

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий

(наименование института полностью)

Кафедра «Прикладная математика и информатика»

(наименование кафедры)

01.03.02 Прикладная математика и информатика

(код и наименование направления подготовки, специальности)

Системное программирование и компьютерные технологии

(направленность (профиль)/специализация)

## БАКАЛАВРСКАЯ РАБОТА

на тему: «Разработка алгоритма диагностики сварки на основе кластерного анализа данных»

Студент

М.М. Аврангшоев

(И.О. Фамилия)

(личная подпись)

Руководитель

В.С. Климов

(И.О. Фамилия)

(личная подпись)

Консультанты

М.А. Четаева

(И.О. Фамилия)

(личная подпись)

**Допустить к защите**

Заведующий кафедрой к.т.н., доцент, А.В. Очеповский

(ученая степень, звание, И.О. Фамилия)

(личная подпись)

« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_\_ г.

Тольятти 2018

## АННОТАЦИЯ

Тема бакалаврской работы: «Разработка алгоритма диагностики сварки на основе кластерного анализа данных».

В данной бакалаврской работе исследуются способы применения алгоритмов машинного обучения в задаче диагностики контактной сварки.

В работе предложен алгоритм построения классификатора на основе результатов кластерного анализа данных. При этом для кластерного анализа данных используется алгоритм k-means. С помощью предложенного алгоритма построен классификатор качества контактной сварки и исследована точность его работы на реальных данных. Разработано программное обеспечение, реализующее предложенные алгоритмы.

При проведении исследований, представленных в бакалаврской работе, было опубликовано 2 статьи.

Структура бакалаврской работы представлена введением, тремя главами, заключением, списком литературы.

Во введении описывается актуальность проводимого исследования, дается краткая характеристика проделанной работы.

В первой главе проводится обзор методов интеллектуального анализа, применяемых для прогнозирования качества сварных соединений, формулируется цель и ставятся задачи.

Во второй главе описывается математический аппарат алгоритма построения классификатора на основе результатов кластерного анализа данных.

В третьей главе приведены примеры использования предложенных подходов и описано разработанное программное обеспечение.

В заключении представлены выводы по проделанной работе.

В работе использовано 4 таблицы, 15 рисунков, список литературы содержит 20 литературных источников. Общий объем выпускной квалификационной работы составляет 42 страниц.

## **ABSTRACT**

The topic of bachelor's work: "Development of a welding algorithm based on cluster analysis of data."

In this bachelor's work the methods of application of machine learning algorithms in the problem of diagnostics of contact welding are investigated.

The algorithm for constructing a classifier based on the results of cluster data analysis is proposed. In this case, the k-means algorithm is used for cluster data analysis. With the help of the proposed algorithm, a quality classifier for contact welding was constructed and the accuracy of its work on real data was investigated. The software that implements the proposed algorithms is developed..

During the research presented in the bachelor's work, 2 articles were published.

The structure of the bachelor's work is represented by an introduction, three chapters, a conclusion, a list of literature.

In the introduction, the urgency of the research is described, a brief description of the work done is given.

The first chapter provides an overview of the methods of intellectual analysis used to predict the quality of welded joints, formulates a goal and sets tasks.

The second chapter describes the mathematical apparatus of the algorithm for constructing a classifier based on the results of cluster analysis of data.

The third chapter gives examples of the use of the proposed approaches and describes the developed software.

In conclusion conclusions on the work done are presented.

The work uses 4 tables, 25 figures, the list of literature contains 20 literature sources. The total volume of the final qualifying work is 42 pages.

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .....	5
1 АНАЛИЗ СОСТОЯНИЯ ВОПРОСА .....	7
2 РАЗРАБОТКА АЛГОРИТМА КЛАССИФИКАЦИИ НА ОСНОВЕ РЕЗУЛЬТАТОВ КЛАСТЕРНОГО АНАЛИЗА .....	10
2.1 Кластерный анализ с использованием алгоритма k-means.....	10
2.2 Анализ данных процесса контактной сварки.....	17
2.3 Разработка алгоритма построения классификатора на основе результатов кластерного анализа и алгоритма классификации объектов.	20
3 ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ПРЕДЛОЖЕННЫХ РЕШЕНИЙ.....	25
3.1 Пример использования алгоритма построения классификатора.....	25
3.2 Пример классификации объекта .....	30
3.2 Программная реализация .....	31
ЗАКЛЮЧЕНИЕ .....	38
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ .....	40

## ВВЕДЕНИЕ

В настоящее время алгоритмы машинного обучения широко применяются в системах интеллектуальной диагностики различных технологических процессов.

Интеллектуальные системы диагностики с точки зрения математики решают задачу классификации, когда по измеряемым параметрам протекания технологического процесса прогнозируется класс качества получаемой продукции.

Так, например, в интеллектуальных системах диагностики процесса контактной сварки такие технологии искусственного интеллекта, как нейронные сети различной конфигурации, деревья классификации и регрессии, различные индуктивные алгоритмы построения классификаторов.

Существуют такие алгоритмы машинного обучения, возможности которых не изучены в полной мере. Например, алгоритм кластеризации данных k-means.

Алгоритм k-means предназначен для автоматизированной группировки объектов по кластерам. При этом кластерная структура подбирается таким образом, чтобы объекты внутри одного кластера были максимально похожи друг на друга. Недостатком алгоритма k-means является, что он не предназначен для решения задач классификации.

Вопросами использования кластерного анализа при решении практических задач занимаются такие исследователи, как Attila Gursoy, Piotr Płoński, Krzysztof Zaremba, López-Escobar Saúl, J. A. Carrasco-Ochoa, Martínez-Trinidad J. Fco, Dong-Jun Xin, Yen-Wei Chen, Le Wang, Li Tian, Yan Jia, Weihong Han, Ling Li, Chunguang Li, Yongxiu Lai, Guoling Shi, Dezhong Yao, Juan Carlos, Rojas Thomas и др [2-18].

В ходе проведенных исследований предложена идея построения классификатора на основе алгоритма k-means. Идея заключается в том, чтобы сначала поделить объекты обучающей выборки на кластеры, и, изучив состав каждого кластера сформировать стратегию классификации данных. Изучение

состава кластеров необходимо для того, чтобы определить, как распределены метки классов по кластерам. Затем на основе изученных свойств кластеров предложено классифицировать новые (ранее неизвестные) объекты, оценивая их близость до центров кластеров.

В исследовании проверяется гипотеза о том, возможности построения классификатора данных на основе результатов кластерного анализа.

Целью работы является разработка алгоритма построения классификатора данных на основе результатов кластерного анализа для диагностики качества контактной сварки.

В ходе выполнения исследований была разработана программа, позволяющая, с использованием предложенных подходов, по данным экспериментальных сварок производить построения классификатора и производить с его помощью классификацию данных.

Предложенные подходы были протестированы на экспериментальных данных контактной сварки. Полученный в ходе исследований классификатор качества позволяет по измеряемым в процессе сварки параметрам оценивать класс качества сварных соединений с точность 81,36%.

В ходе выполнения бакалаврской работы было опубликовано две статьи на тему практического применения алгоритмов машинного обучения [19-20].

Основные результаты работы были доложены на IV Международной научно-практической конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук»

## 1 АНАЛИЗ СОСТОЯНИЯ ВОПРОСА

Технологии искусственного интеллекта получили широкое распространение в системах управления и диагностики, применяемых на производствах в различных отраслях.

Благодаря технологии индуктивного машинного обучения разработчикам интеллектуальных систем не требуется заниматься проектированием математических моделей физических явлений, лежащих в основе исследуемого технологического процесса.

Технология индуктивного машинного обучения основана на накоплении выборки экспериментальных данных при реализации технологического процесса и автоматизированный анализ данных с целью:

- восстановления зависимости между наблюдаемыми параметрами технологического процесса и получаемыми характеристиками продукции;
- отделение значимых с точки зрения диагностики параметров и малозначимых;
- автоматизированное формирование алгоритма управления и диагностики технологическим процессом.

Контактная сварка – наиболее распространённый технологический процесс автоматизированного получения неразъемных соединений листовых деталей. Так как процесс контактной сварки сложноформализуем с точки зрения математического описания физических процессов лежащих в его основе (электрических тепловых металлургических процессов), то при создании систем для контроля качества сварных соединений применяются методы машинного обучения.

Интеллектуальные системы диагностики с точки зрения математики решают задачу классификации, когда по измеряемым параметрам протекания технологического процесса прогнозируется класс качества получаемой продукции. В таблице 1.1 приведен обзор классификаторов качества контактной точечной сварки, основанных на машинном обучении.

Таблица 1.1 – Классификаторы качества сварки, основанные на различных технологиях искусственного интеллекта [1].

Авторы	Используемый математический аппарат	Описание работы классификатора
В.С. Климов, А.С. Климов, С.А. Петрушкин, А.Д. Шижгин.	Деревья принятия решений	На основе измеренных параметров сварки начиная от корня дерева, совершая последовательный переход от узла к узлу дерева с учетом располагающихся в них условий, осуществляется спуск к одному из листов. В листе дерева содержится метка предполагаемого класса качества сварного соединения.
V. Balendran, A. Aravinthan, D. Al-Dabass, K. Sivayoganathan.	Нейронная сеть прямого распространения	Данные об изменении сопротивления между электродами в процессе сварки подаются в двухслойную нейронную сеть прямого распространения. Во втором слое нейронной сети находится один нейрон выходной сигнал, которого указывает на прогнозируемое значение прочности соединения. В зависимости от прогнозируемой прочности делается заключение о классе качества сварки.
J. Rönning, P. Laurinen, L. Tuovinen, H. Junno.	Байесовские сети доверия	Характеристики осциллограмм сварочного тока и напряжения и напряжения подаются на вход Байесовские сети доверия. На выходе сети вычисляется вероятность принадлежности соединения к классу качественных сварок.
R.V. Chinnam, D. Filev, M. El-Banna.	Нейронные сети LVQ	Вектор измеренных значений сопротивления на участке между электродами подаются на вход нейронной сети. Входные сигналы



		взвешиваются с помощью весовых коэффициентов и передаются от слоя к слою сети. На выходе нейронной сети активизируется нейрон с номером класса качества сварного соединения
--	--	---

Существуют такие алгоритмы машинного обучения, возможности которых не изучены в полной мере. Например, алгоритм кластеризации данных k-means.

Алгоритм k-means предназначен для автоматизированной группировки объектов по кластерам. При этом кластерная структура подбирается таким образом, чтобы объекты внутри одного кластера были максимально похожи друг на друга. Недостатком алгоритма k-means является, что он не предназначен для решения задач классификации. Недостатком алгоритма k-means является, что он не предназначен для решения задач классификации.

Поэтому целью данной работы является разработка алгоритма построения классификатора данных на основе результатов кластерного анализа для диагностики качества контактной сварки.

Для достижения данной цели требуется решение в бакалаврской работе следующих задач:

1. Провести анализ состояния вопроса теме использования алгоритмов машинного обучения в системах диагностики сварки.
2. Разработать алгоритм построения классификатора данных на основе результатов кластерного анализа. Протестировать предложенный алгоритм на данных контактной сварки.
3. Спроектировать и разработать программное обеспечение, реализующее предложенный алгоритм.

## 2 РАЗРАБОТКА АЛГОРИТМА КЛАССИФИКАЦИИ НА ОСНОВЕ РЕЗУЛЬТАТОВ КЛАСТЕРНОГО АНАЛИЗА

### 2.1 Кластерный анализ с использованием алгоритма k-means

Кластерный анализ – многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы.

Одним из наиболее распространённых алгоритмов неиерархической кластеризации является алгоритм k-means (Mac-Queen, 1967).

Исходные данные для алгоритма k-means:

1. Обучающая выборка  $X^m$ , состоящая из объектов  $x_1 \dots x_m$  с одинаковым набором атрибутов. Для всех объектов должны быть известны значения атрибутов  $P_1, \dots, P_n$ :

$$x_i = (P_1, P_2, \dots, P_n) \quad (2.1)$$

где  $n$  – количество атрибутов.

Атрибуты могут быть числовыми или категориальными. Таким образом, обучающая выборка задана так:

$$X^m = \{x_1, \dots, x_m\} \quad (2.2)$$

2. Метрика  $\rho(x, x')$  расчёта расстояний между объектами. Можно использовать одну из известных метрик – Евклида, Чебышева, расстояние Манхэттена и др.

Евклидово расстояние (норма/метрика  $\|x\|_2$ ). Евклидово расстояние между двумя объектами, один из которых описывается вектором  $x$ , а второй - вектором  $x'$ , будет рассчитываться так:

$$\|x - x'\|_2 = ((P_1 - P'_1)^2 + (P_2 - P'_2)^2 + \dots + (P_n - P'_n)^2)^{\frac{1}{2}}. \quad (2.3)$$

Множество точек, равноудаленных от некоторого центра при использовании евклидовой метрики будет образовывать круг в двумерном пространстве.

