{Y}), \quad (1.1)$$

где n_1, n_2 – число объектов, входящих в кластеры I и J соответственно.

В качестве метода связи кластеров выбран метод Уорда.

Вывод по разделу 1: в ходе обзорно-аналитической части ВКР были выявлены аналоги приложений, в функционал которых входит проведение кластерного анализа.

Также был проведен обзор методов кластерного анализа, мер сходства между кластерами и методов объединения схожих кластеров. В результате был выбран иерархический агломеративный метод кластерного анализа.

Мерой сходства между кластерами выбран квадрат евклидова расстояния, а методом объединения схожих кластеров выбран метод Уорда.

2 ПРОЕКТНЫЙ РАЗДЕЛ

2.1 Кластеризация полным перебором

Наиболее прямым способом проведения кластерного анализа является метод полного перебора. Суть его заключается в переборе всех допустимых разбиений на группы и поиске такого разбиения, которое ведет к оптимальному значению целевой функции. Такая процедура выполнима при малом количестве объектов n и кластеров m [14].

Число разбиений $S(n, m)$ называется числом Стирлинга второго рода и вычисляется по формуле

$$S(n, m) = \frac{1}{m!} \sum_{k=0}^m C_m^k \cdot (-1)^k (m-k)^n, \quad (2.1)$$

где m – количество кластеров;

n – количество объектов;

$k = 0, 1, \dots, m$;

C_m^k – число сочетаний, вычисляемое по формуле

$$C_m^k = \frac{m!}{k!(m-k)!}.$$

Стоит отметить, что число разбиений по формуле (2.1) можно определить, если известно число кластеров m . В ином случае, когда число кластеров неизвестно, то число всевозможных разбиений определяется по формуле

$$\sum_{m=1}^n S(n, m) = \sum_{m=1}^n \left[\frac{1}{m!} \sum_{k=0}^m C_m^k \cdot (-1)^k (m-k)^n \right].$$

При наличии целевой функции, например ВСК, которая эквивалентна евклидовому расстоянию, оптимальное решение кластерного анализа теоретически может быть получено путем перебора всех допустимых вариантов.

При оптимальном разбиении получается и оптимальное значение целевой функции. Однако, стоит отметить, что данный процесс маловероятен, кроме условия, когда количество объектов n небольшое.

Также при процедуре полного перебора некоторые кластеры для различных вариантов разбиения совпадают. Таким образом, это приводит к дополнительным, а самое главное, лишним вычислениям.

Исходя из этого, при кластерном анализе методом полного перебора желательно иметь такой алгоритм, которые бы минимизировали излишние вычисления.

При кластерном анализе методом полного перебора требуется только матрица исходных данных. Все вычисления основаны только на ней. Однако объем таких вычислений огромен, и решение задачи остается трудоемким даже при высоком быстродействии современных вычислительных устройств.

2.2 Алгоритм иерархической последовательной кластеризации

Схема последовательной кластеризации может быть описана следующим образом. Рассмотрим $I = \{I_1, I_2, \dots, I_n\}$ как множество кластеров $\{I_1\}, \{I_2\}, \dots, \{I_n\}$; выберем кластеры I_i и I_j , которые наиболее близки друг к другу, и объединим их в один кластер.

Новое множество кластеров, состоящее уже из $n-1$ кластеров, будет $\{I_1\}, \{I_2\}, \dots, \{I_i, I_j\}, \dots, \{I_n\}$ [14].

Повторяя процесс, мы получим последовательные множества кластеров, количество кластеров которых уменьшается на единицу. В конце этой процедуры получится кластер, состоящий из n объектов и совпадающий с первоначальным множеством. $I = \{I_1, I_2, \dots, I_n\}$ [14].

В качестве меры расстояния был принят квадрат евклидова расстояния. Для наглядности вычислим матрицу

$$D = \{d_{ij}^2\},$$

где d_{ij}^2 – квадрат расстояния евклидова между I_i и I_j .

В таблице 1 представлена матрица квадратов евклидова расстояния [14].

Таблица 1 – Матрица квадратов евклидова расстояния

	I_1	I_2	I_3	...	I_n
I_1	0	d_{12}^2	d_{13}^2	...	d_{1n}^2
I_2		0	d_{23}^2	...	d_{2n}^2
I_3			0	...	d_{3n}^2
...			
I_n					0

Допустим, что

$$d_{pq}^2 = \min \{d_{ij}^2, i \neq j\}.$$

Вследствие этого кластеры I_i и I_j образуют новую группу $\{I_i, I_j\}$ [14].

И теперь строится матрица квадратов евклидова расстояния с учетом образованного кластеры. Размерность такой матрицы стала $(n-1) \times (n-1)$. Матрица расстояний представлена в таблице 2 [14].

Таблица 2 – Матрица квадратов евклидова расстояния на втором шаге

	$\{I_i, I_j\}$	I_1	I_2	I_3	...	I_n
$\{I_i, I_j\}$	0	d_{ij1}^2	d_{ij2}^2	d_{ij3}^2	...	d_{ijn}^2
I_1		0	d_{12}^2	d_{13}^2	...	d_{1n}^2
I_2			0	d_{23}^2	...	d_{2n}^2
I_3				0	...	d_{3n}^2
...				
I_n						0

Все элементы матрицы, не связанные с двумя объединенными объектами, остались такими же. Вычисления сводятся к минимуму, если можно определить d_{ijk}^2 , где $k = 1, 2, \dots, n; k \neq i \neq j$.

Ланс и Уильямс предложили вариант, в котором вычисления матрицы расстояний опираются только на значения расстояний в предыдущей матрице. Данный вариант предполагает использование всех вариантов расстояний, представленных в подразделе 1.4 [14], [15].

Таким образом выше написанное можно объединить в алгоритм, состоящий из 5 шагов, который представлен в работе [16].

1. Находится минимальное значение в матрице расстояний

$$d_{pq}^2 = \min (d_{ij}^2), \quad (2.2)$$

где $i = 1, \dots, j - 1$;

$j = 2, \dots, n$;

$n_i > 0$;

$n_j > 0$.

2. Увеличивается целевая функция по формуле

$$W_{i,j} = \frac{d_{i,j}^2}{2}. \quad (2.3)$$

Данный пункт производится при объединении двух кластеров I_p и I_q .

3. Кластер I_p заменяется на $I_p \cup I_q$. Производится перерасчет матрицы расстояний. Расстояние образованного кластера с другими кластерами осуществляется по формуле (для метода Уорда)

$$d_{ip}^2 = \frac{1}{n_i + n_r} \left[(n_i + n_p) d_{ip}^2 + (n_i + n_q) d_{iq}^2 - n_i d_{pq}^2 \right], \quad (2.4)$$

где $n_r = n_p + n_q$.

4. Полагается, что $n_p = n_p + n_q$ и $n_q = 0$, строка и столбец объекта I_q исключаются.

5. Записываются элементы кластера S_q в кластер S_p , возвращаются к первому пункту и процедура повторяется $p-2$ раз.

Ланс и Уильямс получили уравнение, аналогичное (2.4) и верное для всех пяти процессов кластеризации, описанных в подразделе 1.4. Это уравнение имеет вид

$$d_{hk}^2 = \alpha_i d_{hi}^2 + \alpha_j d_{hj}^2 + \beta d_{ij}^2 + \gamma |d_{ip}^2 - d_{jp}^2|, \quad (2.5)$$

где α_i , α_j , β и γ — параметры, задающие вид процесса.

Значения параметров, входящих в формулу (2.5), соответствующие шести различным методам кластеризации, представлены ниже:

— минимальное локальное расстояние

$$\alpha_i = \alpha_j = 1/2,$$

$$\beta = 0,$$

$$\gamma = -1/2;$$

— максимальное локальное расстояние

$$\alpha_i = \alpha_j = 1/2,$$

$$\beta = 0,$$

$$\gamma = 1/2;$$

— медиана

$$\alpha_i = \alpha_j = 1/2,$$

$$\beta = -1/4,$$

$$\gamma = 0;$$

— среднее группы

$$\alpha_i = n_i / n_k,$$

$$\alpha_j = n_j / n_k,$$

$$\beta = \gamma = 0;$$

Изм	Лист	№ докум.	Подпись	Дата
-----	------	----------	---------	------

— центроидный метод

$$\alpha_i = n_i / n_k,$$

$$\alpha_j = n_j / n_k,$$

$$\beta = -\alpha_i \alpha_j,$$

$$\gamma = 0;$$

— метод Уорда

$$\alpha_i = (n_h + n_i) / (n_h + n_k),$$

$$\alpha_j = (n_h + n_j) / (n_h + n_k),$$

$$\beta = -n_h / (n_h + n_k),$$

$$\gamma = 0.$$

Представленный в этом разделе алгоритм последовательной кластеризации лег в основу разработки алгоритма приложения кластерного анализа.

2.3 Разработка алгоритма работы приложения кластерного анализа

Основываясь на алгоритме, представленном в подразделе 2.2, разработан алгоритм работы приложения кластерного анализа. Он состоит из следующих шагов. Таблицы 3-7 и рисунок 8 являются примером вывода результатов данного шага в консольное приложение.

1. Исходные данные вносятся в матрицу $C[m,n]$, каждая строка m которой является отдельным кластером. Также определяется размер матрицы (Таблица 3).

Таблица 3 – Пример исходных данных $C[m,n]$

Объекты	Признаки											
	1	2	3	4	5	6	7	8	9	10	11	12
1	2	3	4	5	6	7	8	9	10	11	12	13
0	3	3	1	3	1	4	1	1	5	3	5	1
1	3	5	2	2	5	4	4	5	3	3	5	2
2	2	4	6	4	6	5	4	6	6	5	4	4
3	5	6	4	7	7	6	5	6	6	4	4	6
4	7	6	7	6	6	7	7	6	7	6	5	6

Окончание таблицы 3

Объекты	Признаки											
	1	2	3	4	5	6	7	8	9	10	11	12
1	2	3	4	5	6	7	8	9	10	11	12	13
5	6	8	8	6	7	8	8	6	8	7	8	6
6	6	7	6	8	7	6	9	6	6	9	9	6
7	8	7	7	9	7	9	9	9	8	8	7	7
8	8	7	7	10	10	8	7	10	7	8	8	7
9	1	4	10	5	9	6	6	4	5	9	3	7

2. Вычисляется расстояние (в данной работе – квадрат евклидова расстояния) между отдельными кластерами и образуется матрица расстояний, а также внутригрупповой суммы квадратов (ВСК), равной половине евклидова расстояния (Таблица 4).

Таблица 4 – Матрица ВСК $W[m,m]$

Объекты	0	1	2	3	4	5	6	7	8	9
0	0	26	51,5	73,5	101,5	138,5	141	188	207	134
1		0	21,5	35,5	59,5	86,5	88	126	134	86
2			0	17	27	47	56,5	76,5	83,5	31,5
3				0	13	32	36,5	43,5	47,5	48,5
4					0	10	19,5	19,5	32,5	41,5
5						0	11,5	14,5	25,5	50,5
6							0	17	23	57
7								0	9	78
8									0	82
9										0

3. В матрице расстояний определяется минимальное значение, и отмечаются строки (номера) кластеров I и J , на пересечении которых было найдено данное расстояние (Таблица 5).

Таблица 5 - Вывод минимального значения в матрице $W[m,m]$

Минимальное значение	9,0
Строка	7
Столбец	8

4. Найденное расстояние также записывается в массив расстояний, образованный для записи расстояний на каждом шаге процедуры объединения.

5. Для объединенного кластера пересчитываются расстояния, образуя новую матрицу расстояний размерностью $(m-1) \times (m-1)$ (Таблица 6).

Таблица 6 – Матрица ВСК $W[m-1,m-1]$

Объекты	0	1	2	3	4	5	6	7	8
0	0	26,00	51,50	73,50	101,50	138,50	141,00	260,30	134,00
1		0	21,50	35,50	59,50	86,50	88,00	170,30	86,00
2			0	17,00	27,00	47,00	56,50	103,70	31,50
3				0	13,00	32,00	36,50	57,70	48,50
4					0	10,00	19,50	31,70	41,50
5						0	11,50	23,70	50,50
6							0	23,70	57,00
7								0,00	103,67
8									0

6. Шаги 4-6 повторяются до тех пор, пока не останется один кластер, которым является $C[m,n]$, то есть $m-2$ раз.

7. Выводятся номера объединенных кластеров (исходных) и расстояние, на котором было произведено объединение (Таблица 7).

Таблица 7 – Вывод номеров объединяемых кластеров и расстояния

Объединение	Расстояние
7,8	9,00
4,5	10,00
2,3	17,00
4,5,6	17,33
0,1	26,00
4,5,6,7,8	36,46
2,3,9	47,66
0,1,2,3,9	119,33
0,1,2,3,9,4,5,6,7,8	267,00

8. отображаются результаты кластерного анализа для наглядности процесса объединения кластеров.

9. Варианты графического представления иерархической последовательной кластеризации будут представлены в подразделе 2.4.

Блок схема процедуры кластерного анализа в приложении представлена в приложении А.

2.4 Визуализация результатов кластерного анализа

Одним из вариантов визуального отображения результатов кластерного анализа является диаграмма вложения, пример которой представлен на рисунке 8.

