

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)

02.03.03 Математическое обеспечение и администрирование
информационных систем

(код и наименование направления подготовки, специальности)

Технология программирование

(направленность (профиль)/специализация)

БАКАЛАВРСКАЯ РАБОТА

на тему «Аффинитивный анализ при решении задач построения
классификационных моделей»

Студент

Д.А. Александров

(И.О. Фамилия)

(личная подпись)

Руководитель

В.С. Климов

(И.О. Фамилия)

(личная подпись)

Консультанты

А.В. Москалюк

(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: «Аффинитивный анализ при решении задач построения классификационных моделей».

Данная бакалаврская работа направлена на развитие методов машинного обучения. В ней разрабатывается способ построения классификационных моделей по результатам аффинитивного анализа данных. В работе показаны результаты использования предложенных подходов при формировании классификатора качества сварных соединений. Реализация предложенных подходов представлена в виде программного обеспечения.

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

Во введении описывается актуальность проводимого исследования, формулируется цель и ставятся задачи, которые необходимо решить.

В первой главе работы анализируются алгоритмы машинного обучения, применяемые в системах интеллектуальной диагностики сварки.

Во второй главе представлен математический аппарат аффинитивного анализа данных. Приводится описание предложенного алгоритма построения классификатора на основе результатов аффинитивного анализа. Описывается подготовка экспериментальных данных сварки для анализа с использованием предложенного алгоритма.

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

В заключении представлены выводы по проделанной работе.

В работе использовано 6 таблиц, 13 рисунка, список литературы содержит 20 литературных источников. Общий объем выпускной квалификационной работы составляет 42 страниц.

ABSTRACT

The topic of the bachelor's work: "Affine analysis in solving problems of building classification models."

This bachelor's work is aimed at the development of methods of machine learning. It develops a method for constructing classification models based on the results of affinity analysis of data. The work shows the results of using the proposed approaches for the formation of the quality classifier for welded joints. The implementation of the proposed approaches is presented in the form of software.

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

The introduction describes the relevance of the ongoing research, formulates the goal and sets tasks that need to be addressed.

In the first chapter of the work, machine learning algorithms used in intelligent welding diagnostic systems are analyzed.

The second chapter presents the mathematical apparatus of affinity analysis of data. A description of the proposed algorithm for constructing a classifier based on the results of affinity analysis is given. The preparation of experimental welding data for analysis using the proposed algorithm is described.

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 6 tables, 13 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 ПОСТРОЕНИЯ КЛАССИФИКАТОРА НА ОСНОВЕ АФФИНТИВНОГО АНАЛИЗА	11
2.1 Технология аффинитивного анализа.....	11
2.2 Анализ экспериментальных данных	19
2.3 Алгоритм синтеза классификатора на основе данных аффинитивного анализа	22
3 ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ПРЕДЛОЖЕННЫХ РЕШЕНИЙ.....	28
3.1 Апробация алгоритма на данных контактной сварки	28
3.2 Программная реализация предложенных решений.....	32
ЗАКЛЮЧЕНИЕ	38
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ	40

ВВЕДЕНИЕ

Развитие искусственного интеллекта привело к созданию отдельного класса интеллектуальных систем, предназначенных для диагностики технологическими процессами. В бакалаврской работе в качестве технологического процесса рассматривается контактная точечная сварка.

В интеллектуальных системах диагностики технологических процессов применяются такие методы машинного обучения, как нейронные сети различных видов, деревья принятия решений, индуктивные алгоритмы классификации.

Примечательно, что данные методы направлены на решение задачи классификации – восстановления зависимости между измеряемыми параметрами технологического процесса и получаемым классом качества продукции.

При этом существуют алгоритмы поиска ассоциативных правил, одним из которых является Apriori. Недостатком алгоритма Apriori является то, что он не предназначен для решения задач классификации, он решает задачи аффинитивного анализа.