Расстояние Манхэтгена (норма/метрика  $\|x\|_1$ ). Данная норма имеет следующий вид:

$$\|x - x'\|_1 = (|P_1 - P_1'| + |P_2 - P_2'| + \dots + |P_n - P_n'|) \quad (2.4)$$

Преимущество метрики  $\|x\|_1$  заключается в том, что ее использование позволяет снизить влияние аномальных значений на работу алгоритмов. Множество точек, равноудаленных от некоторого центра при использовании метрики Манхэтгена будет образовывать квадрат в двумерном пространстве.

Расстояние Чебышева (норма/метрика  $\|x\|_\infty$ ). Данная норма имеет следующий вид:

$$\|x - x'\|_\infty = \max(|P_1 - P_1'|, |P_2 - P_2'|, \dots, |P_n - P_n'|) \quad (2.5)$$

Множество точек, равноудаленных от некоторого центра при использовании метрики Чебышева будет образовывать квадрат в двумерном пространстве.

- количество  $k$  кластеров, которое должно быть сформировано из объектов исходной выборки.

Алгоритм состоит из следующих шагов.

1. Случайным образом выбирается  $k$  объектов обучающей выборки, которые будут служить начальными центрами кластеров.

2. Для каждого объектов обучающей выборки определяется ближайший к ней центр кластера. Для этого вычисляется расстояние между объектами и центрами кластеров. Считается, что объект принадлежит тому кластеру, к которому он ближе. В качестве формулы для оценки близости объектов в многомерном пространстве признаков используется одна из известных метрик.

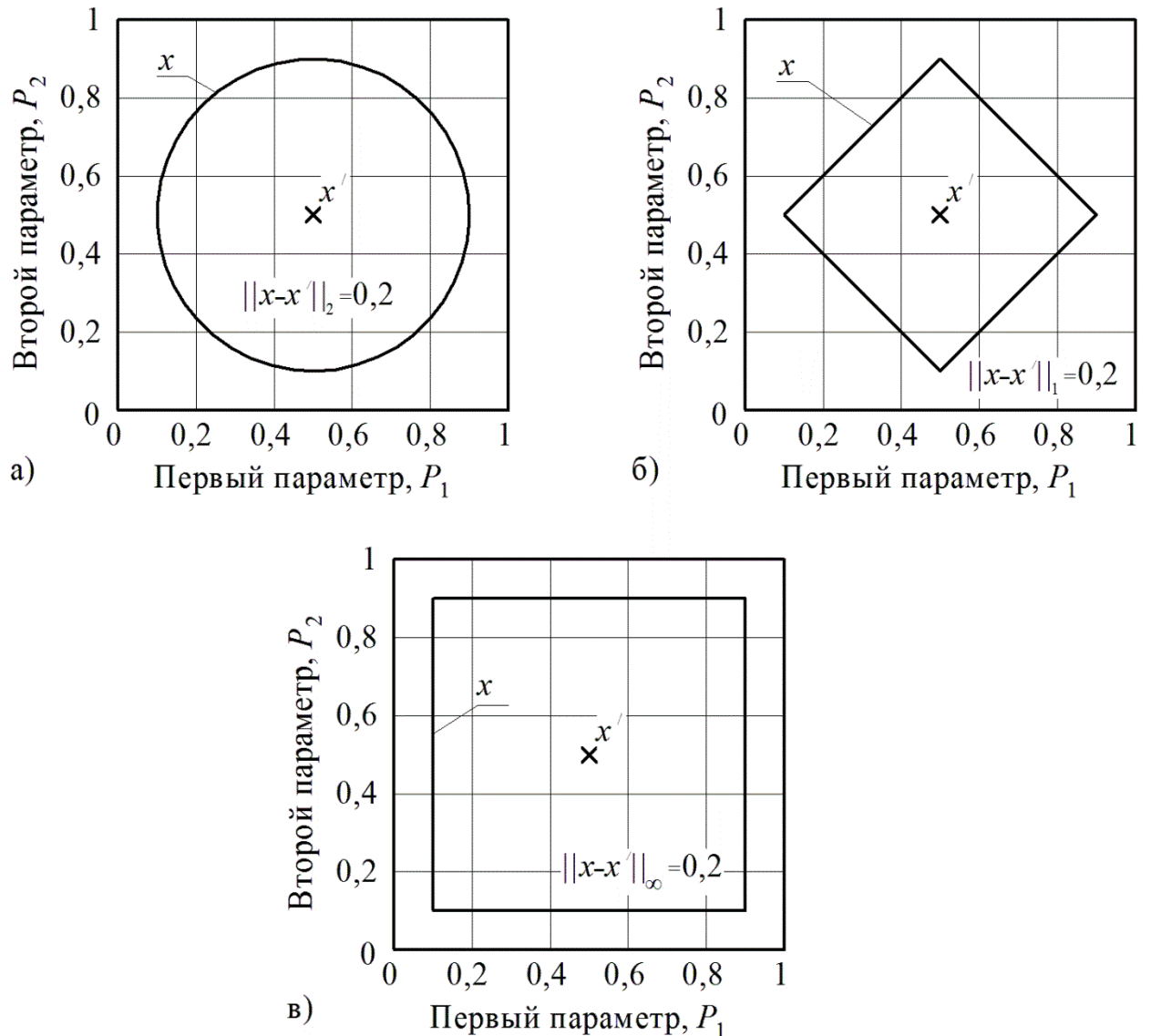


Рисунок 2.1 – Множество равноудаленных объектов  $x$  при использовании : а – метрика Евклида, б – метрика Манхэттена (прямоугольная метрика), в – метрика Чебышева.

3. Как только состав кластеров на данной итерации известен, производится расчёт новых центров кластеров. Это делается путем расчета определения средних значений для каждого числового признака по всем объектом рассматриваемого кластера. Например, в двухмерном пространстве координаты центр кластера на основе вошедших в него  $t$  объектов рассчитывается следующим образом:

$$(P_{2ц}, P_{1ц}) = \left( \frac{\sum_1^t P_1(t)}{t}, \frac{\sum_1^t P_2(t)}{t} \right) \quad (2.6)$$

4. Так как Шаги 2 и 3 повторяются до тех пор, пока не выполнятся один из двух критериев остановки:

- границы кластеров и расположения центров кластеров не перестанет изменяться от итерации к итерации, т.е. на каждой итерации в каждом кластере будет оставаться один и тот же набор записей. На практике алгоритм k-means обычно находит набор стабильных кластеров за несколько десятков итераций.

- достигнут критерий сходимости. Чаще всего используется критерий суммы квадратов ошибок между центром кластера и всеми вошедшими в него объектами:

$$E = \sum_{i=1}^k \sum_{p \in C_i} (p - m_i)^2 \quad (2.7)$$

где  $p \in C_i$  - произвольная точка данных, принадлежащая кластеру  $C_i$ ,  $m_i$  - центр данного кластера. Иными словами, алгоритм остановится тогда, когда ошибка  $E$  достигнет достаточно малого значения.

Пример использования алгоритма k-means для кластеризации объектов с двумя числовыми параметрами ( $P_1$  и  $P_2$ ) представлен на рисунке 2.2. В данном случае количество объектов равно 24, количество кластеров  $k=3$ , метрика для оценки сходства объектов – расстояние Евклида. Пунктиром показаны границы кластеров, которые зависят от:

- расположения центров кластеров;
- метрики для оценки близости объектов.

Символами «×» показаны центры кластеров, расположение которых зависит от:

- объектов обучающей выборки;
- метрики для оценки близости объектов.

К достоинствам алгоритма кластеризации данных k-means можно отнести следующее:

- Умеренные вычислительные затраты, которые растут линейно с увеличением числа записей исходной выборки данных. Вычислительная сложность алгоритма определяется как  $k \times n \times l$ , где  $k$  – число кластеров,  $n$  – число записей и  $l$  – число итераций.
- Результаты его работы не зависят от порядка следования записей в исходной выборке.

К недостаткам алгоритма кластеризации данных k-means относится:

- Чувствительность алгоритма к шумам и аномальным значениям в данных, поскольку они способны значительно повлиять на среднее значение, используемое при вычислении положений центров кластеров.

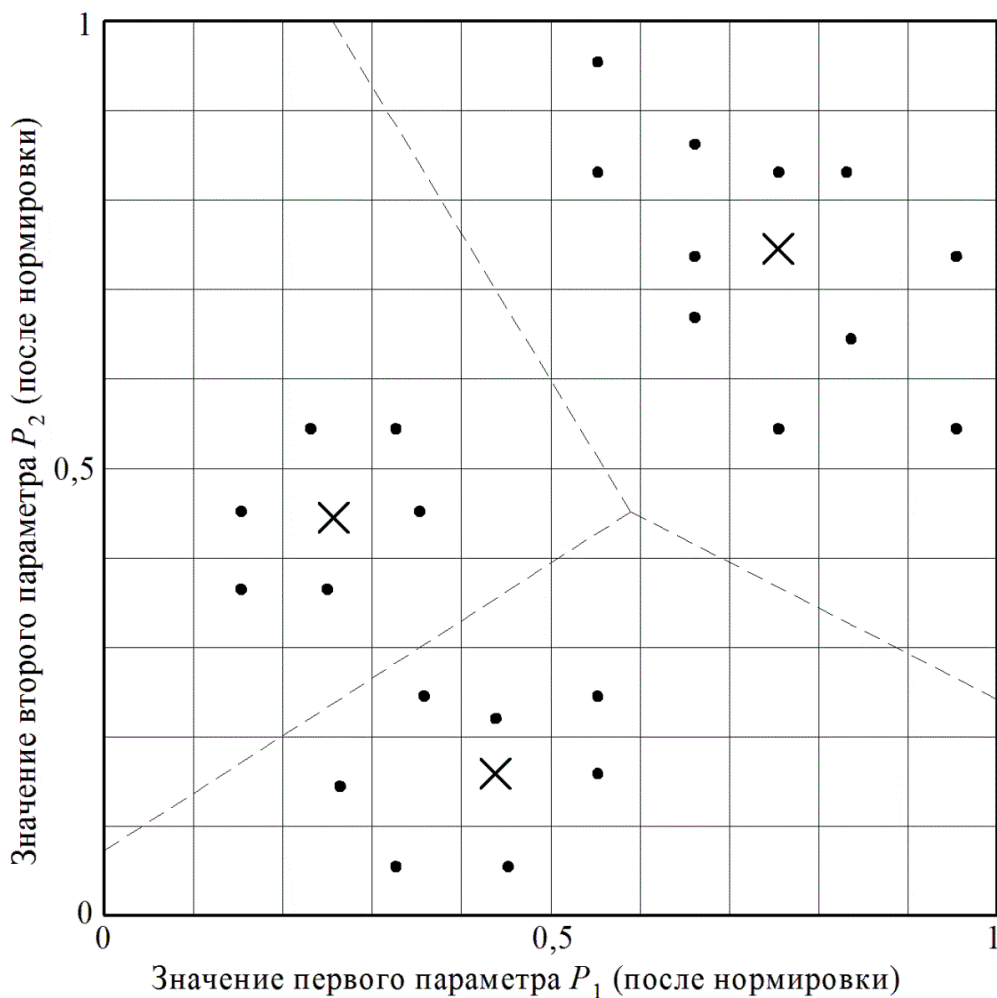


Рисунок 2.2 – Пример кластеризации объектов с двумя числовыми параметрами

Существует ряд проблем алгоритма кластеризации k-means которые до сих пор являются предметом обсуждения в научных статьях. Одной из таких проблем является учет категориальных признаков при сравнении объектов друг с другом. Так как оценка близости объектов относительно друг друга предполагает получение неотрицательного числа, а категориальный признак является нечисловым параметром, то необходимо введение дополнительных функций для сравнения категориальных параметров объектов. Наиболее простым выходом является введение функции отличия  $f_{\text{отл}}$ , которая возвращает ноль, когда сравниваемые категориальные признаки  $P'_c$  и  $P_c$  объектов  $x'$  и  $x$  равны и единицу, когда не равны. Формальная запись такой функции:

$$f_{\text{отл}}(P_c, P'_c) = \begin{cases} 0 & \text{если } P_c = P'_c \\ 1 & \text{если } P_c \neq P'_c \end{cases} \quad (2.8)$$

Для учета значений категориальных признаков функцией отличия модифицируются формулы расчета метрик. В нашем случае в обучающей выборке категориальные признаки отсутствуют, поэтому использование функции отличия не требуется.