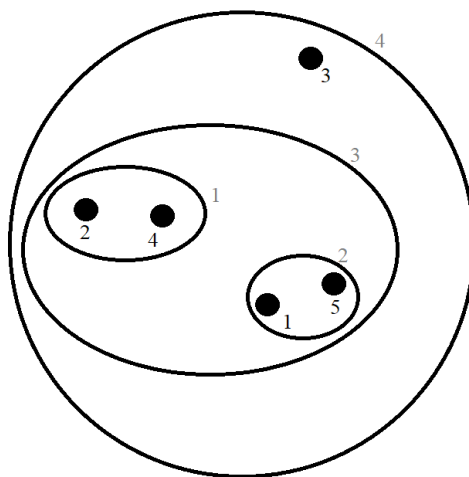


Рисунок 8 – Пример диаграммы вложения

Диаграмма вложения отображает каждым кругом новое объединение. Таким образом, диаметр, охватывающий кластеры, с каждым шагом увеличивается. Исходя из этого, стоит сделать вывод, что данный вариант отображения результатов кластеризации возможен только при небольшом количестве выборки. Также сложна и программная реализация диаграммы вложения.

Таким образом, данный вариант отображения итогов процесса кластеризации не используется в ВКР.

В различных источниках графическое представление иерархических классификаций имеет различное представление [17]. Некоторые варианты представлены на рисунке 9.

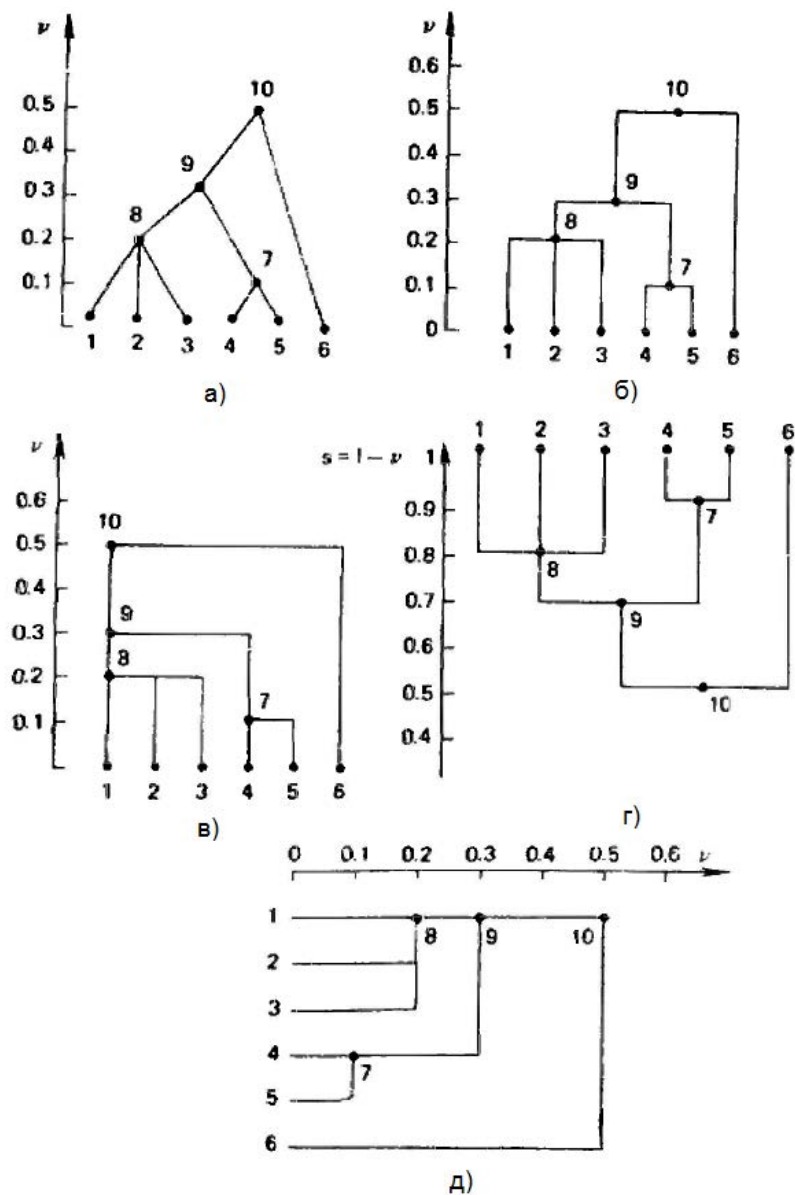


Рисунок 9 – Варианты графического представления кластеризации [17]

Такие графические представления иерархических классификаций называют дендрограммой.

Дендрограммой называется древовидная диаграмма, отображающая объединения объектов в кластеры на каждом шаге последовательной кластеризации.

Пример часто применяемого вида дендрограммы представлен на рисунке 10.

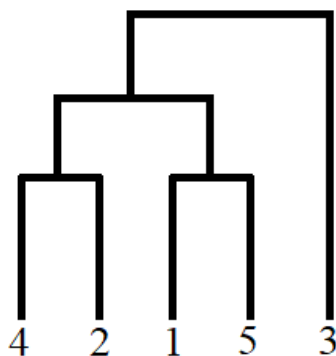


Рисунок 10 – Пример часто применяемого вида дендрограммы

Также он представлен на рисунке 9, б.

В дендрограмме объекты могут располагаться вертикально или горизонтально.

Дендрограмму в разных источниках называют также деревом объединения кластеров или иерархической структуры, древовидной схемой.

Существует много способов построения дендрограмм. Один из вариантов предстоит разработать в рамках разработки приложения кластерного анализа.

В выпускной квалификационной работе будет использоваться дендрограмма с вертикальным расположением объектов.

Вывод по разделу 2: были рассмотрен метод кластеризации полным перебором и получена формула расчета общего числа разбиений. Также была рассмотрена рекурсивная процедура иерархического последовательного кластерного анализа Ланса-Уильямса и определена формула для расчета расстояний между кластерами для метода Уорда на следующих шагах объединения. Был разработан алгоритм работы приложения кластерного анализа, основанный на рассмотренном алгоритме последовательной кластеризации.

Также были рассмотрены варианты визуализации результатов кластерного анализа. В итоге результаты кластерного анализа будут отображены графически посредством построения дендрограммы с вертикальным расположением объектов.

3 РАЗДЕЛ РАЗРАБОТКИ ПРИЛОЖЕНИЯ

3.1 Разработка листинга программы

В разделе 2 был разработан алгоритм процедуры кластерного анализа.

Реализуем данный алгоритм в программном коде на языке C# в интерфейсе программирования приложений (API) Windows Forms, входящий в состав среды разработки MS Visual Studio.

Microsoft Visual Studio является средой разработки для написания, отладки, тестирования и развертывания кода на любой платформе [18].

C# является универсальным, безопасным в отношении типов, объектно-ориентированным языком программирования, целью которого является обеспечение продуктивности работы программистов. В данном языке программирования существует баланс между производительностью, простотой и выразительностью, что является одним из его достоинств [19].

Вся процедура кластерного анализа и построения дендрограммы содержится в функции Cluster(). Стоит отметить, что для работоспособности данной программы необходимо разрешить небезопасный код. На рисунке 11 представлено окно свойств проекта, в котором устанавливается данное разрешение.

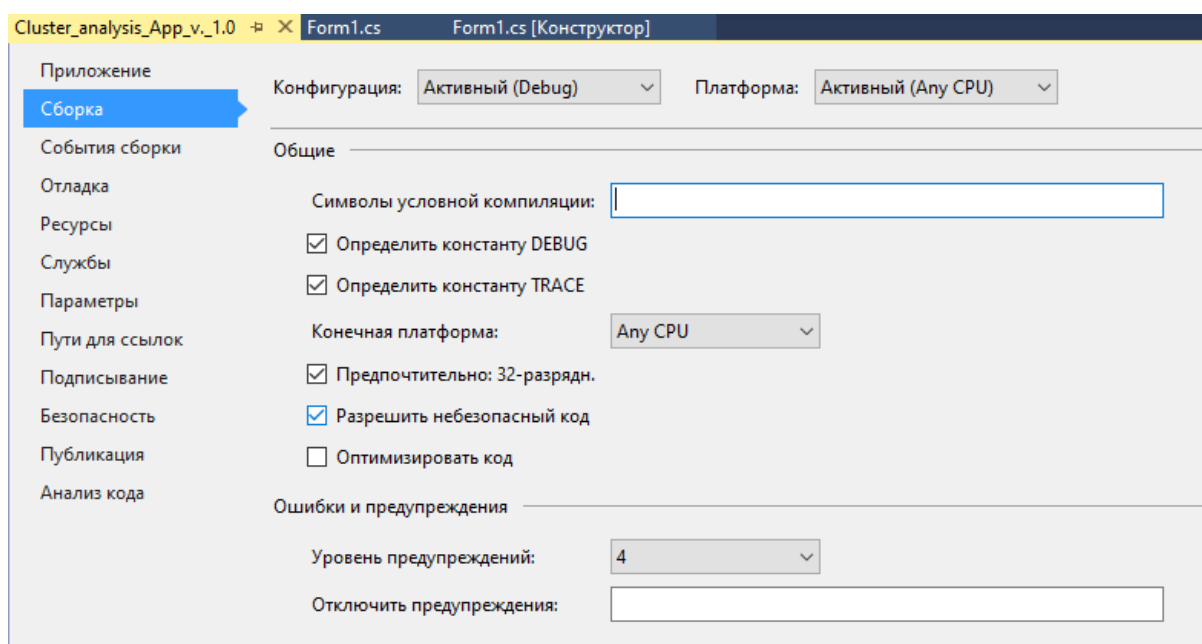


Рисунок 11 – Окно свойств проекта в MS Visual Studio

Выделим основные моменты в написанном коде.

Прежде всего, дадим определения основным элементам, используемым в данном приложении. Их всего три.

Элемент RichTextBox используется для работы с текстом и вывода данных [20].

Элемент DataGridView отображает данные в табличном формате [21].

Элемент Chart — это объект диаграммы, предоставляющий события [22].

Функция начинается с очистки данных с графика и текстового поля и проверки загрузки файла.

```
chart1.Series.Clear();  
richTextBox1.Clear();  
if (openFileDialog1.ShowDialog() == DialogResult.OK)
```

После загрузки окна инициализируются все необходимые переменные. Из файла получаем данные и вносим в массив *C* размерностью *m* x *p*, где *m* - количество объектов, а *p* - количество признаков.

```
string[] lines = File.ReadAllLines(openFileDialog1.FileName);  
int m = lines.Length;  
int p = lines[0].Split('\t').Length;  
double[,] C = new double[lines.Length, lines[0].Split('\t').Length];  
for (i = 0; i < lines.Length; i++)  
{  
    string[] temp = lines[i].Split('\t');  
    for (j = 0; j < temp.Length; j++)  
        C[i,j]=Convert.ToDouble(temp[j]);  
}
```

Затем данная матрица выводится в элемент DataGridView.

```
dataGridView1.RowCount = m + 1;  
dataGridView1.ColumnCount = p + 1;  
dataGridView1.Rows[0].Cells[0].Value = "Об. | пр.";  
dataGridView1.Rows[0].Cells[0].Style.BackColor = Color.FromArgb(255, 255,  
120);
```

```

for (i = 0; i < m + 1; i++)
    if (i != 0)
        for (j = 0; j < p + 1; j++)
            if (j == 0)
                {
                    dataGridView1.Rows[i].Cells[j].Value = i;
                    dataGridView1.Rows[i].Cells[j].Style.BackColor =Color.FromArgb(255,
255, 120);
                }
            else
                dataGridView1.Rows[i].Cells[j].Value = C[i - 1,j - 1];
            else
                for (j = 1; j < p + 1; j++)
                    {
                        dataGridView1.Rows[i].Cells[j].Value = j;
                        dataGridView1.Rows[i].Cells[j].Style.BackColor
=Color.FromArgb(255, 255, 120);
                    }

```

Далее формируется матрица расстояний

```

for (i = 0; i < a; i++)
    {
        for (j = 0; j < a; j++)
            {
                s = 0;
                for (k = 0; k < p; k++)
                    s += (C[i,k] - C[j,k]) * (C[i,k] - C[j,k]);
                D[i, j] = s;
                W[i, j] = n[i] * n[j] / (n[i] + n[j]) * D[i, j];
            }
    }

```

В результате получаем две матрицы: квадратов евклидова расстояния (D) и ВСК (W) размерностью $a \times a$. Расчет матрицы ВСК произведен по формуле (1.1) Переменная a отображает количество оставшихся кластеров и с каждым шагом уменьшается на единицу.

Следующим шагом алгоритма является нахождение минимального значения в матрице расстояний.

```
double min = Math.Pow(10, 10);
for (i = 0; i < a; i++)
{
    for (j = 0; j < a; j++)
    {
        if (D[i, j] < min & i != j)
        {
            min = W[i, j];
            t1 = i;
            t2 = j;
        }
    }
}
r[b] = min; //Запись расстояния, на котором объединяются кластеры
```

Массив $r[m-1]$ необходим для вывода расстояния, на котором происходило объединение двух кластеров, обозначенных $t1$ и $t2$.