В настоящее время исследованиями на тему совершенствования возможностей алгоритмов аффинитивного анализа данных занимаются такие ученые, как Y. Djenouri, M. Comuzzi, P. Arora, S. Singh, X. Zhao, D. Li, Z. Sun, Y. Chen, B. Liao, N. C. Benhamouda, H. Drias, C. Hirèche, S. Dhanya, M. Vysaakan, A. S. Mahesh, María C. Fernández-Baizán, M. Ruiz, J. F. Martínez Sarrías, S. Millán, X. Liang, C. Xue, M. Huang, N. K. Sharma, N. K. Nagwani, J. Agarwal [2-18] и др.

Для преодоления описанной проблемы алгоритма Apriori в бакалаврской работе разрабатывается математическая модель и алгоритм построения классификаторов на основе аффинитивного анализа экспериментальных данных сварки. Построенный классификатор позволяет классифицировать качество сварки на основе измеренных в процессе сварки параметров.

В исследовании проверяется гипотеза о возможности использования аффинитивного анализа для построения классификатора качества сварки.

Целью работы является разработка алгоритма построения классификаторов с использованием аффинитивного анализа на примере экспериментальных данных контактной сварки.

В ходе выполнения бакалаврской работы было опубликовано четыре статьи на тему практического применения алгоритмов машинного обучения [19-22].

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».

Бакалаврская работа «Аффинитивный анализ при решении задач построения классификационных моделей», направленна на разработку математической и алгоритмической модели, поэтому данная тема соответствует требованиям Федерального государственного образовательного стандарта высшего образования по направлению подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем.

1 АНАЛИЗ СОСТОЯНИЯ ВОПРОСА

Классический подход синтеза систем управления и диагностики предполагает на первом этапе сформулировать математическую модель объекта управления или диагностики. Математическая модель объекта нужна для описания связи между измеряемыми параметрами и получаемыми характеристиками.

Если формулирование математической модели объекта (с заданной точностью) не требует значительных временных затрат, то разработка алгоритма диагностики или управления таким объектом является тривиальной задачей.

При этом существуют такие объекты управления и диагностики для которых формулирование адекватной математической модели с учетом всех протыкаемых физических процессов является нерешенной задачей. К таким объектам относится процесс контактной точечной сварки. Это связано с тем, что математическая модель контактной сварки должна включать в себя описание электрических процессов протекающих в контурах сварочного оборудования, металлургические процессов, протекающих в процессе нагрева и плавления металла, находящегося между электродами, процесс кристаллизации металла при формировании литого ядра и т.д.

При синтезе интеллектуальных систем управления и диагностики формулирование математической модели объекта не требуется. Это связано с тем, что применяемые в них индуктивные и дедуктивные методы машинного обучения, позволяют самостоятельно восстанавливать зависимость между измеряемыми параметрами и получаемыми характеристиками.

При использовании в своем составе индуктивных методов машинного обучения для настройки интеллектуальных систем требует сначала накопить экспериментальные данные, которые впоследствии будут объединены в обучающую выборку. За тем на основе анализа данных обучающей выборки будут определены параметры используемой модели при которых достигается

требуемая точность работы системы. К индуктивному машинному обучению относятся нейронные сети различных видов, классификационные и регрессионные деревья принятия решений, индуктивные алгоритмы классификации (kNN, SVM и др.)

При использовании в своем составе дедуктивных методов машинного обучения для настройки системы требуется опросить эксперта из рассматриваемой предметной области, для того чтобы понять как те или иные факторы влияют на результат выполнения технологического процесса. К дедуктивному машинному обучению относятся системы, основанные на нечеткой логике, системы, основанные на рекурсивно-логическом программировании.

Задача интеллектуальной диагностики контактной сварки многими исследователями рассматривается, как задача классификации [1]:

- R.V. Chinnam, D. Filev, M. El-Banna предложили на основе измеренных значений сварочного тока и напряжения строить график изменения сопротивления на участке между электродами. Затем на основе данных графика с помощью нейронной сети LVQ классифицировать качество сварного соединения.

- V. Balendran, A. Aravinthan, D. Al-Dabass, K. Sivayoganathan предложили на основе измеренных значений сварочного тока и напряжения рассчитывать динамику изменения активного сопротивления сварочного участка. Затем на основе анализа графика изменения активного сопротивления с помощью нейронных сетей прямого распространения прогнозировать прочность сварной точки. После, в зависимости от рассчитанного значения прочности относить сварное соединение к классу качественных сварок или некачественный.