Другой проблемой использование алгоритма k-means является необходимость выполнения нормировки числовых параметров объектов. Это необходимость связана с масштаб изменения числовых параметров объектов. Так при оценке близости объектов наибольшее влияние на конечный результат сравнения оказывает тот числовой параметр, масштаб которого самый большой. Для того чтобы уровнять степень влияние всех числовых признаков на результат сравнения их близости с использованием выбранной метрики, самым простым решением является нормировка всех числовых параметров путем линейного преобразования к диапазону  $[0, 1]$ . Так нормированные значение вектора  $i$ -ого числового параметра по всей обучающей выборке рассчитывается как:

$$P_i^H = \frac{P_i - \min(P_i)}{\max(P_i) - \min(P_i)} \quad (2.9)$$

В нашем случае необходимо проведение нормировки числовых параметров объектов, представленных в обучающей выборке, поэтому она будет проводиться по формуле 2.9.

Другой проблемой использование алгоритма k-means является проблема определения оптимального количества кластеров. Это связано с тем, что количество кластеров получаемой структуры является входным параметром алгоритма. Оптимальное количество кластеров определится путем анализа обучающей выборки с использованием стороннего математического аппарата. Другим вариантом определения оптимального количества кластеров – опытным путем (за счет проведения вычислительных экспериментов с заданием различных значений  $k$ ). В нашем случае мы будем использовать второй способ, выбрав в качестве критерия оптимальности точность классификации объектов.

Другой особенностью использование алгоритма k-means является проблема выбора метрики для оценки сходства объектов. Суть проблемы заключается в том, что использование различных метрик влияет на результат каждой итерации алгоритма k-means. Различия в расчете расстояния между объектами приводят к тому, что при одних и тех же исходных данных объекты начинают относиться к разным кластерам. При этом разные метрики приводят к различному расположению центров кластеров. Чем больше итераций требуется для кластеризации данных, тем сильнее отличаются кластерные структуры, полученные при использовании различных метрик.

Таким образом, оптимальной метрика, позволяющая обеспечить наименьшую ошибку кластеризации, подбирает в зависимости от исходных данных опытным путем. Процедура определения оптимальной метрики является вопросом обсуждения в научных статьях. При этом исследователи сходятся во мнении, что в большинстве случаев достаточным является использование метрики Евклида.



Таким образом, для нашего случая, при кластеризации данных, для использования выбрана метрика Евклида.

## 2.2 Анализ данных процесса контактной сварки

Интеллектуальные системы диагностики технологических процессов с точки зрения математики решают задачу классификации. Т.е. на основе измеряемых параметров такие системы прогнозируют класс качества получаемой на выходе продукции.

В нашем случае таким технологическим процессом является контактная точечная сварка. Результатом процесса сварки является сварная точка, качество которой необходимо прогнозировать.

Выбранный критерием качества сварки – предельная сила разрушения сварного соединения при испытаниях на срез больше 7200 Н. Сварные соединения удовлетворяющие данному критерию – качественные (относятся к классу С1), а сварные соединения, не удовлетворяющие данному критерию, считаются бракованными (относятся к классу С2).

Измеряемыми в процессе каждой сварки параметрами являются – значения коэффициента мощности  $\cos\varphi$  в каждом периоде сетевого напряжения. Т.е. каждая сварка описывается вектором значений  $(\cos\varphi_1, \cos\varphi_2, \dots, \cos\varphi_m)$ .

Существуют исследования [1], направленные на обоснование связи между динамикой изменения коэффициента мощности  $\cos\varphi$  в процессе сварки и металлургическими процессам, происходящими между электродами (нагрев, плавление, кристаллизация). При этом в исследованиях отмечается важность не абсолютных значений  $\cos\varphi$ , а динамика изменений значений в процессе сварки.

Для того чтобы диагностировать сварку по динамике изменения  $\cos\varphi$  в процессе сварки вектор значений  $(\cos\varphi_1, \cos\varphi_2, \dots, \cos\varphi_m)$  нормируется в масштабах одной сварки к диапазону от 0 до 100. Таким образом, получаем

вектор значений  $(P_1, P_2, \dots, P_m)$ , где значение 0 соответствует минимальному значению  $\cos\varphi$ , а 100 максимальному значению  $\cos\varphi$  за сварку (рисунок 2.3).

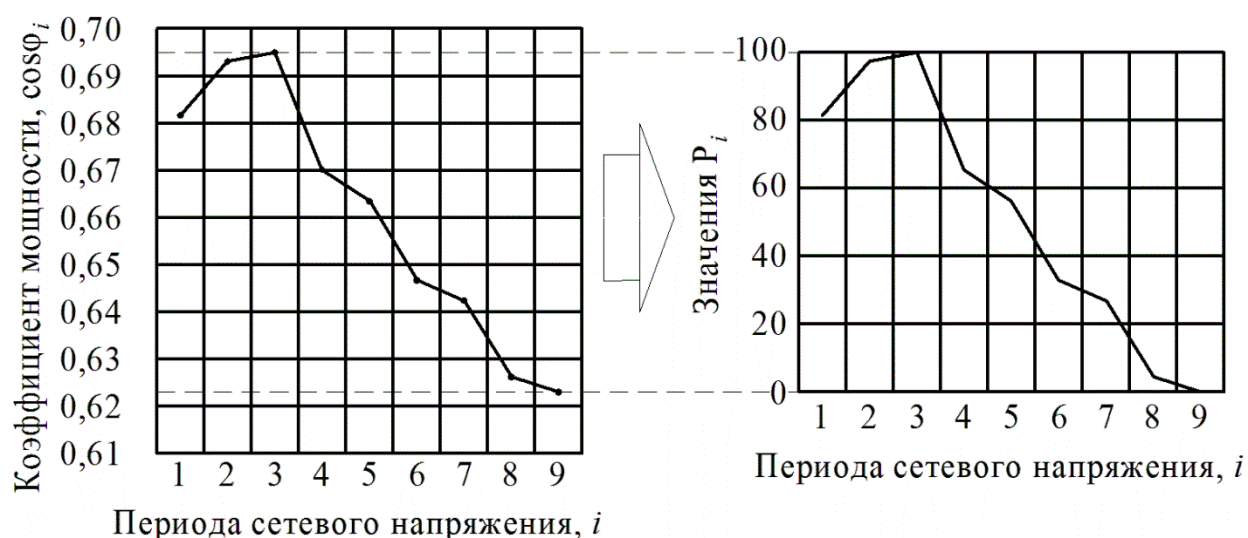


Рисунок 2.3 – Пример нормировки данных сварки

В нашем распоряжении имеются данные сварки пакета заготовок из низкоуглеродистой стали. Пакет состоит из двух листов толщиной по 0,8 миллиметров каждый, расположенные внахлест. При этом длительность сварки – 9 периодов сетевого напряжения (поэтому вектор входных параметров  $(P_1, P_2, \dots, P_9)$  состоит из 9 значений). Все экспериментальные соединения подвергались тестированию на разрывной машине для определения максимальной силы разрушения сварки. Если сила разрушения сварного соединения превышала 7200 Н, то оно соответствовало классу С1 (качественная сварка), в противном случае соединение соответствует классу С2 (некачественная сварка). Пример экспериментальных данных представлен в таблице 2.1.

Таблица 2.1 – Экспериментальные данные процесса сварки

№	P1	P2	P3	P4	P5	P6	P7	P8	P9	C
1	61,84	85,41	100	100	32,54	16,68	0	0	0	C2
2	100	74,67	65,61	65,61	65,61	65,61	21,35	10,94	0	C2
3	83,82	17,6	83,82	100	100	100	100	17,6	0	C2
4	87,51	66,32	100	66,32	59,02	59,02	17,05	8,74	0	C2

5	62,17	93,74	100	68,67	93,74	62,17	15,99	8,19	0	C2
6	93,18	58,8	93,18	100	65,87	51,34	51,34	8,49	0	C2
7	76,27	71,38	100	80,96	76,27	42,13	36,56	30,72	0	C2
8	100	84,52	80,3	80,3	51,66	46,84	11,73	6,01	0	C2
9	94,5	100	100	100	64,34	58,33	14,61	7,48	0	C2
10	74,76	94,61	100	100	63,84	58,03	14,14	7,24	0	C2
11	88,99	100	100	100	63,84	20,71	7,24	0	0	C2
12	94,86	83,99	100	83,99	62,36	36,4	29,23	4,33	0	C2
13	100	100	100	75,18	45,81	36,95	5,3	0	0	C2
14	80,55	97,21	100	60,78	50,59	25,8	21,64	0	12,74	C2
15	93,75	100	100	100	75,75	47,47	13,84	4,82	0	C2
16	93,19	100	82,19	79,85	58,93	37,22	11,28	3,92	0	C2
17	84,36	100	88,49	70,73	49,78	27,45	21,11	17,74	0	C2
18	68,24	100	76,36	54,28	32,82	27,25	6,77	3,46	0	C2
19	100	95,29	95,29	78,55	39,56	28,6	6,91	3,53	0	C2
20	86,05	100	86,05	61,55	35,59	28,64	8,05	0	0	C1
21	83,54	100	86,05	61,55	35,59	28,64	8,05	0	0	C1
22	77,2	95,42	100	64	37,08	11,27	3,92	0	0	C1
23	88,58	100	92,07	69,02	18,99	5,34	0	0	0	C2
24	85,78	100	80,68	9,26	12,6	12,6	12,6	0	10,16	C1
25	100	94,73	83,15	69,97	8,65	2,63	0	0	0	C2
26	100	92,12	0	6,39	6,39	6,39	6,39	6,39	6,39	C1
27	93,17	100	10,16	19,38	19,38	19,38	2,63	0	0	C1
28	97,97	100	90,99	15,97	2,56	11,86	2,56	0	0	C2
29	90,73	100	1,93	0	0	0	0	0	0	C1
30	100	87,75	75,44	62,35	5,86	0	0	0	0	C2
31	100	95,74	76,04	62,12	27,91	17,35	4,49	0	0	C1
32	100	82,87	51,04	42,35	25,61	21,89	6,5	2,25	0	C1
33	100	90,12	56,33	48,79	46,71	29,89	8,1	2,81	0	C1
34	100	68,81	17,47	19,91	19,91	19,91	19,91	4,42	0	C1
35	100	76,61	33,39	36,36	22,71	19,5	5,62	1,95	0	C1
...										
108	100	78,82	45,67	37,82	36,07	21,21	17,28	15,2	0	C1

Представленная выборка данных состоит из результатов исследования сварных соединений, полученных в различных условиях.

Цель построения классификатора, научиться на основе значений  $(P_1, P_2, \dots, P_9)$  прогнозировать класс качества сварного соединения (C1/C2).

## **2.3 Разработка алгоритма построения классификатора на основе результатов кластерного анализа и алгоритма классификации объектов**

Как было сказано выше, алгоритм k-means предназначен для автоматизированной группировки объектов по кластерам. При этом кластерная структура подбирается таким образом, чтобы объекты внутри одного кластера были максимально похожи друг на друга.

Идея построения классификатора заключается в том, чтобы сначала поделить объекты обучающей выборки на кластеры и изучив состав каждого кластера сформировать стратегию классификации данных. Изучение состава кластеров необходимо для того, чтобы определить, как распределены метки классов по кластерам. Затем на основе изученных свойств кластеров предлагается классифицировать новые (ранее неизвестные) объекты. При классификации можно с помощью одной из известных метрик определить близость объекта до каждого из известных центров кластеров. Предположительно метка класса классифицируемого объекта будет та же, что и у большинства объектов находящихся в ближайшем центре кластера.

Таким образом, предложено 2 алгоритма:

1. Алгоритм построения классификатора на основе кластерного анализа данных.
2. Алгоритм классификации объектов с помощью полученных классификаторов.

Алгоритм построения классификатора на основе кластерного анализа данных состоит из следующих этапов:

1. Подготавливается обучающая выборка – данные, на основе которых будет строиться (обучаться) классификатор (рисунок 2.4).
2. Выборка данных подвергается кластеризации с использованием алгоритма k-means. При этом задается количество кластеров  $k$ , выбирается метрика расчета расстояния между объектами. После выполнения данного этапа мы получаем распределение объектов по кластерам и координаты центров каждого кластера (рисунок 2.5).