Далее преобразуем матрицу расстояний

```
for (i = 0; i < a; i++)
{
    if (i == t1)
    {
        for (j = 0; j < a; j++)
        {
            if (j == t1)
```



```

{
    A[i, j] = D[i, j];
}
else if (j == t2)
{
    A[i, j] = 0;
}
else
    A[i, j] = 1 / (n[j] + nr) * ((n[j] + n[t1]) * D[t1, j] + (n[j] + n[t2]) * D[t2, j] -
n[j] * D[t1, t2]);
}
}

else if (i == t2)
{
    for (j = 0; j < a; j++)
        A[i, j] = 0;
}
else
    for (j = 0; j < a; j++)
    {
        if (j == t1)
        {
            A[i, j] = 1 / (n[i] + nr) * ((n[i] + n[t1]) * D[i, t1] + (n[i] + n[t2]) * D[i,
t2] - n[i] * D[t1, t2]);
        }
        else if (j == t2)
        {
            A[i, j] = 0;
        }
    }
else

```

```
A[i, j] = D[i, j];
```

```
}
```

```
}
```

После образования нового кластера, данные кластера t_2 удаляются из матрицы.

```
for (i = 0; i < a; i++)
```

```
{
```

```
  if (i == t2)
```

```
  {
```

```
    for (k = t2; k < a - 1; k++)
```

```
      for (j = 0; j < a; j++)
```

```
        A[k, j] = A[k + 1, j];
```

```
        i = k;
```

```
    }
```

```
  else if (i == a - 1)
```

```
  {
```

```
    for (j = 0; j < a; j++)
```

```
      A[i, j] = 0;
```

```
    }
```

```
  else
```

```
    for (j = 0; j < a; j++)
```

```
      A[i, j] = A[i, j];
```

```
  }
```

```
for (i = 0; i < a; i++)
```

```
{
```

```
  for (j = 0; j < a; j++)
```

```
  {
```

```
    if (j == t2)
```

```
    {
```

```
      for (k = t2; k < a - 1; k++)
```

```

        A[i, k] = A[i, k + 1];
        j = k;
    }
    else if (j == a - 1)
        A[i, j] = 0;
    else
        A[i, j] = A[i, j];
    }
}
a--;

```

Последняя строка обозначает уменьшение переменной a на единицу.

Дальше, начиная с нахождения минимального значения матрицы ВСК, цикл продолжится до тех пор, пока переменная a не станет равна 1.

Вывод результатов последовательного объединения кластеров в RichTextBox осуществляется следующим образом.

```

for (i = 0; i < m - 1; i++)
{
    richTextBox1.Text += Convert.ToString(i + 1) + "-е объединение кластеров: ";
    for (j = 0; j < m; j++)
    {
        if (x1[i, j] != 0)
            richTextBox1.Text += Convert.ToString(x1[i, j]) + " ";
        else
            break;
    }
    richTextBox1.Text += "на расстоянии " + Convert.ToString(string.Format(
("{0:0.00}", r[i])) + "\n\n");
}

```

Для создания дендрограммы осуществляются те же шаги, но для избегания пересечений на ней, матрица исходных объектов сортируется.

Настройка осей графика.

```
chart1.ChartAreas[0].AxisY.Maximum = r[m - 2]; // максимальное значение на  
оси Y
```

```
chart1.ChartAreas[0].AxisY.ScaleView.Zoom(0, r[m - 2] / 2);
```

```
chart1.ChartAreas[0].AxisY.Title = "Расстояние";
```

```
chart1.ChartAreas[0].AxisX.Title = "Объекты";
```

```
chart1.ChartAreas[0].AxisY.Minimum = 0; // минимальное значение на оси Y
```

```
chart1.ChartAreas[0].AxisY.Maximum = r[m - 2]; // максимальное значение на  
оси Y
```

```
chart1.ChartAreas[0].AxisY.Interval = (Math.Round(r[m - 2]/10, 0,  
MidpointRounding.AwayFromZero)+1)*10/20; // интервал на оси Y
```

```
chart1.ChartAreas[0].AxisX.Interval = 0.5; // интервал на оси X
```

```
chart1.ChartAreas[0].AxisX.Minimum = 0; // минимальное значение на оси X
```

```
chart1.ChartAreas[0].AxisX.Maximum = m+1; // максимальное значение на оси  
X
```

```
chart1.Legends.Clear();
```

На каждом шаге после выявления минимума строятся линии, которые соединяют два кластера. Вывод одной линии представлен ниже.

```
string q = Convert.ToString("union" + (c + 1) + ".1");
```

```
chart1.Series.Add(new Series(q));
```

```
chart1.Series[q].ChartType =
```

```
System.Windows.Forms.DataVisualization.Charting.SeriesChartType.Line;
```

```
chart1.Series[q].BorderWidth = 3;
```

```
chart1.Series[q].Color = Color.FromArgb(0, 0, 255);
```

```
chart1.Series[q].Points.AddXY(z1[t1], d[t1]);
```

```
chart1.Series[q].Points.AddXY(z1[t1], r[b]);
```

Примеры вывода данных на данные элементы представлены в подразделе 3.2. Листинг программы представлен в приложении Б и имеет 571 строку кода, написанного на языке C#.

3.2 Разработка графического интерфейса приложения

Разработанное приложение кластерного анализа результатов экспертного оценивания получило название «Cluster analysis App v 1.0».

Диалоговый интерфейс приложения кластерного анализа экспертного оценивания, разрабатываемый в интерфейсе программирования приложений (API) Windows Forms [23], входящий в состав среды разработки MS Visual Studio [18].

При запуске программы появляется диалоговое окно, представленное на рисунке 12.

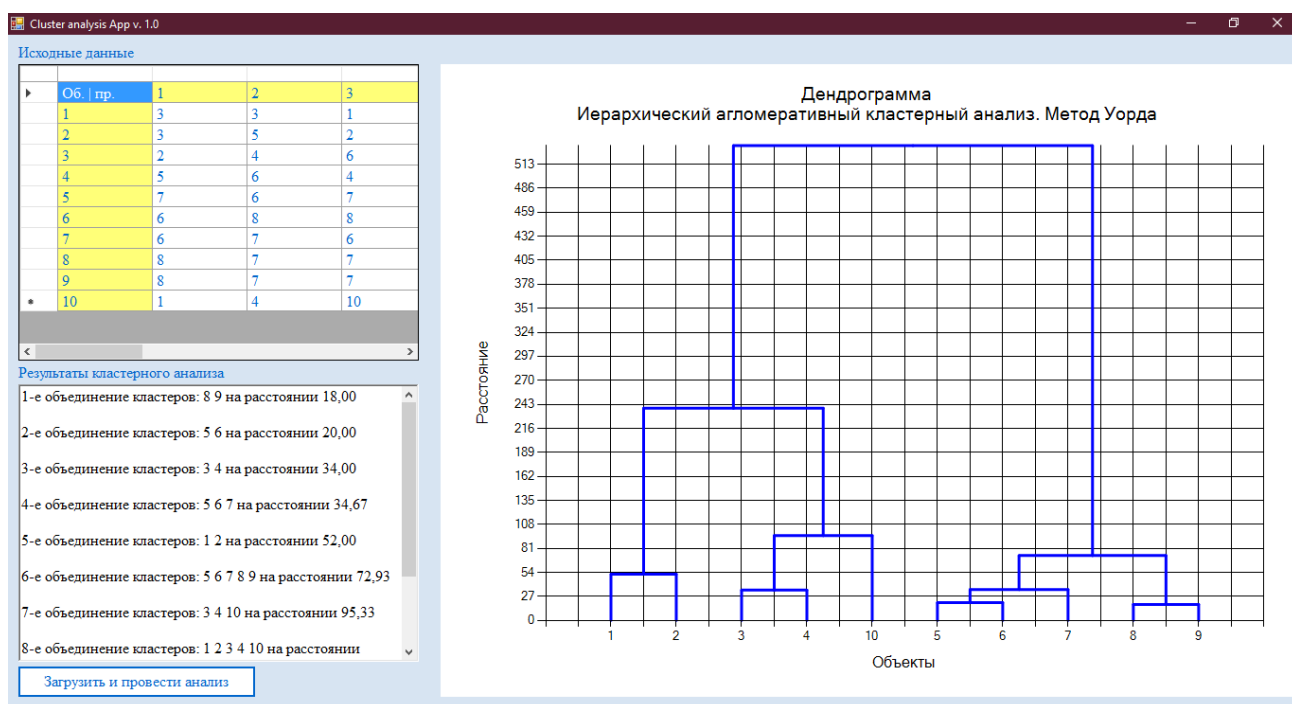


Рисунок 12 – Диалоговое окно приложения

Основным элементом в окне данного приложения, представленного на рисунке 12, является кнопка «Загрузить и провести анализ», которая реализует практически всю программу кластерного анализа.

Исходные данные

	Об. гр.	1	2	3
▶	1	3	3	1
	2	3	5	2
	3	2	4	6
	4	5	6	4
	5	7	6	7
	6	6	8	8
	7	6	7	6
	8	8	7	7
	9	8	7	7
*	10	1	4	10

Рисунок 14 – Элемент DataGridView с исходными данными

В элементе RichTextBox расположенном под надписью «Результат кластерного анализа», выводится результат каждого шага объединения, номера объединяемых кластеров и расстояние, на котором произошло это объединение. Пример представлен на рисунке 15.

Результаты кластерного анализа

1-е объединение кластеров: 8 9 на расстоянии 18,00
 2-е объединение кластеров: 5 6 на расстоянии 20,00
 3-е объединение кластеров: 3 4 на расстоянии 34,00
 4-е объединение кластеров: 5 6 7 на расстоянии 34,67
 5-е объединение кластеров: 1 2 на расстоянии 52,00
 6-е объединение кластеров: 5 6 7 8 9 на расстоянии 72,93
 7-е объединение кластеров: 3 4 10 на расстоянии 95,33
 8-е объединение кластеров: 1 2 3 4 10 на расстоянии

Рисунок 15 – Элемент RichTextBox с выводом результатов пошагового объединения кластеров

Справа в элементе Chart выводится дендрограмма – визуализация результатов кластерного анализа, представленных в элементе RichTextBox. Пример представлен на рисунке 16.

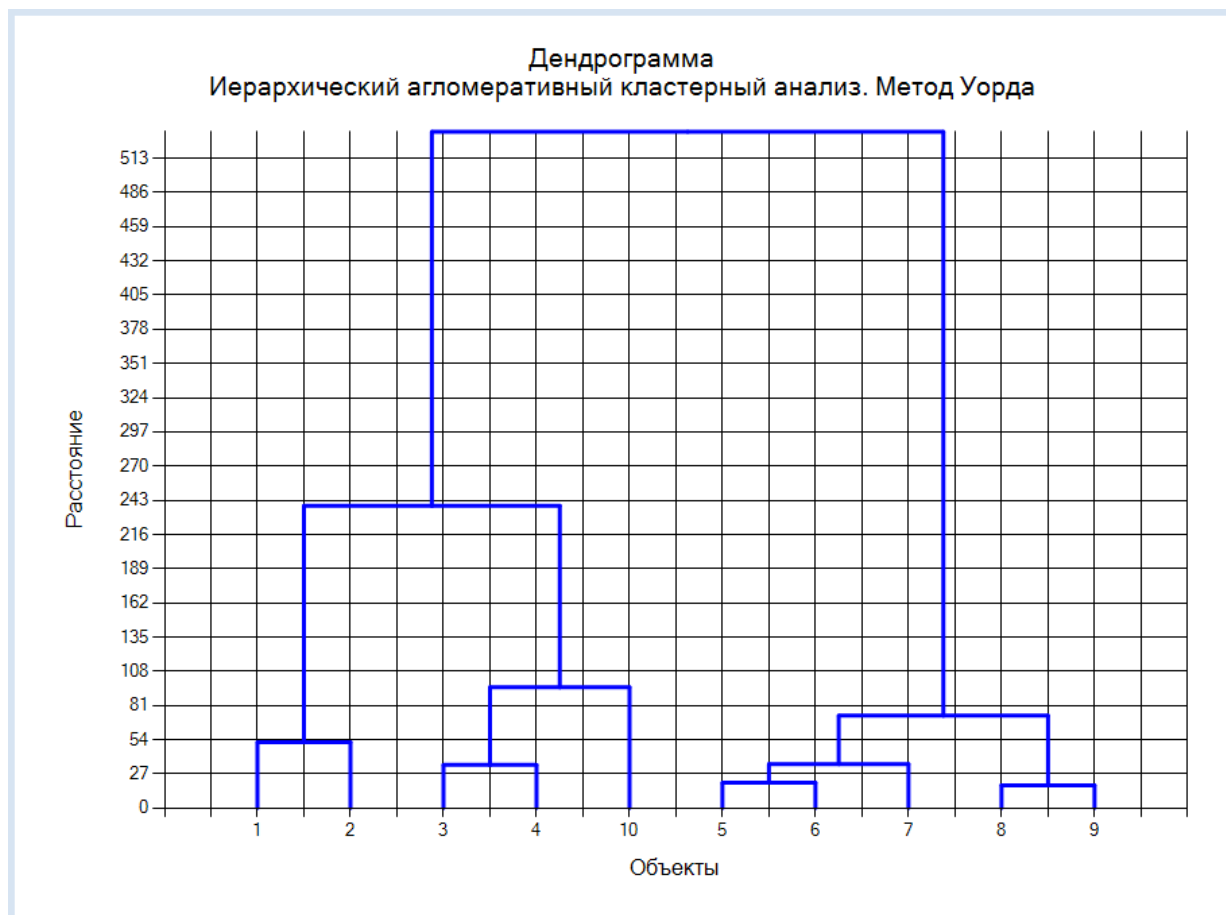


Рисунок 16 – Элемент Chart с выводом дендрограммы

Данное приложение предназначено для проведения иерархического агломеративного кластерного анализа с методом связи кластеров Уорда.

Необходимые условия:

- количество признаков для всех объектов одинаковое;
- загружаются уже нормированные выборки;
- наличие в текстовом файле только значений.

Последний пункт говорит о том, что необходимо приготовить файл формата «*.txt» с выборкой данных в виде матрицы перед тем, как начинать кластерный анализ (Рисунок 17).

4 РАЗДЕЛ ТЕСТИРОВАНИЯ

4.1 Тестирование приложения на примере множества точек в двумерном пространстве

Проведем проверку работы приложения кластерного анализа на примере множества точек в двумерном пространстве. Заполним случайным образом значения координат данных точек.