- J. Röning, P. Laurinen, L. Tuovinen, H. Junno.предложили производить измерение осциллограмм сварочного тока и напряжения. Затем на основе анализа характеристик осциллограмм с помощью Байесовских сетей доверия классифицировать качество сварки.

- В.С. Климов, А.С. Климов, С.А. Петрушкин, А.Д. Шижгин предложили на основе данных об изменении коэффициента мощности сварки с помощью деревьев принятия решений прогнозировать качество сварного соединения.

Таким образом видно, что развитие методов машинного обучения с точки зрения диагностики и правления технологическими процессами является актуальной задачей.

Существуют алгоритмы поиска ассоциативных правил, одним из которых является Apriori. Недостатком алгоритма Apriori является то, что он не предназначен для решения задач классификации, он решает задачи аффинитивного анализа.

В настоящее время исследованиями на тему совершенствования возможностей алгоритмов аффинитивного анализа данных занимаются такие ученые, как Y. Djenouri, M. Comuzzi, P. Arora, S. Singh, X. Zhao, D. Li, Z. Sun, Y. Chen, B. Liao, N. C. Benhamouda, H. Drias, C. Hirèche, S. Dhanya, M. Vysaakan, A. S. Mahesh, María C. Fernández-Baizán, M. Ruiz, J. F. Martínez Sarrías, S. Millán, X. Liang, C. Xue, M. Huang, N. K. Sharma, N. K. Nagwani, J. Agarwal [2-18] и др.

Для преодоления описанной проблемы алгоритма Apriori в бакалаврской работе разрабатывается математическая модель и алгоритм построения классификаторов на основе аффинитивного анализа экспериментальных данных сварки. Построенный классификатор позволяет классифицировать качество сварки на основе измеренных в процессе сварки параметров.

Целью работы является разработка алгоритма построения классификаторов с использованием аффинитивного анализа на примере экспериментальных данных контактной сварки.

Поставленная цель достигается за счет последовательного решения следующих задач:

1. Анализ состояния вопроса об использовании алгоритмов машинного обучения в системах диагностики сварки.

2. Разработка алгоритма синтеза классификатора на основе результатов аффинитивного анализа экспериментальных данных.

3. Тестирование предложенных подходов на реальных данных, а также проектирование, разработка, апробация программного обеспечения по реализации разработанного алгоритма.

2 РАЗРАБОТКА АЛГОРИТМА ПОСТРОЕНИЯ КЛАССИФИКАТОРА НА ОСНОВЕ АФФИНТИВНОГО АНАЛИЗА

2.1 Технология аффинитивного анализа

Аффинитивный анализ (affinity analysis) — набор методов интеллектуального анализа данных, направленный на поиск и исследование взаимной связи (ассоциаций) между событиями, происходящими совместно, и количественную оценку таких связей.

Название такого подхода происходит от английского слова *affinity*, означающее в переводе «близость», «сходство». Целью анализа является обнаружение ассоциации между различными событиями, то есть найти правила для количественного описания взаимной связи между двумя или более событиями. Такие правила называются ассоциативными правилами (*association rules*).

Исходными данными для проведения аффинитивного анализа является набор транзакций T , состоящий из элементарных транзакций t :

$$T = \{t_1, t_2, \dots, t_m\}, \quad (2.1)$$

где m – количество транзакций.

В каждой транзакции содержится набор событий, происходящих одновременно. Набор всех событий задан множеством I :

$$I = \{i_1, i_2, \dots, i_n\} \quad (2.2)$$

где, n – количество элементарных событий.

T	Транзакции
t_1	$\{i_3, i_9, i_{10}, i_{13}\}$
t_2	$\{i_1, i_3, i_{10}\}$
t_3	$\{i_2, i_9\}$
...	...
t_m	$\{i_3, i_9, \dots, i_{10}, i_{13}\}$

Рисунок 2.1 – Исходные данные для аффинитивного анализа

На множестве транзакций T , с использованием множества элементарных событий I можно сформировать ассоциативные правила. Ассоциативное правило – это импликация (бинарная логическая связка) вида:

$$\begin{aligned} X &\Rightarrow Y \\ X \subset I, Y \subset I, X \cap Y &= \emptyset \end{aligned} \quad (2.2)$$

где, I – множество всех событий, X – множество событий, называемых условием (antecedent), Y – множества событий, называемых следствием (consequent).