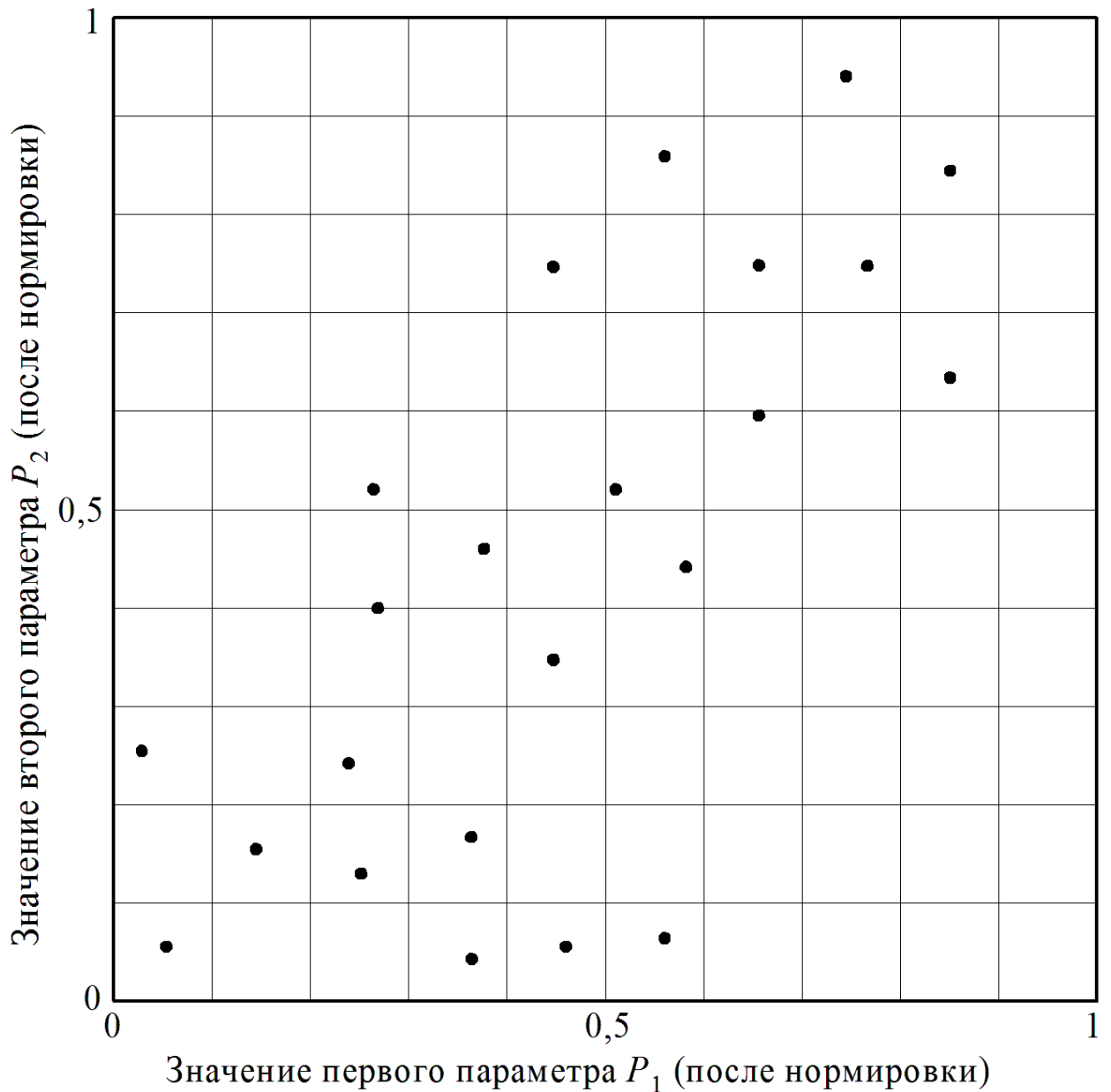


Рисунок 2.4 – Первый этап – подготовка данных для анализа (в т.ч. нормировка)

3. Производится статистический анализ каждого кластера. Целью анализа является исследование распределение классов по кластерам. При этом подсчитывается количество объектов в кластере, относящихся к каждому из классов. Впоследствии для каждого кластера определяется доминирующий класс (рисунок 2.6).

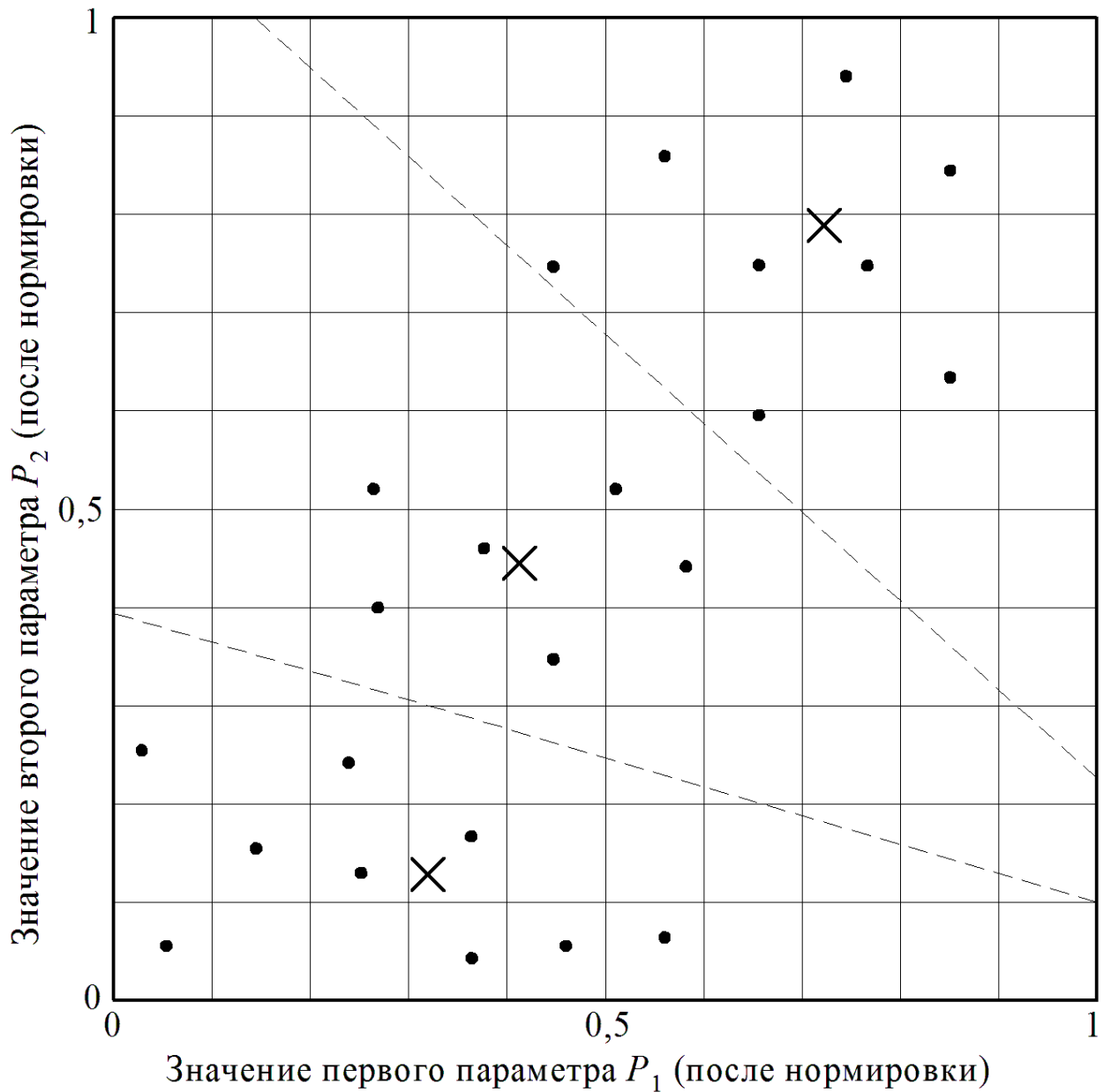
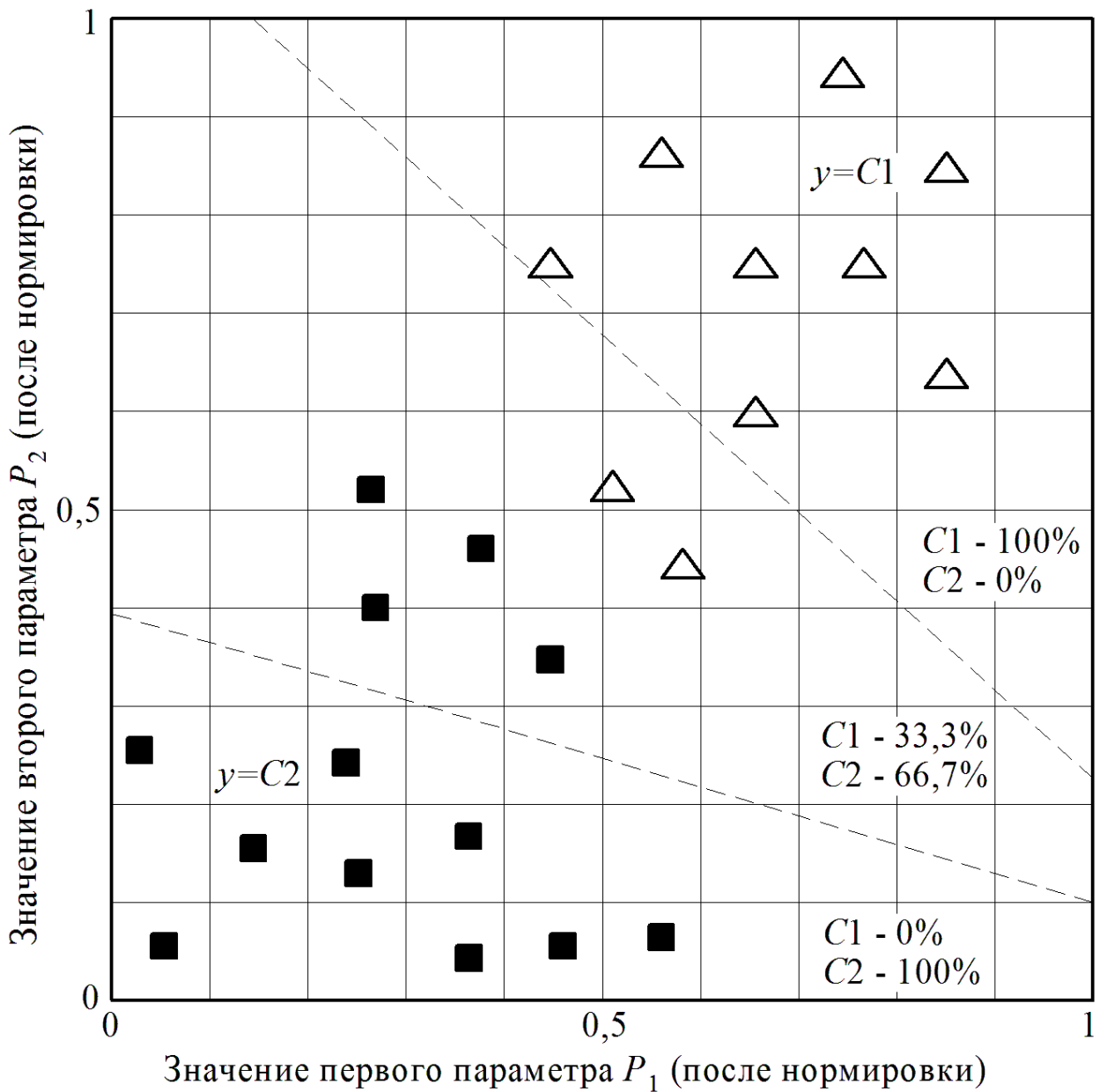


Рисунок 2.5 – Второй этап – кластеризация данных методом k-means с заданным значением  $k$  количества кластеров. Знаками «x» обозначены центры кластеров. Пунктирными линиями обозначены границы кластеров.

4. Производится объединение полученных на предыдущих этапах данных в классификатор. Таким образом, классификатор состоит из координат центров кластеров и метки класса, возвращаемой каждым центром кластера.



Алгоритм классификации объекта с помощью полученного классификатора состоит из следующих этапов:

1. С помощью выбранной метрики рассчитывается расстояние между вектором входных параметров исследуемого объекта и центрами кластеров, хранящихся в классификаторе. Определяется ближайший центр кластера.

2. Исследуемый объект будет отнесен к тому классу, который доминирует в кластере с ближайшим центром.

Графическая интерпретация классификации объекта представлена на рисунке 2.7.