Расположение их в декартовой системе координат представлено на рисунке 18.

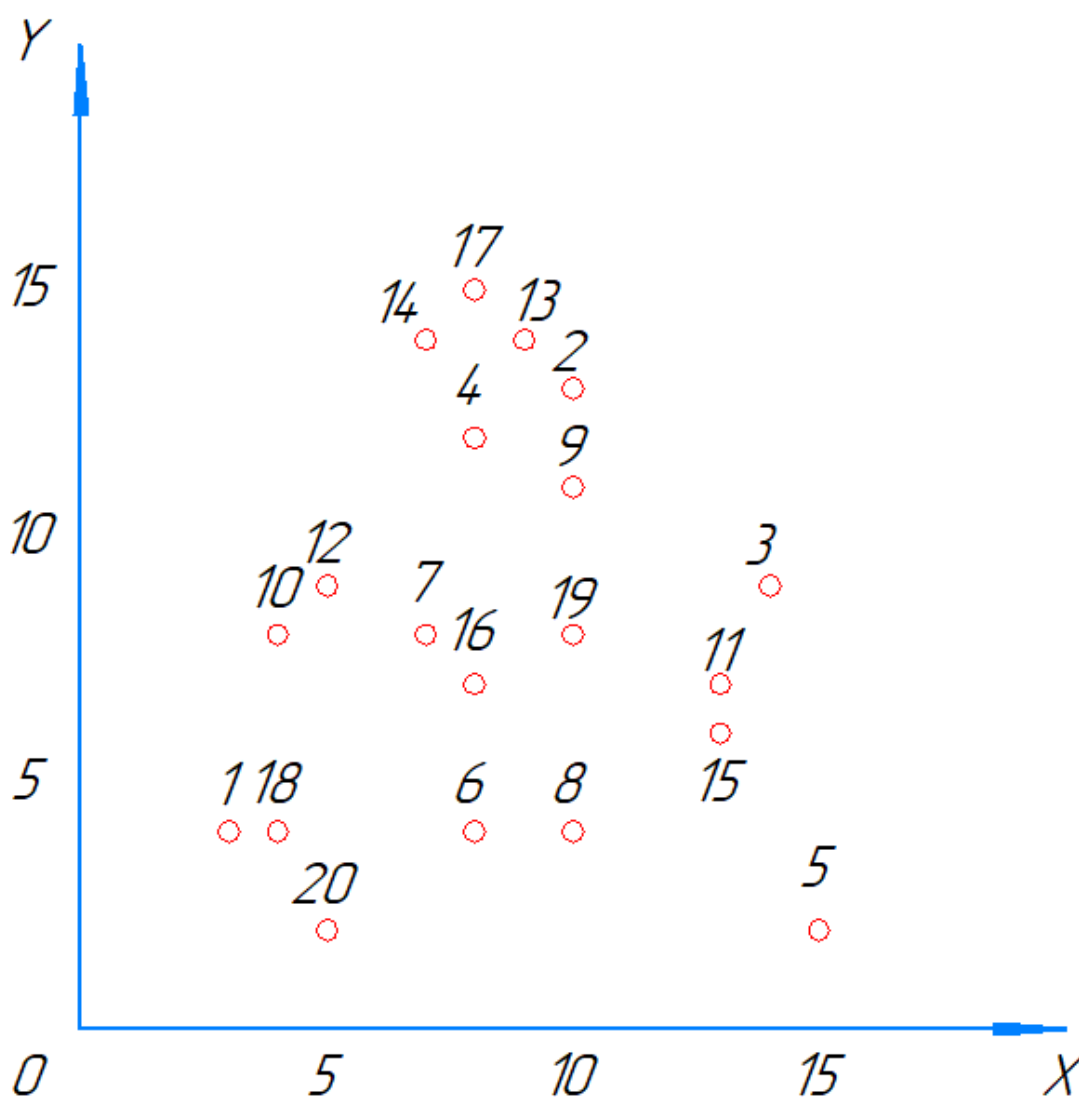


Рисунок 18 – Точки, расположенные на плоскости координат XY

Заполним массив в текстовый документ и проведем кластерный анализ, на рисунке 19 представлен результат кластерного анализа.



Рисунок 19 – Дендрограмма результатов кластерного анализа массива точек

Дендрограмма в масштабе по оси расстояний представлена на рисунке 20.

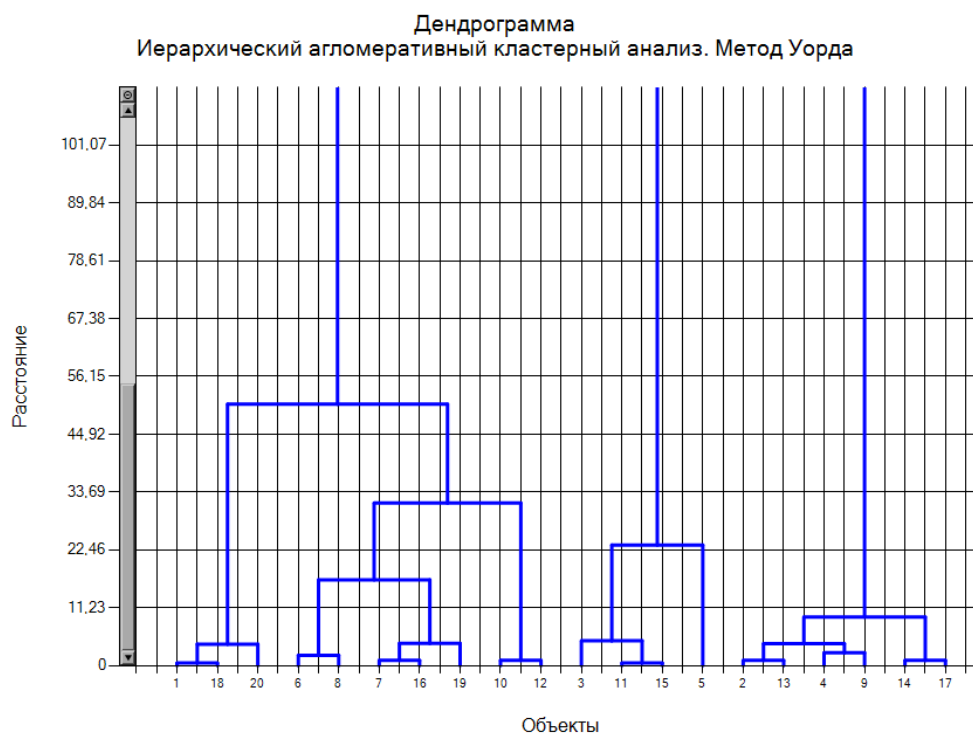


Рисунок 20 – Дендрограмма результатов кластерного анализа массива точек (в масштабе)

Изм	Лист	№ докум.	Подпись	Дата

Из получившихся результатов на рисунках 19-20 видим, что образовались две крупные группы точек. Отообразим точки, вошедшие в тот или иной кластер, в таблице 8.

Таблица 8 – Разбиение объектов на кластеры

№ кластера	Объекты, вошедшие в кластер
1	1 18 20 6 8 7 16 19 10 12
	3 11 15 5
2	2 13 4 9 14 17

Также отобразим данные группы на графике, который был представлен на рисунке 18. Процесс объединения посредством построения графов представлен на рисунке 21.

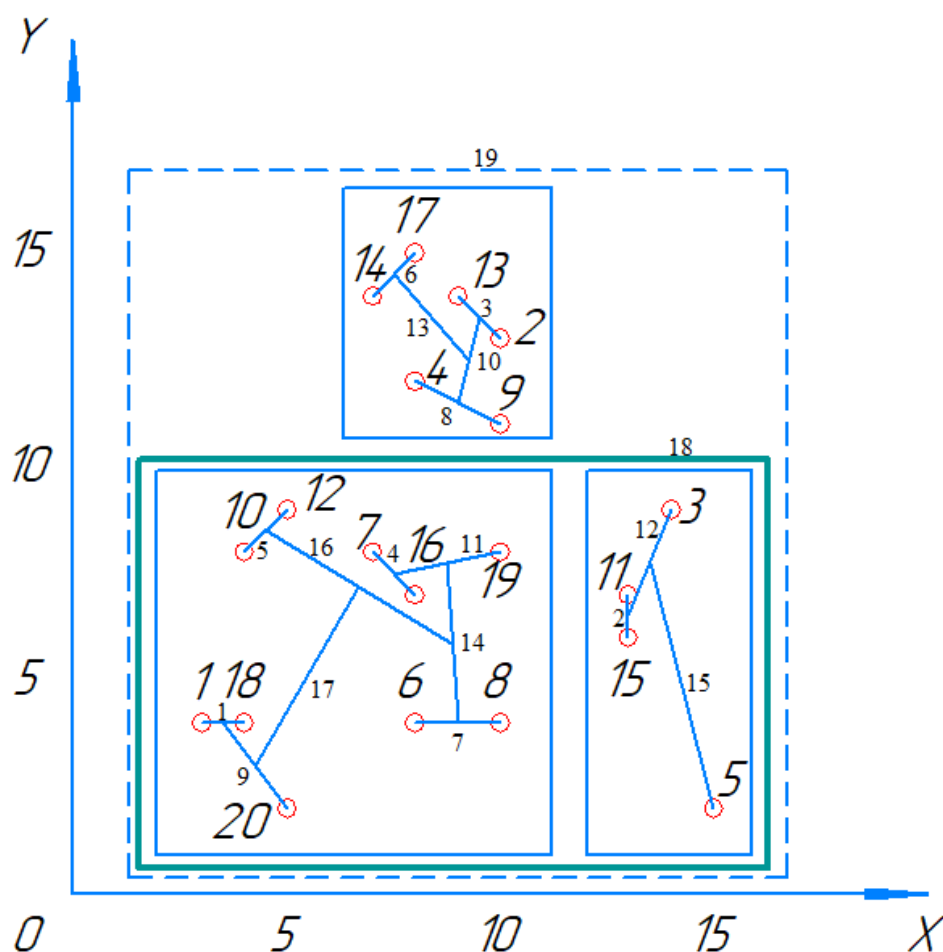


Рисунок 21 – Процесс объединения точек в кластеры с помощью графов на координатной плоскости XY

Таким образом, из полученных результатов видно, что путем кластерного анализа можно образовать группы близко расположенных друг к другу точек.

4.2 Тестирование приложения в многомерном пространстве

Проведем следующий опыт в 10-мерном пространстве. Образует выборку из 10 объектов и 45 признаков. Заполним следующим образом, представленным на таблице 9. Таблица с исходными данными представлена в приложении В.

Таблица 9 – Правила заполнения двумерного массива случайными числами

Объекты	Минимальное значение	Максимальное значение
1,2	1	3
3,4	4	6
5,6	7	8
7,8	9	10
9,10	1	10

Таким образом, формируются строго обособленные кластеры. Их количество должно быть равно четырем. Также присутствуют два объекта, которые заполнены случайными числами от 1 до 10, то есть любое значения. Необходимость нормирования отсутствует, поскольку значения получены в одном диапазоне.

Проведем кластерный анализ. Результат пошагового объединения кластеров представлен в таблице 10.

Таблица 10 – Результат объединения кластеров

Шаг объединения	Кластер 1	Кластер 2	Расстояние
1	2	3	4
1	7	8	9,50
2	5	6	11,50
3	3	4	35,50

Окончание таблицы 10

Шаг объединения	Кластер 1	Кластер 2	Расстояние
1	2	3	4
4	1	2	38,50
5	5,6	7,8	181,50
6	3,4	9	238,50
7	3,4,9	10	340,25
8	1,2	3,4,9,10	685,92
9	1,2,3,4,9,10	5,6,7,8	2123,83

Результат представлен на рисунке 22.