Ассоциативные правила описывают связь между наборами предметов, соответствующими условию и следствию. Эта связь характеризуется такими показателями (метриками), как поддержка *supp*, достоверность *conf*, лифт *lift*, леввередж *levr*.

Поддержка *supp* ассоциативного правила — это отношение числа транзакций, которые содержат как условие, так и следствие к общему количеству транзакций.

$$supp(X \Rightarrow Y) = \frac{|\{t \in T; (X \cup Y) \subseteq t\}|}{|T|}, \quad (2.3)$$

где T – множество транзакций, t – транзакция.

Для любого предметного набора A также может быть рассчитана поддержка следующим образом:

$$supp(A) = \frac{|\{t \in T; A \subseteq t\}|}{|T|} \quad (2.4)$$

Достоверность *conf* ассоциативного правила представляет собой меру точности правила и определяется как отношение количества транзакций, содержащих и условие, и следствие, к количеству транзакций, содержащих только условие:

$$conf(X \Rightarrow Y) = \frac{supp(X \Rightarrow Y)}{supp(X)} \quad (2.5)$$

Лифт *lift* — это отношение частоты появления условия в транзакциях, которые также содержат и следствие, к частоте появления следствия в целом:

$$lift(X \Rightarrow Y) = \frac{supp(X \Rightarrow Y)}{supp(X) \times supp(Y)} \quad (2.6)$$

Значения лифта большие, чем единица, показывают, что условие чаще появляется в транзакциях, содержащих следствие, чем в остальных. Можно сказать, что лифт является обобщенной мерой связи двух предметных наборов: при значениях лифта > 1 связь положительная, при 1 она отсутствует, а при значениях < 1 — отрицательная.

Левередж *levr* — это разность между наблюдаемой частотой, с которой условие и следствие появляются совместно (то есть поддержкой ассоциации), и произведением частот появления (поддержек) условия и следствия по отдельности:

$$levr(X \Rightarrow Y) = supp(X \Rightarrow Y) - supp(X) \cdot supp(Y) \quad (2.7)$$

Чем выше значение *lev*, тем сильнее (значимей) ассоциативное правило.

Анализ ассоциативных правил используется при решении задач из различных областей, например:

- выявление наборов товаров, которые в супермаркетах часто покупаются вместе или никогда не покупаются вместе;
- определение доли клиентов, положительно относящихся к нововведениям в их обслуживании;
- определение профиля посетителей веб-ресурса;
- определение доли случаев, в которых новое лекарство показывает опасный побочный эффект.

Методом перебора на основе множества транзакций *T* получать все возможные ассоциативные правила не целесообразно по следующим причинам:

1. Перебирая все возможные сочетания элементарных событий в условии и следствии ассоциативного можно получать неоправданно большое количество ассоциативных правил.

2. Ценностью обладают значимые ассоциативные правила с высокими показателями поддержки *supp*, достоверности *conf*, лифта *lift*, леввереджа *levr*, которые составляют малую долю от общего числа всех вариантов ассоциативных правил.

Чтобы сократить пространство поиска значимых ассоциативных правил на основе набора транзакций T применяется алгоритм Apriori.

В основе алгоритма лежит правило антимонотонности, которое утверждает, если предметный набор Z не является частым, то добавление некоторого нового предмета A к набору Z не делает его более частым. Другими словами, если Z не является частым набором, то и набор $Z \cup A$ также не будет являться таковым. Данное свойство значительно уменьшает пространство поиска ассоциативных правил.

Использование алгоритма Apriori для поиска значимых ассоциативных правил на основе множества транзакций состоит из 2 шагов:

1. Поиск частых предметных наборов.
2. Генерирование ассоциативных правил на основе частых предметных наборов.
3. Оценка сгенерированных ассоциативных правил с использованием описанных выше метрик.
4. Ранжирование правил по значимости с учетом их количественных показателей.