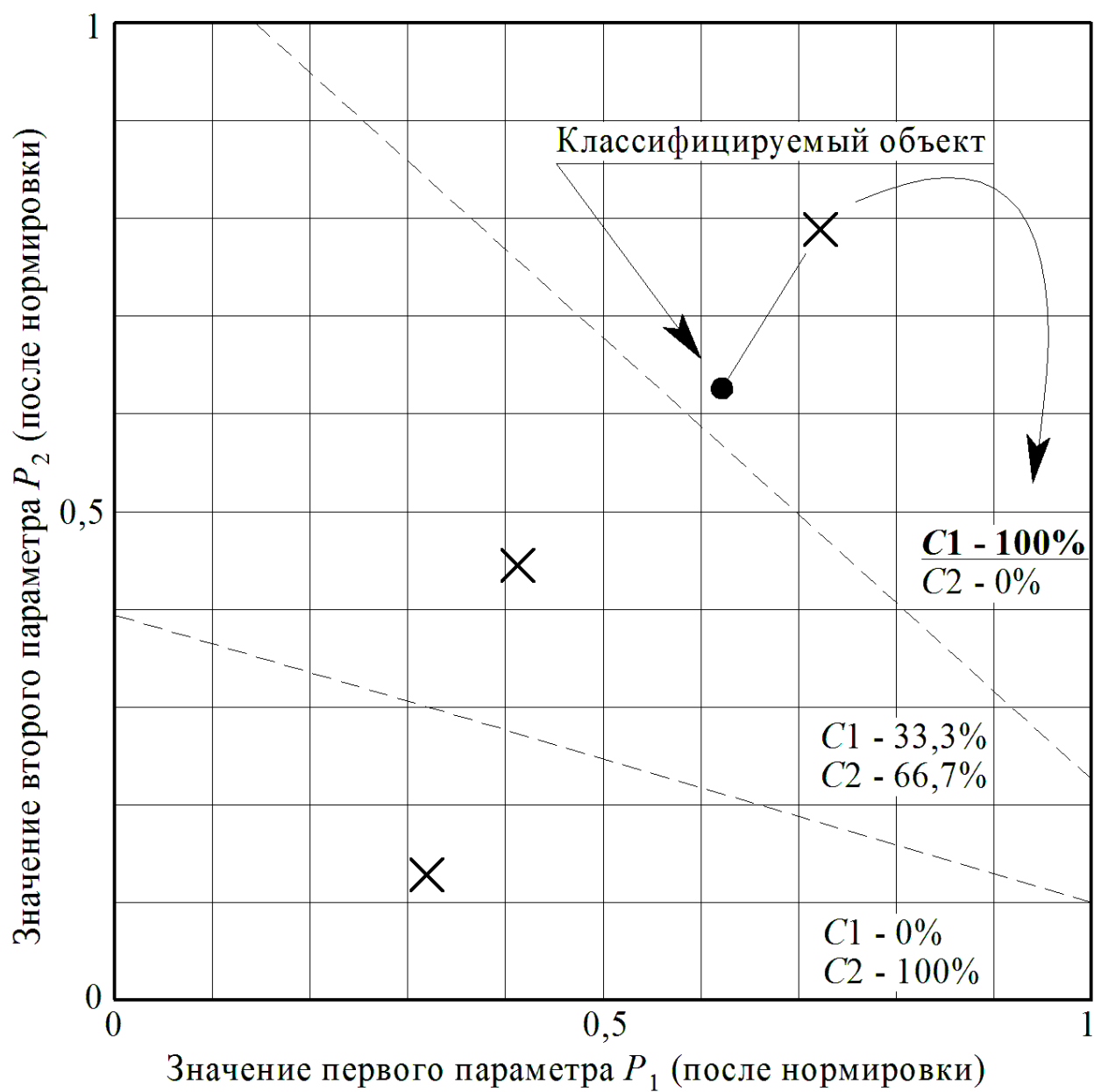


Рисунок 2.7 – Классификация объекта путем определения ближайшего центра кластера



### **3 ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ПРЕДЛОЖЕННЫХ РЕШЕНИЙ**

#### **3.1 Пример использования алгоритма построения классификатора**

Проведем тестирование предложенного алгоритма построения классификатора на основе кластерного анализа данных и исследуем точность работы классификатора.

Для этого возьмем экспериментальные данные, полученные при выполнении 118 сварных соединений. Формат данных, значение входных и выходных параметров описаны в подпункте бакалаврской работы «Анализ данных процесса контактной сварки».

Данные 118 сварок разделены на 2 группы по 59 сварок в каждой. Первая группа данных будет использована для построения классификаторов качества сварки, а вторая группа данных будет использована для проверки точности работы классификаторов.

На одних и тех же данных будет построено 5 классификатор, их отличие будет заключаться в том, что на этапе кластеризации будет выбрано различное количество кластеров: 2, 3, 4, 5 и 6

На первом этапе построения классификаторов в соответствии с предложенным алгоритмом производится кластеризация данных методом k-means. Результат кластеризации данных 59 сварок при различном количестве кластеров представлен в таблице 3.1

Таблица 3.1 – Результат выполнения кластеризации при различных заданных значениях  $k$  (количество кластеров)

Количество кластеров $k$	№ центра кластера	Координаты центров кластеров
k=2	I	$P_I = (86,48451; 88,74093; 92,581772; 77,352448; 56,802238; 40,259346; 20,004526; 7,493294; 0,318383)$
	II	$P_{II} = (95,69094; 92,79401; 56,200043; 36,538143; 20,486128; 15,614007; 8,836353; 4,094865; 0,871150)$
k=3	I	$P_I = (86,56165; 51,69834; 92,494995; 95,238953; 75,288399; 61,207108; 57,805912; 25,033615; 0,000000)$
	II	$P_{II} = (96,05547; 91,17162; 48,072388; 31,049023; 17,801304; 14,350900; 7,915840; 4,377267; 1,103456)$
	III	$P_{III} = (87,26076; 93,45891; 92,000153; 73,540787; 52,328815; 36,173706; 15,452765; 5,293518; 0,318383)$
k=4	I	$P_I = (83,82353; 17,59962; 83,823532; 100,000000; 100,000000; 100,000000; 100,000000; 17,599621; 0,000000)$
	II	$P_{II} = (95,50500; 89,92255; 34,945366; 21,324818; 16,527857; 15,278824; 9,354776; 6,285310; 0,639039)$
	III	$P_{III} = (74,72267; 95,60790; 91,408646; 57,665508; 33,126686; 18,720758; 8,988028; 0,691789; 4,579352)$
	IV	$P_{IV} = (89,88422; 91,11301; 90,250031; 74,110397; 51,870090; 36,294144; 16,967068; 6,828431; 0,000000)$
k=5	I	$P_I = (83,82353; 17,59962; 83,823532; 100,000000; 100,000000; 100,000000; 100,000000; 17,599621; 0,000000)$
	II	$P_{II} = (95,50500; 89,92255; 34,945366; 21,324818; 16,527857; 15,278824; 9,354776; 6,285310; 0,639039)$

	III	$P_{III} = (83,16751; 98,60444; 90,340591; 35,021648; 31,597050; 19,199280; 17,122169; 0,000000; 11,448380)$
	IV	$P_{IV} = (94,73449; 90,01495; 88,441071; 73,159386; 48,892857; 34,069389; 16,050776; 5,523622; 0,000000)$
	V	$P_V = (74,34259; 94,15794; 94,785652; 75,995079; 54,877266; 37,545208; 16,189636; 8,594682; 0,000000)$
k=6	I	$P_I = (83,82353; 17,59962; 83,823532; 100,000000; 100,000000; 100,000000; 100,000000; 17,599621; 0,000000)$
	II	$P_{II} = (95,50500; 89,92255; 34,945366; 21,324818; 16,527857; 15,278824; 9,354776; 6,285310; 0,639039)$
	III	$P_{III} = (83,16751; 98,60444; 90,340591; 35,021648; 31,597050; 19,199280; 17,122169; 0,000000; 11,448380)$
	IV	$P_{IV} = (94,79133; 91,01486; 88,294891; 72,293556; 48,566254; 33,511616; 15,170963; 4,304225; 0,000000)$
	V	$P_V = (74,19459; 95,91010; 94,384552; 75,613503; 53,231930; 37,192471; 14,622588; 6,892562; 0,000000)$
	VI	$P_{VI} = (84,61945; 65,19872; 96,486145; 90,477905; 67,642120; 46,745495; 39,943100; 37,023590; 0,000000)$

На втором этапе предложенного алгоритма построения классификатора статистический анализ объектов каждого кластера. При этом исследуется распределение классов по кластерам: подсчитывается количество объектов в кластере, относящихся к каждому из классов. Впоследствии для каждого кластера определяется доминирующий класс. Результат данного этапа представлен в таблице 3.2.

Таблица 3.2 – Определение преобладающего класса в каждом кластере

Количество кластеров $k$	№ центра кластера	Всего элементов выборки	Количество элементов (%) класса C1	Количество элементов (%) класса C2	Класс, возвращаемый кластером
k=2	I	59	6 (15%)	34 (85%)	C2
	II		14 (73,68%)	5 (26,32%)	C1
k=3	I	59	0 (0%)	4 (100%)	C2
	II		11 (73,33%)	4 (26,67%)	C1
	III		9 (22,5%)	31 (77,5%)	C2
k=4	I	59	0 (0%)	1 (100%)	C2
	II		9 (75%)	3 (25%)	C1
	III		2 (33,33%)	4 (66,67%)	C2
	IV		9 (22,5%)	31 (77,5%)	C2
k=5	I	59	0 (0%)	1 (100%)	C2
	II		8 (80%)	2 (20%)	C1
	III		1 (50%)	1 (50%)	C2
	IV		9 (28,12%)	23 (71,88%)	C2
	V		2 (14,29%)	12 (85,71%)	C2
k=6	I	59	0 (0%)	1 (100%)	C2
	II		8 (80%)	2 (20%)	C1
	III		1 (50%)	1 (50%)	C2
	IV		9 (29,03%)	22 (70,97%)	C2
	V		2 (15,38%)	11 (84,62%)	C2
	VI		0 (0%)	2 (100%)	C2

После того, как классификаторы при различных значениях параметра  $k$  построены протестируем их точность работы на примере других данных 59 сварок (классификатор состоит из параметров центров кластеров и метки

класса, возвращаемой каждым кластером). Результаты точности определения классов качества сварных соединений представлены на рисунке 3.1

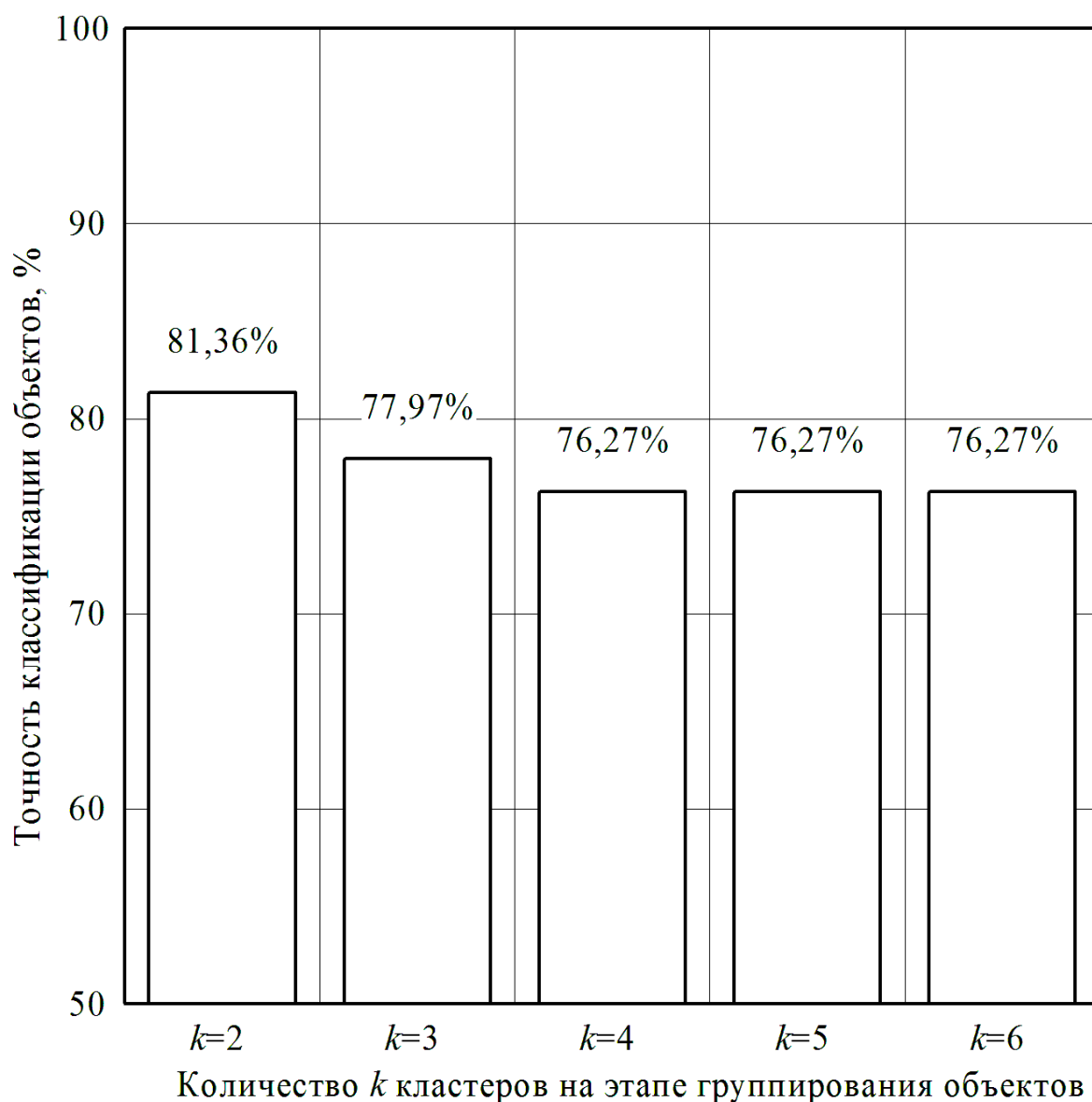


Рисунок 3.1 – Результат исследования точности классификации объектов на одних и тех же данных при различном значении  $k$

Хоть на рисунке 3.1 с увеличением количества кластеров и происходит падение точности работы классификатора, нельзя утверждать, что такая тенденция будет сохранена и для других данных (из другой предметной области). Оптимальное количество кластеров в большей мере зависит от природы данных и в каждом конкретном случае требует отдельного исследования. Точность классификации качества сварки в данном случае можно оценить как удовлетворительную (точность 76,27...81,36 %).