Дендрограмма
Иерархический агломеративный кластерный анализ. Метод Уорда

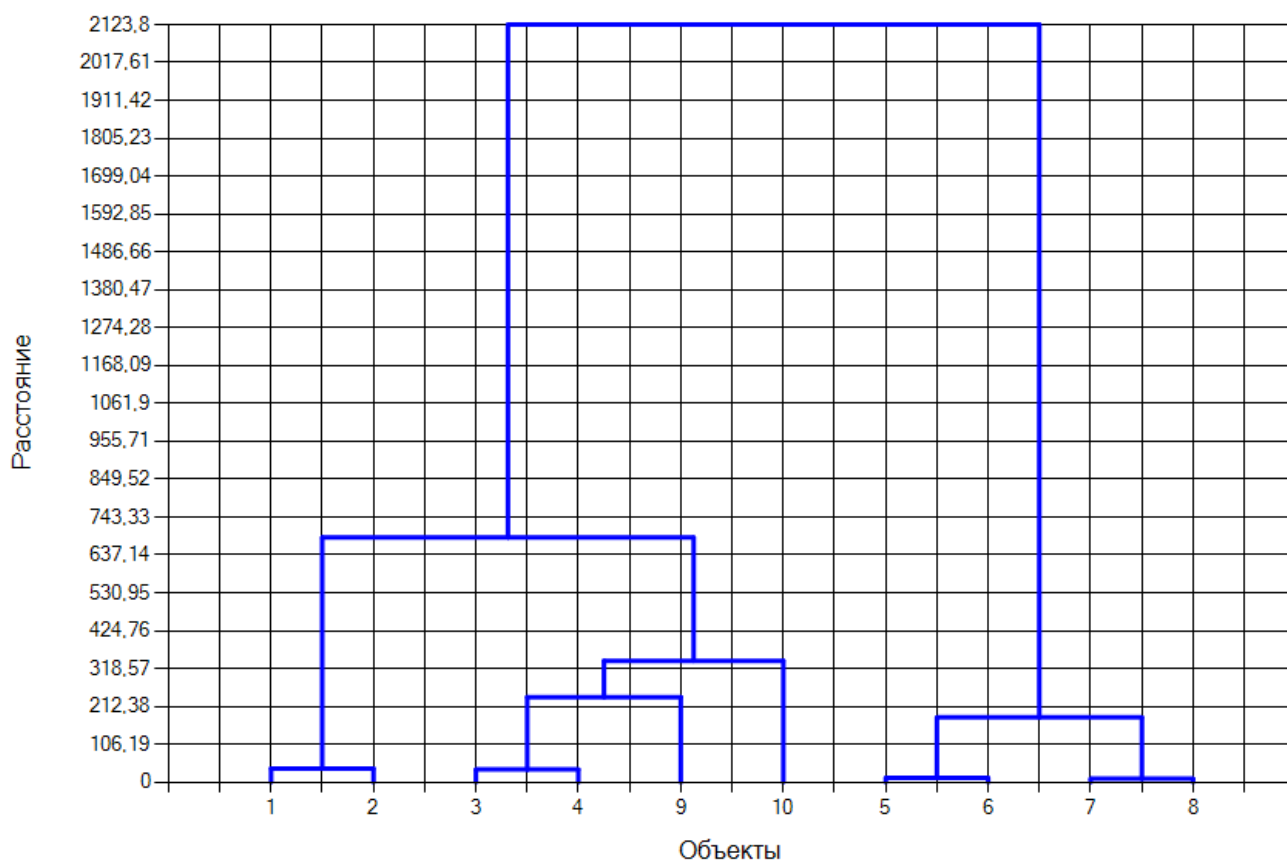


Рисунок 22 – Дендрограмма результатов кластерного анализа

Исходя из данных, полученных в таблице 10 видно, что в первую очередь происходит объединение тех кластеров, для которых был задан строго определенный диапазон, то есть объединение объектов 1,2; 3,4; 5,6; 7,8.

Объекты под номером 9 и 10 могли присоединиться к любой из 4 групп кластеров. Таким образом, объект 9 был присоединен к кластеру {3,4}, а объект 10 следом был присоединен к образованному кластеру {3,4,9}.

Вполне возможно выдвинуть гипотезу, что если объекты имеют значения по всему диапазону (в данном случае от 1 до 10), то с большей вероятностью среднее значение по всем признакам данного объекта будет в районе середины диапазона.

4.3 Проведение кластерного анализа результатов анкетирования студентов посредством приложения

Сформируем выборку данных из результатов проведенного анкетирования студентов-выпускников бакалавриата по вопросам компетенций по направлению 12.03.01 «Приборостроение», представленные в [24], на котором они в данное время обучаются, созданного в Google Forms [25]. Тема опроса, ряд вопросов и внешний вид формы представлены на рисунках 23-24.

Как вы оцениваете овладение ключевыми компетенциями по своей специальности за время обучения в университете?

Доброго времени суток, уважаемые студенты. Вам предлагается ответить на следующие вопросы. Обращаем внимание, что анкета анонимная.

* Обязательно

1. Способность формировать мировоззренческую позицию на основе философских знаний *

0 1 2 3 4 5 6 7 8 9 10

Рисунок 23 – Форма анкетирования студентов, образованная в Google Forms

2. Способность анализировать основные этапы и закономерности исторического развития общества для формирования гражданской позиции *

0 1 2 3 4 5 6 7 8 9 10

3. Способность использовать основы экономических знаний в различных сферах деятельности *

0 1 2 3 4 5 6 7 8 9 10

4. Способность использовать основы правовых знаний в различных сферах деятельности *

0 1 2 3 4 5 6 7 8 9 10

Рисунок 24 – Ряд некоторых вопросов в анкете в Google Forms

Перечень всех вопросов анкетирования представлен в приложении В.

Для проведения последующего анализа результатов кластерного анализа были заданы дополнительные вопросы, которые не учтены в образовании выборки: пол, курс, форма обучения, оценка посещаемости, работал или работает ли студент, занимается ли внеучебной деятельностью.

Выборка результатов анкетирования представлена в приложении Г.

Проведем кластерный анализ по данной выборке. Результат кластерного анализа представлен в виде дендрограммы, изображенной на рисунке 25.

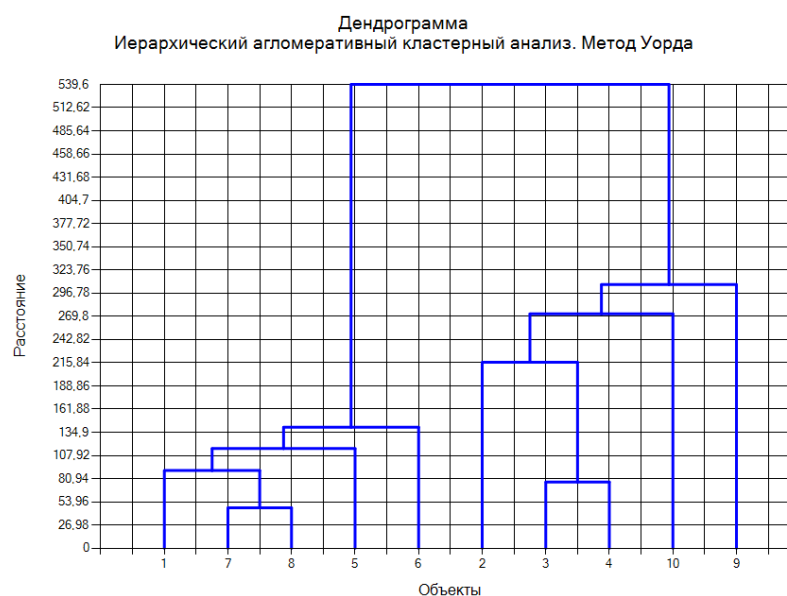


Рисунок 25 – Дендрограмма результатов кластерного анализа

Таким образом, исходя из данных на рисунке 25, видно, что было образовано две крупные группы студентов.

4.4 Анализ результатов кластерного анализа

В подразделе 4.3 были получены результаты кластерного анализа на основе данных анкетирования студентов-выпускников бакалавриата по вопросам компетенции по направлению 12.03.01 Приборостроение, на котором они обучаются.

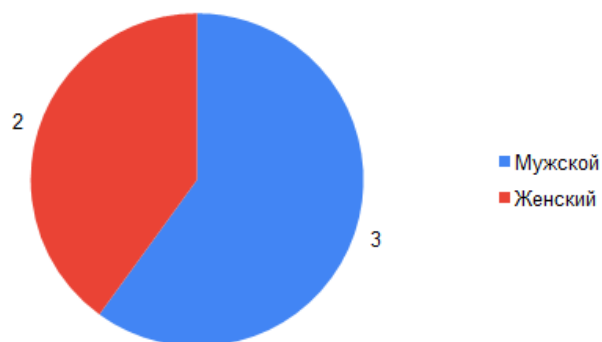
Проведем анализ результатов кластерного анализа. Из графика, представленного на рисунке 25, видно, что были образованы две группы студентов, в каждой из которой по пять человек. Также на последнем шаге это отображено в таблице 11.

Таблица 11 – Результат пошагового объединения кластеров

Шаг объединения	Кластер 1	Кластер 2	Расстояние
1	7	8	47,00
2	3	4	77,00
3	1	7,8	90,33
4	1,7,8	5	115,92
5	1,7,8,5	6	140,75
6	2	3,4	216,33
7	2,3,4	10	272,67
8	2,3,4,10	9	306,80
9	1,7,8,5,6	2,3,4,10,9	539,60

Проведем анализ полученных групп. Выведем в таблицу средние значения по группе и также средние значения по ожидаемому среднему баллу в дипломе. Результат представлен в таблице 12.

Распределение группы 1 по половому признаку



Распределение группы 2 по половому признаку

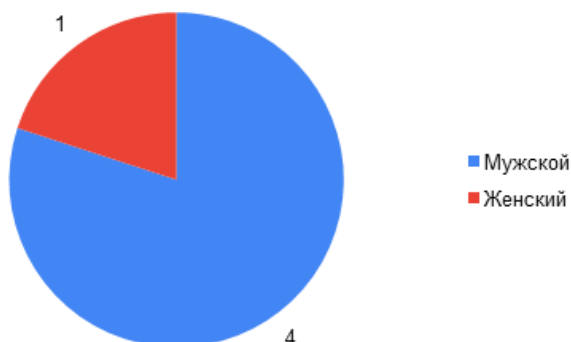


Рисунок 26 – Диаграмма соотношения количества студентов в каждой группе по половому признаку

На основе полученных данных на рисунке 26 нельзя однозначно утверждать, что по половому признаку можно отметить чье-либо преимущество, хоть и во второй группе преобладают студенты мужского пола.

На рисунке 27 представлены диаграммы распределения по признаку наличия опыта работы на предприятиях во время учебы в каждой группе.



Рисунок 27 – Диаграмма соотношения количества студентов в каждой группе по признаку наличия опыта работы во время учебы

На основе полученных данных на рисунке 27 можно обозначить, что студенты второй группы в большинстве своем сталкивались с реальным производством.

Некоторые параллельно с учебным процессом проводят часть времени на предприятии. Таким образом, стоит отметить, что у данной группы студентов, вероятно, более объективный взгляд на их уровень подготовки. В первой группе студентов большинство не работало на предприятии или работало не по специальности. Тем самым студенты не могут более объективно оценить свои способности, несмотря на проведение практических и лабораторных занятий.

Также подтверждаются слова известного философа Сократа «Я знаю, что я ничего не знаю», или более точная формулировка Альберта Эйнштейна: «Чем больше я учусь, тем больше понимаю, как много я не знаю».

На рисунке 28 представлены диаграммы распределения по признаку наличия внеучебной деятельности у студентов в каждой группе.



Рисунок 28 – Диаграмма соотношения количества студентов в каждой группе по признаку наличия внеучебной деятельности

Исходя из результатов, представленных на рисунке 28, нельзя однозначно утверждать о том, как влияет внеучебная деятельность на оценивание своих способностей в рамках специальности. Стоит предположить, что студенты, участвующие не только в учебном процессе, но и в других различных мероприятиях, могут иметь более широкий кругозор, в зависимости от деятельности, и их активное участие в жизни университета и за его пределами придают ощущение уверенности в своих способностях. Стоит отметить, что

Средний балл результатов анкетирования

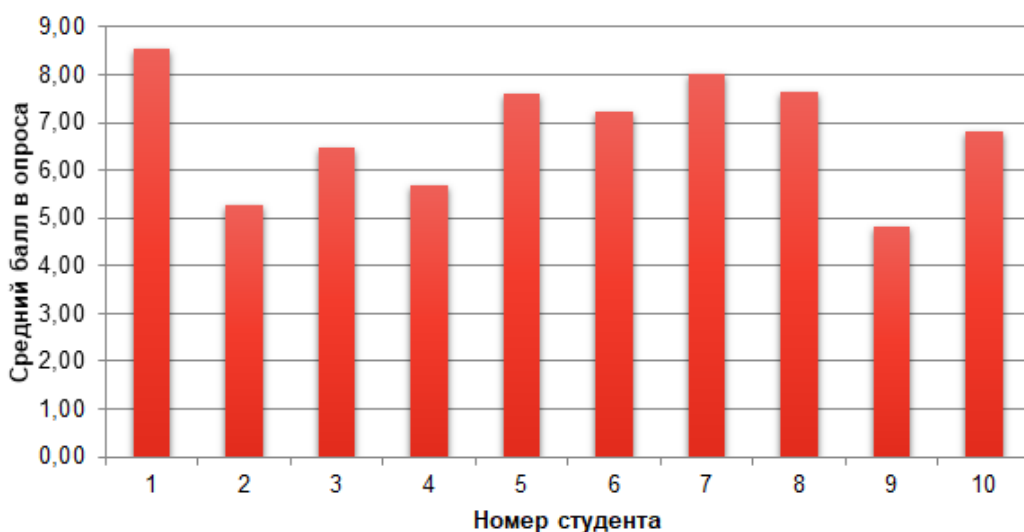


Рисунок 29 – Гистограмма средних значений опроса

По гистограмме на рисунке 29 можно сделать вывод о том, что при отсутствии какого-либо анализа или классификации данных результатов невозможно выявить какую-либо закономерность или разделение по группам. Отобразим результаты кластерного анализа по спискам студентов, которые сформировались в таблицах 12 и 14, в гистограммах на рисунках 30-31.