Псевдо код алгоритма Apriori для поиска частых предметных наборов представлен ниже. Входными параметрами являются множество транзакций T и порог поддержки e . Алгоритм возвращает массив частых предметных наборов (2-предметных наборов L_2 , частых 3-предметных наборов L_3 и т.д.):

Apriori(T, e)

$L_1 \leftarrow \{l \text{ arg } e \text{ 1-itemsets}\}$

$k \leftarrow 2$

while $L_{k-1} \neq \emptyset$

$C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a\} - \{c \mid \{s \mid s \subseteq c \wedge |s|=k-1\} \not\subseteq L_{k-1}\}$

for transactions $t \in T$

$C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}$

for candidates $c \in C_t$

$count[c] \leftarrow count[c] + 1$

$L_k \leftarrow \{c \mid c \in C_k \wedge count[c] \geq e\}$

$k \leftarrow k + 1$

return $\bigcup_k L_k$

Таким образом, поиск частых предметных наборов происходит следующим образом. В массив 1-предметных наборов L_1 попадут все события, для которых $supp \geq e$.

Массив 2-предметных наборов L_2 формируется следующим образом. Путем перебора всех сочетаний элементов набора L_1 формируются пары предметных наборов. Каждая полученная пара проверяется на условие $supp \geq e$. Пары, для которых не выполняется данное условие, исключаются из массива L_2 .

Массивы 3-предметных наборов и более (4-предметных, 5-предметных и т.д.) формируются следующим образом. Из элементов предметного набора находящихся на одну ступень ниже формируются пары. Пары можно сформировать из тех элементов набора, у которых все элементарные события кроме последнего идентичны. Объединив элементарные события этих пар можно получить предметный набор для текущей ступени. Получившийся предметный набор для включения его в массив L_k должен соответствовать условию $supp \geq e$.

Пример формирования частых предметных наборов в соответствии с алгоритмом *Apriori* представлен в таблице 2.1.

Таблица 2.1 – Пример формирование частый предметных наборов в соответствии с алгоритмом Apriori

Номер итерации k выполнение алгоритма				
$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Частые предметные наборы, сформированные на каждой итерации				
$L_1 = \begin{Bmatrix} i_2 \\ i_5 \\ i_6 \\ i_{11} \end{Bmatrix}$	$L_2 = \begin{Bmatrix} i_2, i_5 \\ i_2, i_6 \\ i_2, i_{11} \\ i_5, i_6 \\ i_5, i_{11} \end{Bmatrix}$	$L_3 = \begin{Bmatrix} i_2, i_5, i_6 \\ i_2, i_5, i_{11} \\ i_5, i_6, i_{11} \end{Bmatrix}$	$L_4 = \{i_2, i_5, i_6, i_{11}\}$	$L_5 = \emptyset$
Условие для включения предметного набора в множество L_k				
$supp(L_{1,i}) \geq e$	$supp(L_{2,i}) \geq e$	$supp(L_{3,i}) \geq e$	$supp(L_{4,i}) \geq e$	-
Включение в массив $D = \bigcup_k L_k$				
Нет	Да	Да	Да	Нет

На основе всех найденных частых предметных наборов D генерируются ассоциативные правила.

$$D = \bigcup_k L_k \quad (2.8)$$

Для этого к каждому частому предметному набору S из массива D , применяется следующий алгоритм:

1) На основе набора элементарных событий S генерируются все возможные поднаборы $SubS$.

2) Если поднабор $SubS$ является непустым поднабором S , то в множество ассоциативных правил R добавляется правило вида:

$$SubS \Rightarrow (S \setminus SubS), \quad (2.9)$$

где $S \setminus SubS$ представляет собой набор S без поднабора $SubS$.

Данная процедура повторяется для каждого подмножества $SubS$ из S .

После выполнение данной процедуры на всем множестве D будет получен набор ассоциативных правил R вида:

$$R = \left\{ \begin{array}{l} X_1 \Rightarrow Y_1 \\ X_2 \Rightarrow Y_2 \\ \dots \\ X_q \Rightarrow Y_q \end{array} \right\}, \quad (2.10)$$

где q – количество найденных ассоциативных правил. Стоит отметить, что всегда $|R| > |D|$. Пример генерации ассоциативных правил представлен в таблице 2.2.