### 3.2 Пример классификации объекта

Приведем расчёты при классификации качества сварки на примере данных одной из сварок, содержащихся в тестовой выборке. При этом будем использовать классификатор полученный при  $k=2$  (количество кластеров). На основе анализа измеренный в процессе сварки параметров был получен вектор значений  $\mathbf{P}_x$ :

$$\mathbf{P}_x = (81,42; 97,33; 100; 65,50; 56,17; 32,89; 25,12; 4,44; 0)$$

Определим, к какому классу относится сварное соединение с таким вектором параметров. В нашем случае присутствует 2 класса: С1 – качественная сварка (сила разрушения сварного соединения  $>7200\text{Н}$ ) и С2 – некачественная сварка.

Для классификации сначала воспользуемся кластерной структурой состоящей из двух кластеров, полученной путем анализа алгоритмом k-means экспериментальных данных процесса сварки. Координаты первого  $\mathbf{P}_I$  и второго  $\mathbf{P}_{II}$  центра кластеров (таблица 3.1, при  $k=2$ ):

- $\mathbf{P}_I = (86,48451; 88,74093; 92,581772; 77,352448; 56,802238; 40,259346; 20,004526; 7,493294; 0,318383)$

- $\mathbf{P}_{II} = (95,69094; 92,79401; 56,200043; 36,538143; 20,486128; 15,614007; 8,836353; 4,094865; 0,871150).$

Рассчитаем, к какому центру кластера исследуемый объект  $\mathbf{P}_x$  ближе, для этого воспользуемся метрикой Евклида. Расстояние между  $\mathbf{P}_x$  и  $\mathbf{P}_I$ :

$$\begin{aligned} \|\mathbf{P}_x - \mathbf{P}_I\|_2 = & [(81,42 - 86,48451)^2 + (97,33 - 88,74093)^2 + (100 - 92,581772)^2 + \\ & + (65,50 - 77,352448)^2 + (56,17 - 56,802238)^2 + (32,89 - 40,259346)^2 + \\ & + (25,12 - 20,004526)^2 + (4,44 - 7,493294)^2 + (0 - 0,318383)^2]^{\frac{1}{2}} = 19,63 \end{aligned}$$

$$\begin{aligned} \|\mathbf{P}_x - \mathbf{P}_{II}\|_2 = & [(81,42 - 95,69094)^2 + (97,33 - 92,79401)^2 + (100 - 56,20004)^2 + \\ & + (65,50 - 36,53814)^2 + (56,17 - 20,48613)^2 + (32,89 - 15,61401)^2 + \\ & + (25,12 - 8,836353)^2 + (4,44 - 4,094865)^2 + (0 - 0,87115)^2]^{\frac{1}{2}} = 69,43 \end{aligned}$$

Как видно из расчётов, вектор  $P_x$  ближе к первому центру кластера, чем ко второму, поэтому обратим к статистическим данным первого кластера (первая строчка таблицы 3.2). Как видно, в первом кластере содержится 40 объектов обучающей выборки из которых большинство (34 объекта) относятся к классу C2. Это значит, что сварка, описываемая вектором значений  $P_x$  относится к классу C2 (некачественная сварка).

### 3.2 Программная реализация

Для реализации алгоритмов построения классификатора и классификации объектов было разработано приложение, внешний вид которого представлен на рисунке 3.2.

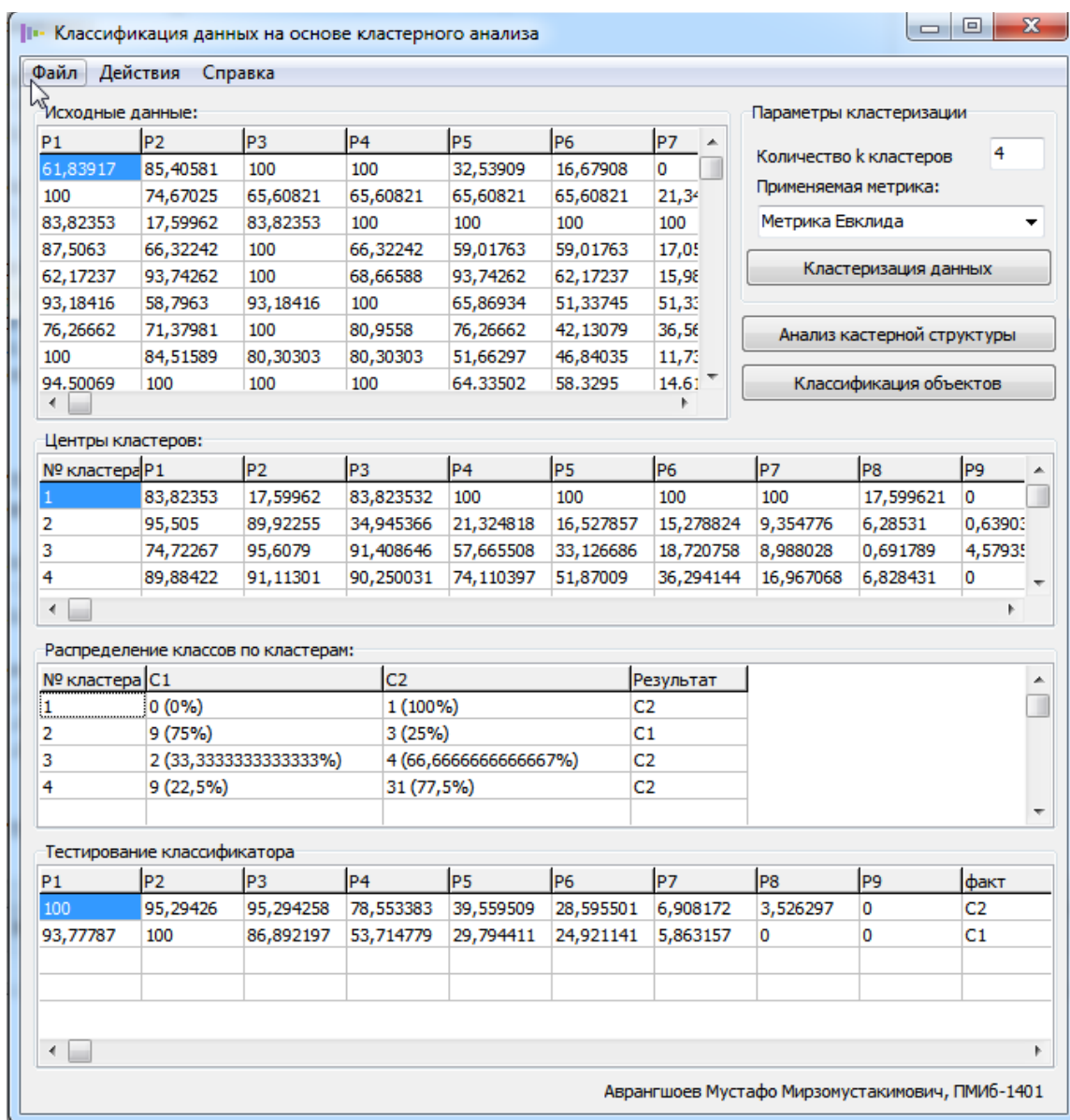


Рисунок 3.2 – Внешнее окно программы

Разработанное приложение обладает следующими функциональными особенностями:

- реализация предложенного алгоритма построения классификатора на основе результатов кластерного анализа исходных данных;
- реализация алгоритма классификации объектов, основанный на определении близости объектов к центрам кластеров и распределении классов по кластерам.
- импорт исходных данных для построения классификатора из xslx-файла (Excel) и их просмотр в приложении;
- возможность сохранения параметров центров кластеров в xslx-файл (Excel);
- возможность сохранения статистических данных по распределению классов по кластерам в xslx-файл (Excel);
- наличие удобного пользовательского графического интерфейса.

Созданное приложение работает следующим образом. Сначала необходимо загрузить в экспериментальные данные сварки в программу. Для этого необходимо выбрать подпункт «Открыть исходные данные» пункта «файл» (рисунок 3.3). В результате выполнения этого действия данные будут импортированы из выбранного xslx-файла.



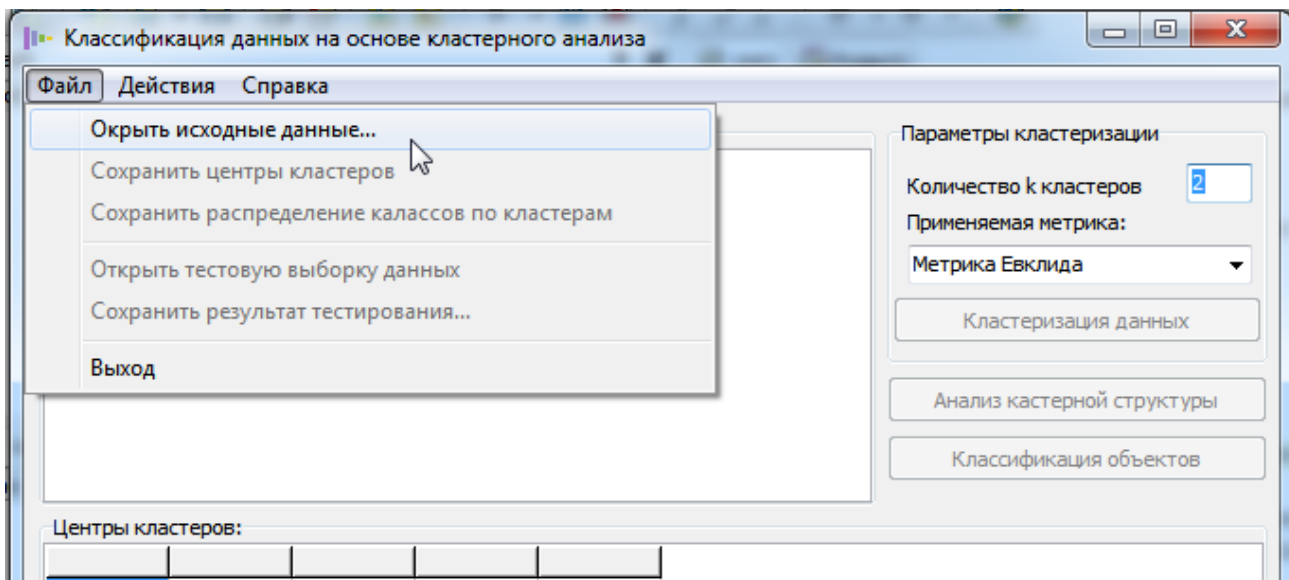


Рисунок 3.3 – Содержимое пункта меню «Файл»

Импортированные исходные данные будут отображены в таблице «исходные данные», расположенной в левой верхней части окна программы (рисунок 3.4).