Средний балл результатов анкетирования

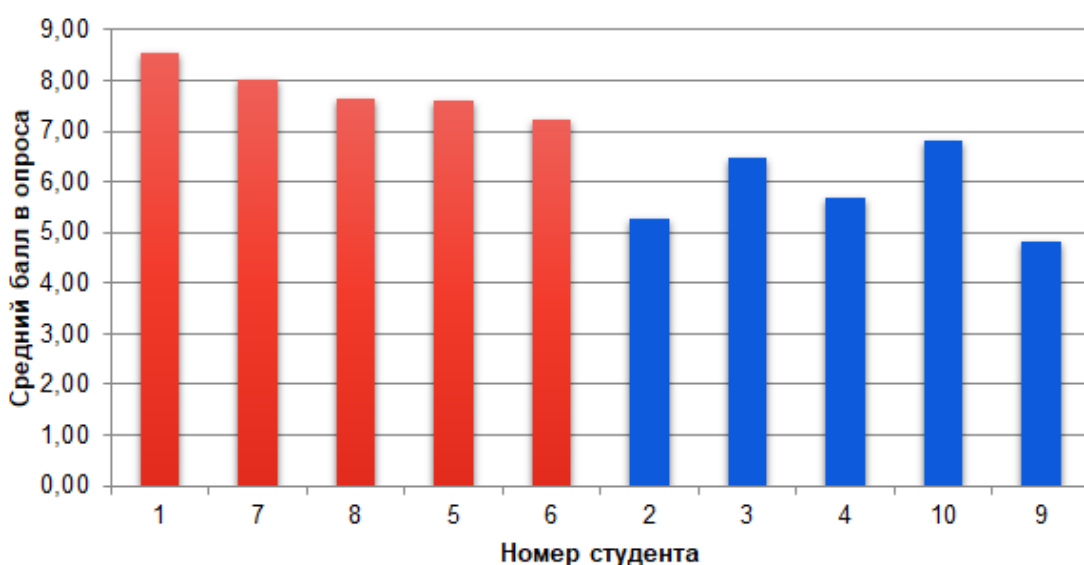


Рисунок 30 – Гистограмма средних значений опроса при кластеризации
1, 7, 8, 5, 6 – студенты первой группы; 2, 3, 4, 10, 9 – студенты второй группы

Изм	Лист	№ докум.	Подпись	Дата
-----	------	----------	---------	------

Средний балл результатов анкетирования

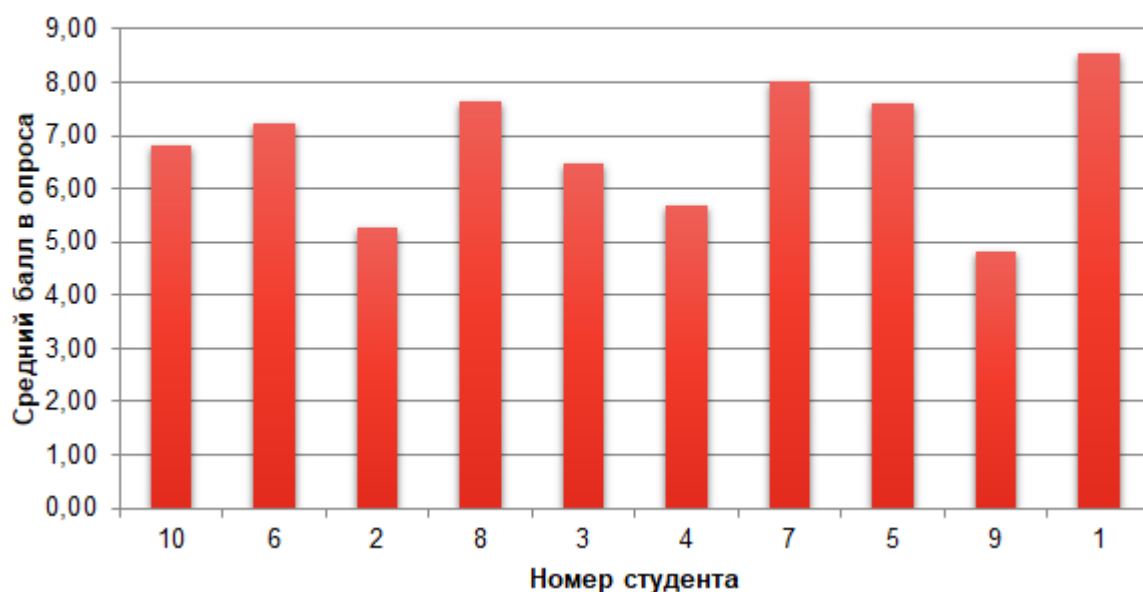


Рисунок 31 – Гистограмма средних значений опроса при сортировке по среднему баллу в дипломе

На рисунке 30 можно отметить более четкое разделение между выявленными двумя группами. На рисунке 31 такого четкого разделения не наблюдается, аналогично гистограмме на рисунке 29, то есть определить какую-либо закономерность затруднительно.

Таким образом, можно сделать вывод, что посредством кластерного анализа можно выделить иерархически структурированные группы. Это важно при обработке большого количества данных, когда из всего потока необходимо выделить наиболее однородные группы. Кластерный анализ может являться качественной основой для последующего анализа как в целом по группам, так и внутри самих групп.

Вывод по разделу 4: проведено тестирование работы приложения кластерного анализа на различных примерах, а именно на примере массива точек в двумерном пространстве, затем на примере специально разобранных по группам объектов в многомерном пространстве.

Результат кластерного анализа данных на примере массива точек в двумерном пространстве показал правильное разбиение, так как точки в образованных

четкое разделение, что говорит о том, что распределение более верное, нежели при сортировке по оценкам в дипломе.

Таким образом, тестирование приложения кластерного анализа показало его работоспособность, а также выявило качественность полученных результатов.

					<i>ЮУрГУ – 12.03.01.2020.272 ВКР</i>	<i>Лист</i>
<i>Изм</i>	<i>Лист</i>	<i>№ докум.</i>	<i>Подпись</i>	<i>Дата</i>		61

ЗАКЛЮЧЕНИЕ

В ходе выпускной квалификационной работы были выполнены следующие задачи.

Был проведен обзор методов кластерного анализа, мер сходства между кластерами и методов объединения схожих кластеров. В результате был выбран иерархический агломеративный метод кластерного анализа. Мерой сходства между кластерами выбран квадрат евклидова расстояния, а методом объединения схожих кластеров выбран метод Уорда.

Также была рассмотрена рекурсивная процедура иерархического последовательного кластерного анализа Ланса-Уильямса, и на основе этого был разработан алгоритм работы приложения кластерного анализа, основанный на рассмотренном алгоритме последовательной кластеризации.

Также был выбран вариант отображения результатов кластерного посредством построения дендрограммы с вертикальным расположением объектов.

Разработан листинг программы приложения, а также описаны основные моменты кода. Разработан диалоговый интерфейс данного приложения, соответствующий требованиям технического задания.

Проведено тестирование работы приложения кластерного анализа на примере массива точек в двумерном пространстве, затем на примере специально разобранных по группам объектов в многомерном пространстве. Результат кластерного анализа данных на примере массива точек в двумерном пространстве показал правильное разбиение, так как точки в образованных кластерах были близки друг к другу на каждом шаге объединения, а сами кластеры не пересекались друг с другом.

Также был проведен кластерный анализ по образованной выборке посредством анкетирования студентов-выпускников бакалавриата по вопросам освоения ими компетенций направления 12.03.01 «Приборостроение», на котором они обучаются. Его результат был проанализирован на основе дополнительных данных.

									Лист
									62
Изм	Лист	№ докум.	Подпись	Дата	ЮУрГУ – 12.03.01.2020.272 ВКР				

На основе сравнения результатов кластерного анализа и среднего балла по результатам опроса каждого студента, сделан вывод о том, что кластерный анализ учитывает не только средний результат, но и результаты ответов на вопросы.

Было выяснено, что студенты группы, средний балл анкетирования которых больше, чем у студентов другой группы, в основном получают менее высокий средний балл в дипломе. Также большинство студентов группы, результат опроса которых ниже, чем у другой группы, имели опыт с реальным производством в рамках своего направления, что приводит к мнению, что данные студенты более объективны в своих способностях.

Проведено сравнение результатов распределения студентов по результатам кластерного анализа и при сортировке их по ожидаемому среднему баллу в дипломе. Выяснено, что на гистограмме среднего балла результатов анкетирования при сортировке студентов кластерным анализом видно более четкое разделение, что говорит о том, что распределение более верное.

Тестирование приложения кластерного анализа показала его работоспособность, а также выявило качество полученных результатов.

Таким образом, цель выпускной квалификационной работы достигнута.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. ГОСТ Р ИСО 9241-110-2016 Эргономика взаимодействия человек-система. Часть 110. Принципы организации диалога - М.: Стандартинформ, 2016. - 28 с.
2. Trion, R. G. Cluster analysis / R. G. Trion — L.: Ann Arbor Edwards Bros., 1939. - 139 p.
3. Мандель, И. Д. Кластерный анализ / И.Д. Мандель — М.: Финансы и статистика.1988. — 176 с.
4. Steinhaus, H. Sur la division des corps materiels en parties / H. Steinhaus // Bull. Acad. Polon. Sci., C1. III , 1956. - vol IV. - p. 801 - 804.
5. Метод k-средних · Loginom Wiki - <https://wiki.loginom.ru/articles/k-means.html>.
6. Большой Российский энциклопедический словарь. - Репр. изд. - Москва: Большая Российская энцикл., 2009 - 1887 с.
7. НОУ ИНТУИТ | Лекция | Методы кластерного анализа. Иерархические методы - <https://www.intuit.ru/studies/courses/6/6/lecture/182?page=2>.
8. Williams, W. T. Multivariate Methods in Plant Ecology / W. T. Williams, J. M. Lambert, G. N. Lance // J. Ecology, 1966. - Vol. 54 - p. 427.
9. MacNaughton-Smith, P. Some statistical and other numerical techniques for classifying individuals / P. MacNaughton-Smith - London: H. M. S. O, 1965.
10. Sorenson T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application for analysis of the vegetation on Danish commons / T. Sorenson // Biol. Skr. 5, 1968. - p. 1 - 34.
11. Sokal, R. R. A statistical method for evaluating systematic relationships / R. R. Sokal, C. D. Michener // University of Kansas Sci. Bull Vol. 38 - Lawrence: University of Kansas, 1958. - p. 1409 - 1438.
12. Sneath, P.H. Numerical taxonomy: the principles and practice of numerical classification / P.H. Sneath, R.R. Sokal // San Francisco: Freeman, 1973. - 573 p.

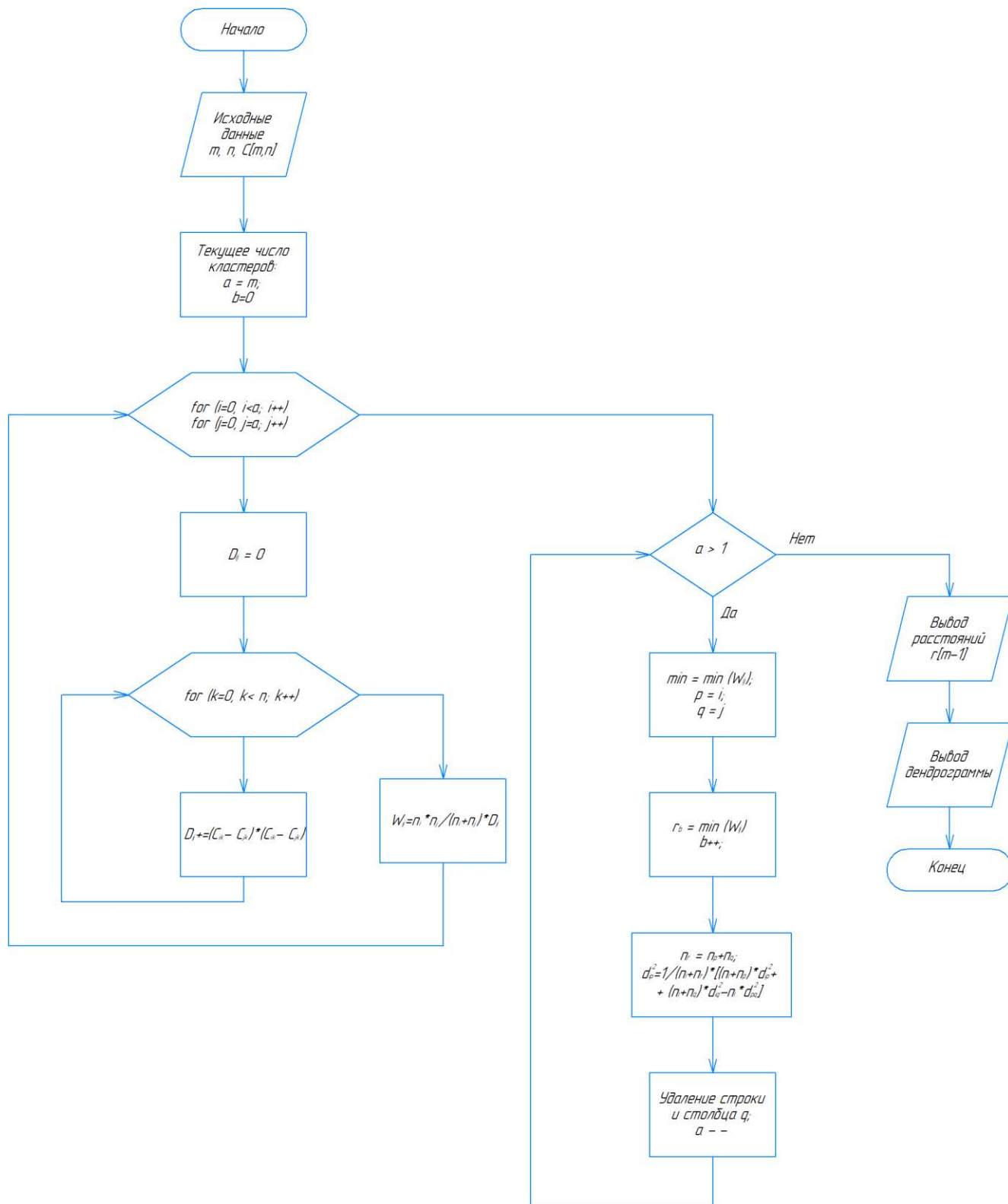
Изм	Лист	№ докум.	Подпись	Дата

24. Об утверждении федерального государственного образовательного стандарта высшего образования по направлению подготовки 12.03.01 Приборостроение (уровень бакалавриата): Приказ Министерства образования и науки РФ от 3 сентября 2015 г. N 959 - М., 2015.