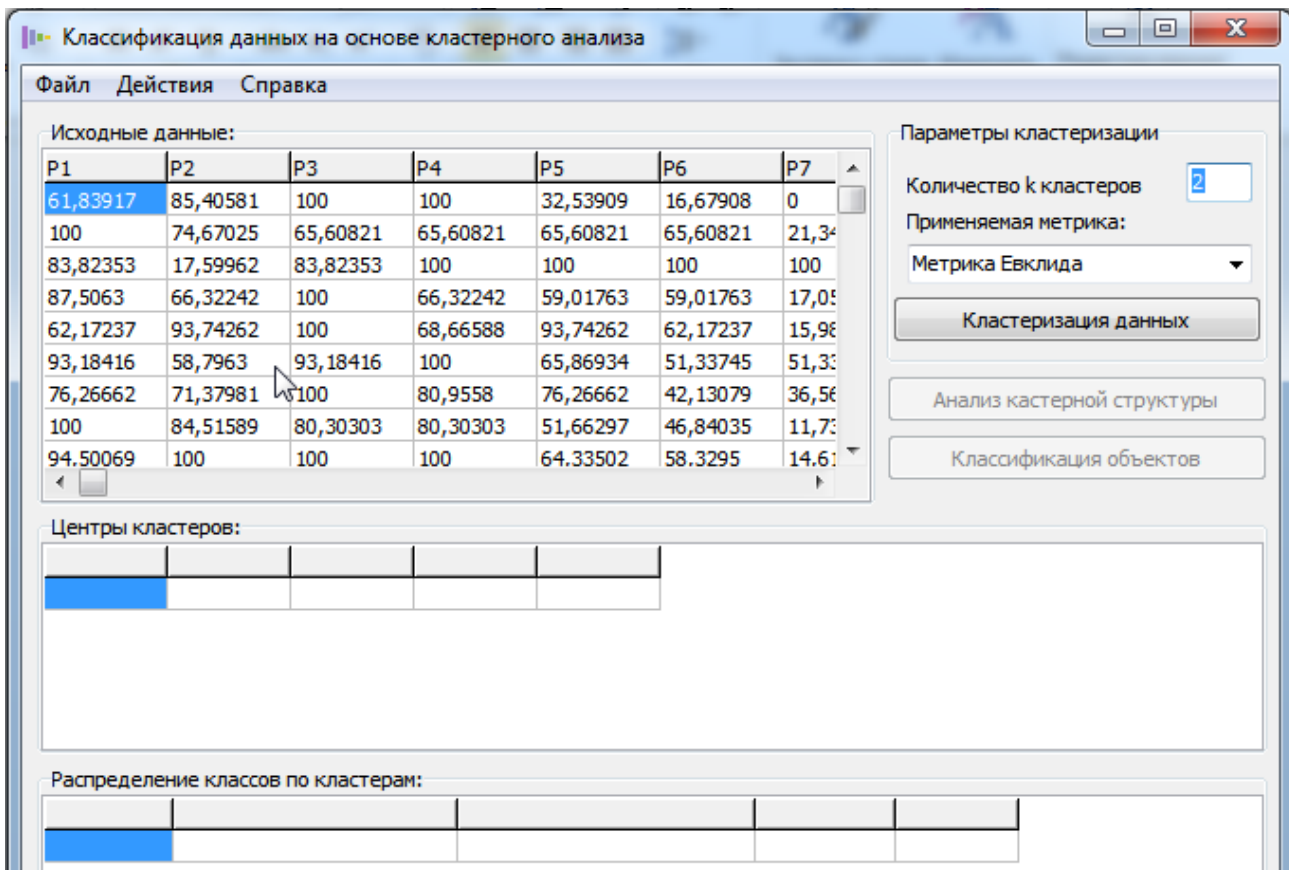


Рисунок 3.4 – Результат загрузки экспериментальных данных сварки

Первым шагом построения классификатора данных (описанного во второй главе бакалаврской работы) является кластеризация данных. Для ее выполнения в блоке «Параметры кластеризации» необходимо задать настройки кластеризации: количество кластеров, метрику для оценки расстояния между объектами. При этом надо учитывать, что параметры кластеризации влияют на точность получаемого классификатора (см. пример на рисунке 3.1).

После того, как параметры кластеризации будут заданы для ее запуска необходимо нажать на кнопку «Кластеризация данных».

В результате кластеризации будут определены центры кластеров. Рассчитанные параметры центров кластеров будут отображены в таблице «Центры кластеров». Результат выполнения данного этапа показан на рисунке 3.5

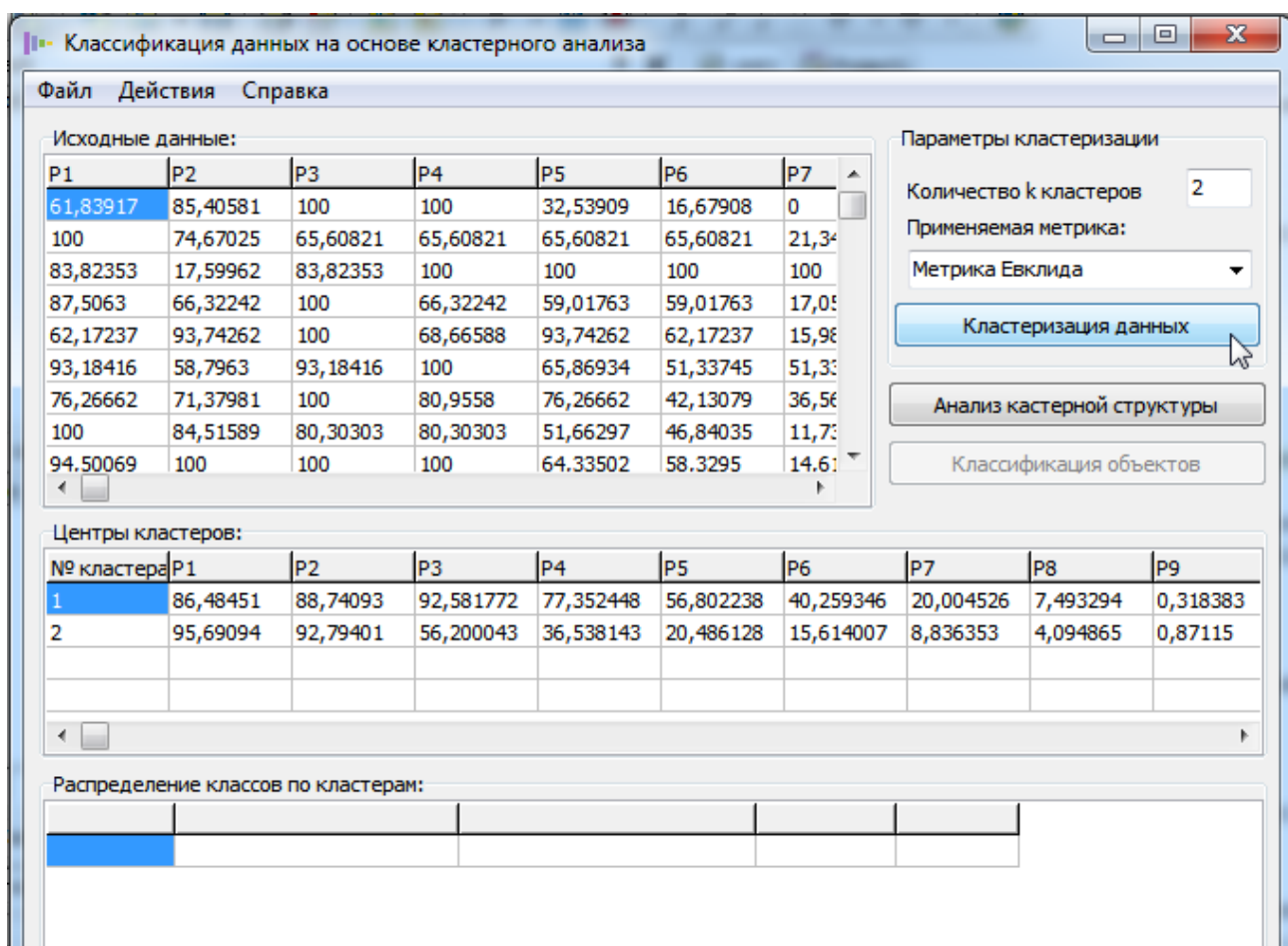


Рисунок 3.5 – Результат кластеризации исходных данных

Следующим этапом построения классификатора является исследование распределение классов по кластерам. При этом подсчитывается количество объектов в кластере, относящихся к каждому из классов. Впоследствии для каждого кластера определяется доминирующий класс.

Для запуска процедуры является исследования распределения классов по кластерам необходимо нажать на кнопку «Анализ кластерной структуры». Результат выполнения данного этапа представлен на рисунке 3.6.

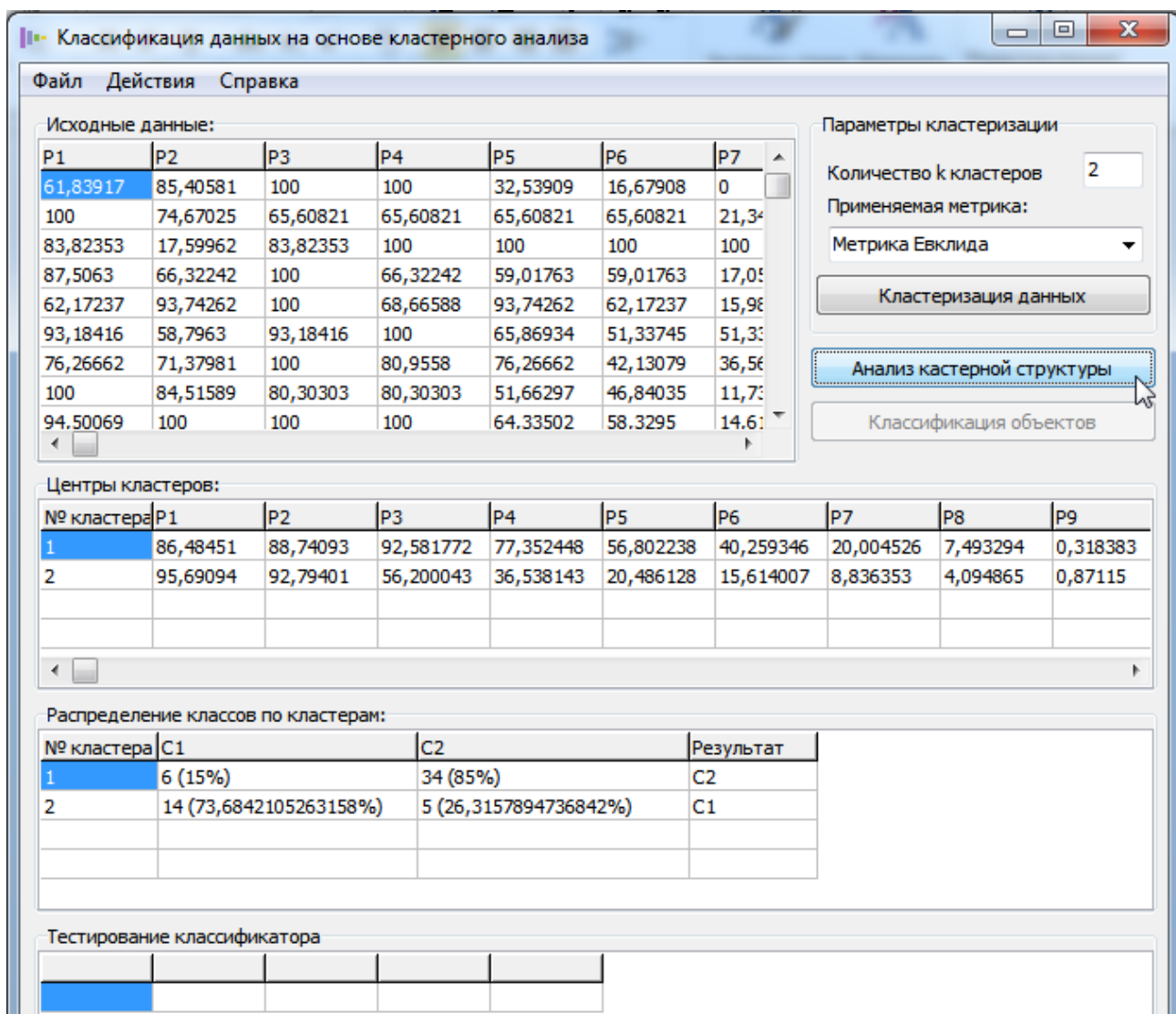


Рисунок 3.6 – Результат оценки соотношения классов внутри каждого кластера

После того, как классификатор построен можно приступить к классификации объектов (классификатор состоит из параметров центров кластеров и метки класса, возвращаемой каждым кластером). Для этого сначала необходимо загрузить в программу выборку объектов, подвергаемую классификации. Это можно сделать, выбрав подпункт «открыть тестовую выборку данных» пункта «файл» (рисунок 3.3).

После этого тестовая выборка данных отобразится в таблице «Тестирование классификатора», расположенной в нижней части главного окна программы (рисунок 3.7).