25. Как вы оцениваете овладение ключевыми компетенциями по своей специальности за время обучения в университете? - <https://docs.google.com/forms/d/e/1FAIpQLSfjQ9e7cMYtNOUFNWxUEQnCKtg2gkUjYUvu0PE5MEGv6-4kzw/viewform>.

ПРИЛОЖЕНИЕ А

Алгоритм работы приложения кластерного анализа



ПРИЛОЖЕНИЕ Б

Листинг программы приложения кластерного анализа

```
////////////////////////////////////
// Cluster analysis App v. 1.0           //
// Made by student group KE-414 Alexandrov Dmitry //
////////////////////////////////////

// использование необходимых объектов
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;

namespace Cluster_analysis_App_v._1._0
{
    public partial class Form1 : Form
    {
        public Form1()
        {
            InitializeComponent();
        }

        private void Form1_Load(object sender, EventArgs e)
        {
        }

        private void btnClusteranalysis_Click(object sender, EventArgs e)
        {
            Cluster();
        }

        private void Cluster()
        {
            chart1.Series.Clear(); // очистка графика
            richTextBox1.Clear(); // очистка текстового поля
            if (openFileDialog1.ShowDialog() == DialogResult.OK) // загрузка файла
            {
                // ОБОЗНАЧЕНИЕ ПЕРЕМЕННЫХ
                int i, j, k;
                int b = 0, c = 1, t1 = 0, t2 = 0;
                double s, nr;
                string[] lines = File.ReadAllLines(openFileDialog1.FileName); // количество
                строк в файле
                double[,] C = new double[lines.Length, lines[0].Split('\t').Length]; //
                матрица исходных данных (строки - объекты, столбцы - признаки)
                double[] n = new double[lines.Length]; // массив, в который записывается
                количество исходных кластеров, содержащихся в текущих кластерах
                double[,] D = new double[lines.Length, lines.Length]; // матрица расстояний
                double[,] W = new double[lines.Length, lines.Length]; // матрица
                внутригрупповых сумм квадратов (ВСК) - целевая функция
            }
        }
    }
}
```

Изм	Лист	№ докум.	Подпись	Дата
-----	------	----------	---------	------

```

double[,] A = new double[lines.Length, lines.Length]; // матрица, используемая
для преобразования матрицы расстояний на каждом шаге
// для дендрограммы
double[,] S = new double[lines.Length, lines[0].Split('\t').Length]; //
матрица исходных данных, отсортированных для исключения пересечения линий на графике
double[] d = new double[lines.Length]; // массив, в который записывается
расстояние для вновь образуемых кластеров
double[] z = new double[lines.Length]; // массив, в который записываются
координаты кластеров на графике по оси X
// Запись исходных данных из файла
for (i = 0; i < lines.Length; i++)
{
    n[i] = 1;
    d[i] = 0;
    string[] temp = lines[i].Split('\t');
    for (j = 0; j < temp.Length; j++)
        C[i, j] = Convert.ToDouble(temp[j]);
}
// Дополнительные обозначения переменных
int m = C.GetLength(0); // количество строк в двумерном массиве C[m x p]
int p = C.GetLength(1); // количество столбцов в двумерном массиве C[m x p]
int a = m; // переменная, обозначающая размерность матрицы расстояний на
каждом шаге
double[] r = new double[a - 1]; // массив расстояний, на которых будут
объединяться пара кластеров
double[,] x1 = new double[a - 1, m]; // матрица, преобразуемая в массив
double[] t = new double[a];
int[,] y = new int[a, m];
int[] y1 = new int[m];
double[] z1 = new double[a];

// вывод матрицы C[m x p] в таблицу
dataGridView1.RowCount = m + 1;
dataGridView1.ColumnCount = p + 1;
dataGridView1.Rows[0].Cells[0].Value = "06. | пр.";
dataGridView1.Rows[0].Cells[0].Style.BackColor = Color.FromArgb(255, 255,
120);

for (i = 0; i < m + 1; i++)
    if (i != 0)
        for (j = 0; j < p + 1; j++)
            if (j == 0)
            {
                dataGridView1.Rows[i].Cells[j].Value = i;
                dataGridView1.Rows[i].Cells[j].Style.BackColor =
Color.FromArgb(255, 255, 120);
            }
            else
                dataGridView1.Rows[i].Cells[j].Value = C[i - 1, j - 1];
        else
            for (j = 1; j < p + 1; j++)
            {
                dataGridView1.Rows[i].Cells[j].Value = j;
                dataGridView1.Rows[i].Cells[j].Style.BackColor =
Color.FromArgb(255, 255, 120);
            }

// Массив номеров элементов
for (i = 0; i < m; i++)
    y[i, 0] = i + 1;

```

```

// процедура кластерного анализа //

// Формирование матрицы квадратов евклидовых расстояний
for (i = 0; i < a; i++)
{
    for (j = 0; j < a; j++)
    {
        s = 0;
        for (k = 0; k < p; k++)
            s += (C[i,k] - C[j,k]) * (C[i,k] - C[j,k]); // Квадрат
евклидового расстояния
        D[i, j] = s;
        W[i, j] = n[i] * n[j] / (n[i] + n[j]) * D[i, j]; // Внутригрупповая
сумма квадратов (ВСК)
    }
}
// Цикл последующих этапов, который будет повторяться до тех пор, пока все
отдельные кластеры не объединятся в один
while (a > 1)
{
    // Объединение кластеров, имеющих минимальное расстояние
    // определение минимального расстояния
    double min;
    for (i = 0; i < a; i++)
    {
        for (j = 0; j < a; j++)
        {
            if (D[i, j] < min & i != j)
            {
                min = D[i, j];
                t1 = i;
                t2 = j;
            }
        }
    }
    r[b] = Math.Round(min, 2, MidpointRounding.AwayFromZero); //Запись
расстояния, на котором объединяются кластеры
    // запись в одну строку объединенных кластеров, массив номеров
    k = 0;
    for (i = 0; i < a; i++)
    {
        if (i == t1)
        {
            for (j = 0; j < m; j++)
            {
                if (y[i, j] == 0)
                {
                    y[i, j] = y[t2, k];
                    k++;
                }
                else
                {
                    y[i, j] = y[i, j];
                }
            }
        }
        else if (i == t2)
        {

```

```

        for (k = t2; k < a - 1; k++)
            for (j = 0; j < m; j++)
                y[k, j] = y[k + 1, j];
        i = k;
    }
    else
    {
        for (j = 0; j < m; j++)
            y[i, j] = y[i, j];
    }
}
for (j = 0; j < m; j++)
    y[a - 1, j] = 0;

for (i = 0; i < m; i++)
    if (y[t1, i] != 0)
        x1[b, i] = y[t1, i];
b++;
nr = n[t1] + n[t2]; // сумма исходных кластеров в объединенном кластере
// Перерасчет матрицы евклидовых расстояний
for (i = 0; i < a; i++)
{
    if (i == t1)
    {
        for (j = 0; j < a; j++)
        {
            if (j == t1)
            {
                A[i, j] = D[i, j];
            }
            else if (j == t2)
            {
                A[i, j] = 0;
            }
            else
                A[i, j] = 1 / (n[j] + nr) * ((n[j] + n[t1]) * D[t1, j] +
(n[j] + n[t2]) * D[t2, j] - n[j] * D[t1, t2]);
        }
    }
    else if (i == t2)
    {
        for (j = 0; j < a; j++)
            A[i, j] = 0;
    }
    else
        for (j = 0; j < a; j++)
        {
            if (j == t1)
            {
                A[i, j] = 1 / (n[i] + nr) * ((n[i] + n[t1]) * D[i, t1] +
(n[i] + n[t2]) * D[i, t2] - n[i] * D[t1, t2]);
            }
            else if (j == t2)
            {
                A[i, j] = 0;
            }
            else
                A[i, j] = D[i, j];
        }
    }

    for (i = 0; i < a; i++)

```

Изм	Лист	№ докум.	Подпись	Дата
-----	------	----------	---------	------

```

{
  if (i == t2)
  {
    for (k = t2; k < a - 1; k++)
      for (j = 0; j < a; j++)
        A[k, j] = A[k + 1, j];
    i = k;
  }
  else if (i == a - 1)
  {
    for (j = 0; j < a; j++)
      A[i, j] = 0;
  }
  else
    for (j = 0; j < a; j++)
      A[i, j] = A[i, j];
}
for (i = 0; i < a; i++)
{
  for (j = 0; j < a; j++)
  {
    if (j == t2)
    {
      for (k = t2; k < a - 1; k++)
        A[i, k] = A[i, k + 1];
      j = k;
    }
    else if (j == a - 1)
      A[i, j] = 0;
    else
      A[i, j] = D[i, j];
  }
}
a--;
// для убедительности, что главная диагональ вспомогательной матрицы
расстояний нулевая
for (i = 0; i < a; i++)
  for (j = 0; j < a; j++)
    if (i == j)
      A[i, j] = 0;
n[t2] = n[t1] + n[t2]; // количество исходных кластеров в новом
// удаление n[t2]
for (i = t2; i < a; i++)
{
  if (i == a - 1)
    n[i] = 0;
  n[i] = n[i + 1];
}
// перезаписывание матрицы расстояний и матрицы ВСК
for (i = 0; i < a; i++)
  for (j = 0; j < a; j++)
  {
    D[i, j] = A[i, j];
    W[i, j] = D[i, j] / 2;
  }
c++; // переход на следующий номер объединения
}

// вывод результата текущего объединения
for (i = 0; i < m - 1; i++)
{

```

```

richTextBox1.Text += Convert.ToString(i + 1) + "-е объединение кластеров:
";
for (j = 0; j < m; j++)
{
    if (x1[i, j] != 0)
        richTextBox1.Text += Convert.ToString(x1[i, j]) + " ";
    else
        break;
}
richTextBox1.Text += "на расстоянии " +
Convert.ToString(string.Format("{0:0.00}", r[i])) + "\n\n";
}
// кластерный анализ завершен

////////////////////////////////////

// построение дендрограммы
chart1.ChartAreas[0].AxisY.Maximum = r[m - 2]; //максимальное значение на оси Y
chart1.ChartAreas[0].AxisY.ScaleView.Zoom(0, (Math.Round(r[m - 2]/2, 2,
MidpointRounding.AwayFromZero)));
chart1.ChartAreas[0].AxisY.Title = "Расстояние";
chart1.ChartAreas[0].AxisX.Title = "Объекты";
chart1.ChartAreas[0].AxisY.Minimum = 0; // минимальное значение на оси Y
chart1.ChartAreas[0].AxisY.Interval = (Math.Round(r[m - 2]/20, 2,
MidpointRounding.AwayFromZero)); // интервал на оси Y
chart1.ChartAreas[0].AxisX.Interval = 0.5; // интервал на оси X
chart1.ChartAreas[0].AxisX.Minimum = 0; // минимальное значение на оси X
chart1.ChartAreas[0].AxisX.Maximum = m+1; // максимальное значение на оси X

c = 0;
for (i = 0; i < m; i++)
{
    n[i] = 1;
}
// переписываем номера исходных кластеров в массив, который будет
использоваться для подписи оси X
// в массиве у расположены номера кластеров в том порядке, при котором на
дендрограмме исключаются пересечения линий
for (i = 0; i < m; i++)
{
    y1[i] = y[0, i];
}
// переписываем матрицу
for (i = 0; i < m; i++)
{
    for (j = 0; j < p; j++)
    {
        S[i, j] = C[y1[i] - 1, j];
    }
}
a = m;
//Матрица номеров элементов у и номеров объектов z
for (i = 0; i < m; i++)
{
    z[i] = i + 1;
    for (j = 0; j < m; j++)
    if (j == 0)
        y[i, j] = i + 1;
    else
        y[i, j] = 0;
}

```

```

// процедура кластерного анализа

// Формирование матрицы квадратов евклидовых расстояний
for (i = 0; i < a; i++)
{
    for (j = 0; j < a; j++)
    {
        s = 0;
        for (k = 0; k < p; k++)
            s += (S[i, k] - S[j, k]) * (S[i, k] - S[j, k]); // Квадрат
евклидова расстояния
        D[i, j] = s;
        W[i, j] = n[i] * n[j] / (n[i] + n[j]) * D[i, j]; // Внутригрупповая
сумма квадратов (ВСК)
    }
}
b = 0;
// Цикл
while (a > 1)
{
    // нахождение минимального расстояния и номеров объединяемых объектов

    for (i = 0; i < a; i++)
    {
        for (j = 0; j < a; j++)
        {
            if (W[i, j] < min & i != j)
            {
                min = W[i, j];
                t1 = i;
                t2 = j;
            }
        }
    }
    // перезаписываем матрицу номеров
    k = 0;
    for (i = 0; i < a; i++)
    {
        if (i == t1)
        {
            for (j = 0; j < m; j++)
            {
                if (y[i, j] == 0)
                {
                    y[i, j] = y[t2, k];
                    k++;
                }
                else
                {
                    y[i, j] = y[i, j];
                }
            }
        }
        else if (i == t2)
        {
            for (k = t2; k < a - 1; k++)
                for (j = 0; j < m; j++)
                    y[k, j] = y[k + 1, j];
            i = k;
        }
    }
}