Центры кластеров:									
№ кластера	P1	P2	P3	P4	P5	P6	P7	P8	P9
1	86,48451	88,74093	92,581772	77,352448	56,802238	40,259346	20,004526	7,493294	0,318383
2	95,69094	92,79401	56,200043	36,538143	20,486128	15,614007	8,836353	4,094865	0,87115

Распределение классов по кластерам:			
№ кластера	C1	C2	Результат
1	6 (15%)	34 (85%)	C2
2	14 (73,6842105263158%)	5 (26,3157894736842%)	C1

Тестирование классификатора									
P1	P2	P3	P4	P5	P6	P7	P8	P9	факт
100	95,29426	95,294258	78,553383	39,559509	28,595501	6,908172	3,526297	0	C2
93,77787	100	86,892197	53,714779	29,794411	24,921141	5,863157	0	0	C1

Аврангшоев Мустафо Мирзомустакимович, ПМИБ-1401

Рисунок 3.7 – Результат загрузки объектов для тестирования классификатора

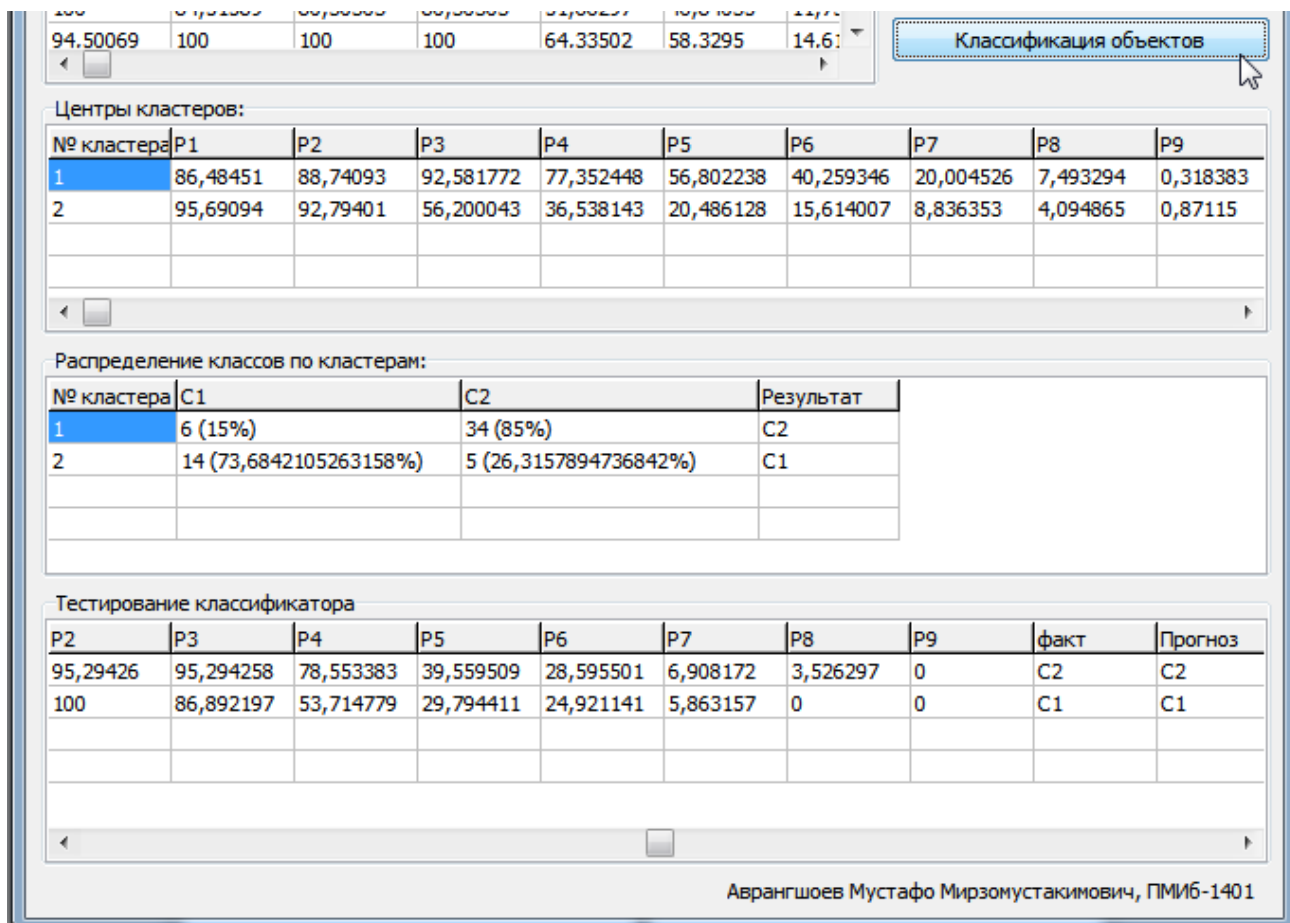


Рисунок 3.8 – Результат классификации объектов

Для запуска классификации объектов необходимо нажать на кнопку «Классификация объектов». При этом каждый объект тестовой выборки будет сравниваться с центром кластера и путем расчета расстояния до выбранной метрики будет определен ближайший центр кластера. Исследуемый объект будет отнесен к тому классу, который доминирует в кластере с ближайшим центром.

Результат классификации объектов будет отображен в столбце «Прогноз» таблицы «Тестирование классификатора».

Результаты классификации объектов тестовой выборки данных можно экспортировать в Excel-файл для дальнейшего исследования точности работы классификатора. Для этого необходимо выбрать подпункт «сохранить результат тестирования» пункта «файл» (рисунок 3.3).

## ЗАКЛЮЧЕНИЕ

В результате анализа выполненных бакалаврской работе исследований были сформулированы следующие выводы:

1. Обзор отечественных и зарубежных литературных источников показал, что современные системы управления и диагностики контактной сварки основаны на использовании алгоритмов машинного обучения. Это позволяет автоматизировать процесс настройки систем, так зависимость между измеряемыми в процессе сварки параметрами и получаемыми прочностными характеристиками восстанавливается алгоритмами машинного обучения самостоятельно (на основе анализа экспериментальных данных).

2. Анализ отечественных и литературных источников показывает, что актуальными являются научные исследования, направлены на расширения возможностей алгоритма кластеризации данных k-means.

3. В ходе выполнения бакалаврской работы разработан алгоритм построения классификатора данных на основе результатов кластерного анализа. Данный алгоритм сначала подбирает кластерную структура для исходных данных, затем проводит статистические исследования каждого кластера для определения преобладающего в нем класса. Затем при классификации объекта определяется его принадлежность к одному из кластеров путем расчёта расстояния от объекта до центра кластеров. Считается, что исследуемый объект относится к тому кластеру, расстояние до центра которого наименьшее. Исследуемому объекту присваивается метка преобладающего в данном кластере класса.

4. Предложенный алгоритм классификации был проверен на данных контактной сварки. В результате был получен классификатор качества сварки, позволяющий на основе измеряемых в процессе сварки параметров (изменение коэффициента мощности  $\cos\varphi$ ) классифицировать качество получаемых соединений (класс С1 – качественная сварка с силой разрушения  $>7200$  Н, класс С2 – некачественная сварка).

5. Опытным путем оценена точность полученного классификатора качества, которая в зависимости от параметра  $k$  составляет 76,27...81,36 % (рисунок 3.1)

6. В ходе выполнения бакалаврской работы было разработано и протестировано программное обеспечение реализующее классификацию данных на основе кластерного анализа (рисунок 3.2).

В ходе выполнения бакалаврской работы было опубликовано две статьи на тему практического применения алгоритмов машинного обучения [19-20].

Основные результаты работы были доложены на IV Международной научно-практической конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук»

## СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Управление тепловыми и металлургическими процессами при контактной сварке с применением систем искусственного интеллекта : отчет о НИР (промежуточ.) / Тольяттинский государственный университет; рук. Климов В. С. – Тольятти, 2016. – 50 с. – Исполн.: Климов А. С., Кудинов А. К. – № ГР 115020230051.
2. Gursoy, A. Data Decomposition for Parallel K-means Clustering [Text] / Attila Gursoy // 5th International Conference on Parallel Processing and Applied Mathematics (PPAM 2003). – Springer-Verlag Berlin Heidelberg, 2003. – pp. 241-248.
3. Płoński, P. Full and Semi-supervised k-Means Clustering Optimised by Class Membership Hesitation [Text] / Piotr Płoński, Krzysztof Zaremba // International Conference on Adaptive and Natural Computing Algorithms (ICANNGA 2013). – Springer-Verlag Berlin Heidelberg, 2013. – pp. 218-225
4. Saúl, L.E. Fast Global k-Means with Similarity Functions Algorithm [Text] / López-Escobar Saúl, J. A. Carrasco-Ochoa, Martínez-Trinidad J. Fco // International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2006). – Springer-Verlag Berlin Heidelberg, 2006. – pp. 512-521
5. Xin, D.J. SOR Based Fuzzy K-Means Clustering Algorithm for Classification of Remotely Sensed Images [Text] / Dong-Jun Xin, Yen-Wei Chen // International Symposium on Neural Networks: Advances in Neural Networks (ISNN 2013). – Springer-Verlag Berlin Heidelberg, 2013. – pp. 375-382
6. Wang, L. A Hybrid Algorithm for Web Document Clustering Based on Frequent Term Sets and k-Means [Text] / Le Wang, Li Tian, Yan Jia, Weihong Han // Advances in Web and Network Technologies, and Information Management. – Springer-Verlag Berlin Heidelberg, 2007. – pp. 198-203
7. Sangita, O. An Improved K-Means Clustering Approach for Teaching Evaluation [Text] / Oswal Sangita, Jagli Dhanamma // International Conference on



Advances in Computing, Communication and Control (ICAC3 2011). – Springer-Verlag Berlin Heidelberg, 2011. – pp. 108-115

8. Li, L. Coherent Sources Mapping by K-Means Cluster and Correlation Coefficient [Text] / Ling Li, Chunguang Li, Yongxiu Lai, Guoling Shi, Dezhong Yao // International Conference on Natural Computation: Advances in Natural Computation (ICNC 2006). – Springer-Verlag Berlin Heidelberg, 2006 – pp. 237-240

9. Ishioka, T. Extended K-means with an Efficient Estimation of the Number of Clusters [Text] / Tsunenori Ishioka // International Conference on Intelligent Data Engineering and Automated Learning: Data Mining, Financial Engineering, and Intelligent Agents (IDEAL 2000). – Springer-Verlag Berlin Heidelberg, 2000 – pp. 17-22

10. Carlos, J. A New Clustering Algorithm Based on K-Means Using a Line Segment as Prototype [Text] / Juan Carlos, Rojas Thomas // Iberoamerican Congress on Pattern Recognition: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (CIARP 2011). – Springer-Verlag Berlin Heidelberg, 2011. – pp. 638-645

11. Wu, J. Advances in K-means Clustering: A Data Mining Thinking [Text] / Junjie Wu. – Springer-Verlag Berlin Heidelberg, 2012. – 178 p.

12. Maulik, U. Multiobjective Genetic Algorithms for Clustering : Applications in Data Mining and Bioinformatics [Text] / Ujjwal Maulik, Sanghamitra Bandyopadhyay, Anirban Mukhopadhyay. – Springer-Verlag Berlin Heidelberg, 2011. – 275 p.

13. Miyamoto, S. Algorithms for Fuzzy Clustering: Methods in c-Means Clustering with Applications [Text] / Sadaaki Miyamoto, Hidetomo Ichihashi, Katsuhiro Honda. – Springer-Verlag Berlin Heidelberg, 2008. – 245 p.

14. Das, S. Metaheuristic Clustering [Text] / Swagatam Das, Ajith Abraham, Amit Konar. – Springer-Verlag Berlin Heidelberg, 2009. – 249 p.

15. Wu, W. Clustering and Information Retrieval [Text] / Weili Wu, Hui Xiong, Shashi Shekhar. – Kluwer Academic Publishers, 2004. – 299 p.

16. Masulli, F. Clustering High – Dimensional Data [Text] / Francesco Masulli, Alfredo Petrosino, Stefano Rovetta. – Springer Heidelberg New York, 2015. – 149 p.

17. Shekhar, S. Encyclopedia of GIS [Text] / Shashi Shekhar, Hui Xiong, Xun Zhou. – Springer International Publishing, 2017. – 392 p.

18. Liu, Y. An Improved Kernel K-means Clustering Algorithm [Text] / Yang Liu, Hong Peng Yin, Yi Chai // Proceedings of 2016 Chinese Intelligent Systems Conference. – Springer Science+Business Media Singapore, 2016. – pp. 275-280

19. Аврангшоев, М.М. Применение алгоритма индуктивного машинного обучения ID3 для диагностики качества сварки [Текст] / С.Р. Никитин, М.М. Аврангшоев // IV Международной научно-практической конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук». – Тольятти: Изд. ТГУ, 2018.

20. Аврангшоев, М.М. Модификация алгоритма для интеллектуальных систем распознавания образов [Текст] / М.М. Аврангшоев, Д.А. Александров // IV Международной научно-практической конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук». – Тольятти: Изд. ТГУ, 2018.