```

```

else
{
    for (j = 0; j < m; j++)
        y[i, j] = y[i, j];
}
}
for (j = 0; j < m; j++)
    y[a - 1, j] = 0;
r[b] = min;
nr = n[t1] + n[t2];
// Перерасчет матрицы евклидовых расстояний
for (i = 0; i < a; i++)
{
    if (i == t1)
    {
        for (j = 0; j < a; j++)
        {
            if (j == t1)
            {
                A[i, j] = D[i, j];
            }
            else if (j == t2)
            {
                A[i, j] = 0;
            }
            else
                A[i, j] = 1 / (n[j] + nr) * ((n[j] + n[t1]) * D[t1, j] +
(n[j] + n[t2]) * D[t2, j] - n[j] * D[t1, t2]);
        }
    }
    else if (i == t2)
    {
        for (j = 0; j < a; j++)
            A[i, j] = 0;
    }
    else
        for (j = 0; j < a; j++)
        {
            if (j == t1)
            {
                A[i, j] = 1 / (n[i] + nr) * ((n[i] + n[t1]) * D[i, t1] +
(n[i] + n[t2]) * D[i, t2] - n[i] * D[t1, t2]);
            }
            else if (j == t2)
            {
                A[i, j] = 0;
            }
            else
                A[i, j] = D[i, j];
        }
    }
for (i = 0; i < a; i++)
{
    if (i == t2)
    {
        for (k = t2; k < a - 1; k++)
            for (j = 0; j < a; j++)
                A[k, j] = A[k + 1, j];
        i = k;
    }
    else if (i == a - 1)
    {

```



```

        for (j = 0; j < a; j++)
            A[i, j] = 0;
    }
    else
        for (j = 0; j < a; j++)
            A[i, j] = A[i, j];
}
for (i = 0; i < a; i++)
{
    for (j = 0; j < a; j++)
    {
        if (j == t2)
        {
            for (k = t2; k < a - 1; k++)
                A[i, k] = A[i, k + 1];
            j = k;
        }
        else if (j == a - 1)
            A[i, j] = 0;
        else
            A[i, j] = A[i, j];
    }
}
// записываем в массив z1 массив z
for (i = 0; i < a; i++)
{
    z1[i] = z[i];
}
// формирование дендрограммы линиями
string q = Convert.ToString("union" + (c + 1) + ".1");
chart1.Series.Add(new Series(q));
string q2 = Convert.ToString("union" + (c - 1) + ".1-5");
chart1.Series.Add(new Series(q2));
chart1.Series[q2].Points.AddXY(z[t1], r[b]);

chart1.Series[q].Points.AddXY(z[t1], d[t1]);
chart1.Series[q].Points.AddXY(z[t1], r[b]);
n[t1] = n[t1] + n[t2];
d[t1] = r[b];
z[t1] = (z[t1] + z[t2]) / 2;

chart1.Series[q2].AxisLabel = " ";
chart1.Series[q2].Points.AddXY(z[t1], r[b]);
string q1 = Convert.ToString("union" + (c + 1) + ".2");
chart1.Series.Add(new Series(q1));

chart1.Series[q1].Points.AddXY(z[t1], r[b]);
chart1.Series[q1].Points.AddXY(z[t2], r[b]);
chart1.Series[q1].Points.AddXY(z[t2], d[t2]);
b++;
// удаление из массивов n,z,d,y1 элемента связанного с t2
for (i = t2; i < a; i++)
{
    if (i == a - 1)
    {
        n[i] = 0;
        z[i] = 0;
        d[i] = 0;
        y1[i] = 0;
    }
}

```

```

        else
        {
            n[i] = n[i + 1];
            z[i] = z[i + 1];
            d[i] = d[i + 1];
            y1[i] = y1[i + 1];
        }
    }
    a--;
    for (i = 0; i < a; i++)
        for (j = 0; j < a; j++)
            if (i == j)
                A[i, j] = 0;
    // переписываем матрицу расстояний и матрицу ВСК
    for (i = 0; i < a; i++)
        for (j = 0; j < a; j++)
        {
            D[i, j] = A[i, j];
            W[i, j] = D[i, j] / 2;
        }
    c++;
}
}

private void chart1_Click(object sender, EventArgs e)
{
}

private void richTextBox1_TextChanged(object sender, EventArgs e)
{
}

private void listView1_SelectedIndexChanged(object sender, EventArgs e)
{
}

private void dataGridView1_CellContentClick(object sender, DataGridViewCellEventArgs
e)
{
}
}
}

```

ПРИЛОЖЕНИЕ В

Исходные данные для тестирования

Объекты	Признаки														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	2	3	2	2	1	1	3	1	1	3	2	1	1	3	1
2	3	2	3	2	1	1	1	3	3	1	1	1	2	3	1
3	4	4	4	5	6	5	6	5	4	4	4	5	6	4	4
4	6	6	4	5	4	4	6	5	5	6	5	4	6	6	6
5	7	7	8	7	7	7	8	8	8	8	7	8	7	8	8
6	8	7	8	7	7	7	7	8	7	7	8	7	7	8	7
7	10	10	9	9	10	9	9	10	10	10	9	9	9	9	9
8	10	10	10	9	10	10	9	10	10	9	9	9	10	9	9
9	5	5	2	1	4	9	6	5	5	5	9	1	5	7	8
10	6	4	1	3	2	5	1	10	6	8	4	5	2	2	10
Объекты	Признаки														
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	3	3	1	2	1	1	3	1	3	1	3	1	2	3
2	3	1	1	3	2	3	3	1	1	3	3	3	1	3	1
3	6	5	4	5	5	6	4	5	4	5	6	4	4	6	5
4	5	4	5	5	6	5	6	6	6	4	5	5	4	6	6
5	8	8	7	7	8	8	7	8	7	8	7	8	7	7	8
6	8	8	7	8	8	7	8	7	8	8	7	7	7	8	7
7	9	9	9	10	9	9	10	10	9	10	9	9	9	9	9
8	9	10	9	10	9	10	10	9	10	9	10	10	9	9	10
9	1	3	1	9	7	6	1	2	6	8	4	7	5	3	8
10	9	6	10	6	4	3	1	8	8	9	10	1	6	4	8
Объекты	Признаки														
	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1	2	1	2	1	2	2	3	1	2	2	1	1	3	2	3
2	3	2	1	3	1	3	3	2	3	3	1	2	2	3	1
3	5	4	4	6	4	4	5	4	6	4	5	4	5	4	4
4	4	5	6	4	4	6	4	4	5	4	6	6	5	6	4
5	8	7	7	7	8	8	8	7	8	8	7	7	8	7	7
6	8	7	7	8	7	8	8	8	8	7	8	8	7	8	7
7	9	10	9	10	9	9	9	9	10	9	10	10	10	10	10
8	9	9	10	9	9	9	9	9	10	10	9	9	10	10	9
9	4	8	10	8	5	2	10	2	10	4	7	3	7	3	9
10	6	3	6	9	5	9	10	8	4	9	1	1	5	4	4

ПРИЛОЖЕНИЕ Г

Перечень вопросов анкетирования студентов

Для создания выборки необходимо ответить на 42 вопроса. Диапазон оценки для каждого ответа от 0 до 10, шаг – один балл.

1. Способность формировать мировоззренческую позицию на основе философских знаний
2. Способность анализировать основные этапы и закономерности исторического развития общества для формирования гражданской позиции
3. Способность использовать основы экономических знаний в различных сферах деятельности
4. Способность использовать основы правовых знаний в различных сферах деятельности
5. Способность к коммуникации в устной и письменной формах на русском и иностранном языках для решения задач межличностного и межкультурного взаимодействия
6. Способность работать в команде, толерантно воспринимая социальные и культурные различия
7. Способность к самоорганизации и самообразованию
8. Способность использовать методы и средства физической культуры для обеспечения полноценной социальной и профессиональной деятельности
9. Способность использовать приемы первой помощи, методы защиты в условиях чрезвычайных ситуаций
10. Способность представлять адекватную современному уровню знаний научную картину мира на основе знания основных положений, законов и методов естественных наук и математики
11. Способность осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий
12. Способность выявлять естественно-научную сущность проблем, возникающих в ходе профессиональной деятельности, привлекать для их решения физико-математический аппарат
13. Способность учитывать современные тенденции развития техники и технологий в своей профессиональной деятельности
14. Способность обрабатывать и представлять данные экспериментальных исследований
15. Способность собирать, обрабатывать, анализировать и систематизировать научно-техническую информацию по тематике исследования
16. Способность использовать современные программные средства подготовки конструкторско-технологической документации
17. Способность использовать нормативные документы в своей деятельности

Изм	Лист	№ докум.	Подпись	Дата

32. Способность к разработке планов конструкторско-технологических работ и контролю их выполнения, включая обеспечение соответствующих служб необходимой технической документацией, материалами, оборудованием

33. Способность разрабатывать оптимальные решения при создании продукции приборостроения с учетом требований качества, стоимости, сроков исполнения, конкурентоспособности и безопасности жизнедеятельности, а также экологической безопасности

34. Способность устанавливать порядок выполнения работ и организацию маршрутов технологического прохождения элементов и узлов приборов и систем в процессе их изготовления

35. Способность к размещению технологического оборудования, технического оснащения и организации рабочих мест, расчету производственных мощностей и загрузки оборудования по действующим методикам и нормативам

36. Способность к организации технического контроля и участию в управлении качеством производства продукции приборостроения, включая внедрение систем менеджмента качества

37. Способность к контролю соответствия разрабатываемых проектов и технической документации стандартам, техническим условиям и другим нормативным документам

38. Способность владеть правилами и методами монтажа, настройки и регулировки узлов приборов и систем, в том числе связанных с включением человека-оператора в контур управления приборами

39. Способность проводить поверку, наладку и регулировку оборудования, настройку программных средств, используемых для разработки, производства и настройки приборной техники

40. Готовность к практическому применению основных правил выполнения ремонта и обслуживания приборов, основ технологии обслуживания приборной техники

41. Способность владеть средствами эксплуатации приборных баз данных, экспертных и мониторинговых систем

42. Готовность составлять заявки на запасные детали и расходные материалы, а также на поверку и калибровку аппаратуры

Форма обучения

- Очная
- Заочная
- Очно-заочная
- Дистанционная

Курс

— 1 (бакалавриат)

— 2 (бакалавриат)

— 3

— 4

— 5

— 6

— 1 (магистратура)

— 2 (магистратура)

Пол

— Мужской

— Женский

—

Ожидаемый средний балл в дипломе (например, 4,5)

Как Вы оцениваете свою посещаемость занятий?

— Почти всегда

— Часто

— 50/50

— Редко

— Почти никогда

Работали/работаете во время учебы?

— Работаю, по специальности

— Работаю, не по специальности

— Работал, по специальности

— Работал, не по специальности

— Не работал

Занимаетесь ли Вы внеучебной деятельностью?

— Да

— Нет

Изм	Лист	№ докум.	Подпись	Дата

ПРИЛОЖЕНИЕ Д

Выборка для проведения кластерного анализа

№	1. Способн	2. Способн	3. Способн	4. Способн	5. Способн	6. Способн	7. Способн	8. Способн	9. Способн	10. Способн	11. Способн	12. Способн	13. Способн	14. Способн
1	7	8	6	3	9	7	10	10	10	8	10	8	10	9
2	7	2	7	1	5	10	5	4	5	5	6	9	5	10
3	5	5	3	1	4	5	9	9	3	7	7	9	7	8
4	2	2	2	1	3	5	7	3	5	5	6	5	6	6
5	5	6	6	6	8	8	10	6	6	8	10	7	8	8
6	8	8	8	7	7	9	7	9	9	9	9	9	7	8
7	8	8	8	6	7	10	7	9	8	9	8	7	9	8
8	8	7	6	7	6	8	8	8	8	6	8	6	6	9
9	8	6	6	6	6	6	5	6	7	6	7	4	5	8
10	3	10	8	0	7	10	7	10	7	7	4	7	9	10

№	15. Способн	16. Способн	17. Способн	18. Способн	19. Готовно	20. Способн	21. Готовно	22. Способн	23. Способн	24. Способн	25. Способн	26. Готовно	27. Способн	28. Способн
1	10	9	10	10	8	9	7	9	10	9	8	10	10	8
2	5	4	6	3	5	9	10	8	3	8	8	4	7	7
3	8	8	8	8	6	7	9	9	6	7	7	7	8	8
4	6	6	8	7	4	7	7	7	6	7	6	7	6	7
5	8	9	10	4	4	8	6	7	7	10	8	7	9	10
6	8	9	8	9	9	9	9	9	6	9	6	5	8	8
7	9	8	8	10	9	5	7	7	7	7	7	9	8	7
8	10	5	9	9	9	9	8	8	6	6	6	9	9	9
9	5	7	9	6	8	3	3	9	2	3	3	3	3	2
10	3	7	10	10	7	7	7	8	8	7	4	10	7	7

№	29. Готовно	30. Способн	31. Готовно	32. Способн	33. Способн	34. Способн	35. Способн	36. Способн	37. Способн	38. Способн	39. Способн	40. Готовно	41. Способн	42. Готовно
1	5	6	10	8	9	9	9	9	10	10	10	10	10	10
2	1	5	5	4	7	7	7	2	5	7	1	6	0	0
3	5	5	7	7	9	7	6	7	7	6	5	5	4	8
4	6	6	6	7	5	6	7	6	8	7	7	7	6	7
5	8	10	9	10	9	7	8	8	9	6	7	9	8	5
6	4	5	4	5	6	7	7	4	8	6	6	4	8	5
7	9	9	8	8	9	8	8	7	9	7	9	8	9	9
8	7	7	9	8	8	8	8	8	7	8	8	7	7	8
9	1	2	4	5	2	1	5	6	4	5	6	1	3	1
10	10	7	7	2	3	4	1	4	7	7	8	10	5	7

Изм	Лист	№ докум.	Подпись	Дата
-----	------	----------	---------	------