Таблица 2.2 – Генерации правил на основе данных таблицы 2.1

Частые предметные наборы (множество D)	Ассоциативные правила (множество R)
$L_2 = \left\{ \begin{array}{l} i_2, i_5 \\ i_2, i_6 \\ i_2, i_{11} \\ i_5, i_6 \\ i_5, i_{11} \end{array} \right\}$	$i_2 \Rightarrow i_5 ; i_5 \Rightarrow i_2 ; i_2 \Rightarrow i_6 ;$ $i_6 \Rightarrow i_2 ; i_2 \Rightarrow i_{11} ; i_{11} \Rightarrow i_2 ; i_5 \Rightarrow i_6 ;$ $i_6 \Rightarrow i_5 ; i_5 \Rightarrow i_{11} ; i_{11} \Rightarrow i_5 .$
$L_3 = \left\{ \begin{array}{l} i_2, i_5, i_6 \\ i_2, i_5, i_{11} \\ i_5, i_6, i_{11} \end{array} \right\}$	$i_2 \Rightarrow i_5, i_6 ; i_5 \Rightarrow i_2, i_6 ; i_6 \Rightarrow i_2, i_5 ;$ $i_5, i_6 \Rightarrow i_2 ; i_2, i_6 \Rightarrow i_5 ; i_2, i_5 \Rightarrow i_6 ;$ $i_2 \Rightarrow i_5, i_{11} ; i_5 \Rightarrow i_2, i_{11} ; i_{11} \Rightarrow i_5, i_2 ;$ $i_5, i_{11} \Rightarrow i_2 ; i_2, i_{11} \Rightarrow i_5 ; i_5, i_2 \Rightarrow i_{11} ;$ $i_5 \Rightarrow i_6, i_{11} ; i_6 \Rightarrow i_5, i_{11} ; i_{11} \Rightarrow i_6, i_5 ;$ $i_6, i_{11} \Rightarrow i_5 ; i_5, i_{11} \Rightarrow i_6 ; i_6, i_5 \Rightarrow i_{11} ;$
$L_4 = \{i_2, i_5, i_6, i_{11}\}$	$i_2 \Rightarrow i_5, i_6, i_{11} ; i_5 \Rightarrow i_2, i_6, i_{11} ; i_6 \Rightarrow i_5, i_2, i_{11} ;$ $i_{11} \Rightarrow i_5, i_6, i_2 ; i_2, i_5 \Rightarrow i_6, i_{11} ; i_2, i_6 \Rightarrow i_5, i_{11} ;$ $i_2, i_{11} \Rightarrow i_6, i_5 ; i_5, i_6 \Rightarrow i_2, i_{11} ; i_5, i_{11} \Rightarrow i_2, i_6 ;$ $i_6, i_{11} \Rightarrow i_2, i_5 ; i_5, i_6, i_{11} \Rightarrow i_2 ; i_2, i_6, i_{11} \Rightarrow i_5 ;$

	$i_5, i_2, i_{11} \Rightarrow i_6 ; i_5, i_6, i_2 \Rightarrow i_{11} .$
--	---

После того, как найдено множество ассоциативных правил R , его исследуют с использованием показателей поддержки $supp$, достоверности $conf$, лифта $lift$, леввереджа lev .

Чаще всего данное исследование заключается в составлении рейтинговой таблицы ассоциативных правил путем их сортировки в порядке убывания значимости. Для этого обычно используют обобщенный показатель, рассчитываемый как произведение поддержки правила $supp(r_i)$ на достоверность $conf(r_i)$.

Графическая интерпретация этапа составления рейтинговой таблицы ассоциативных правил представлена на рисунке 2.2.



Рисунок 2.2 – Сортировка правил по показателю $supp(r_i) \cdot conf(r_i)$

2.2 Анализ экспериментальных данных

Исходными данными для построения классификатора с использованием методов индуктивного машинного обучения является обучающая выборка. В нашем случае обучающей выборкой является экспериментальные данные контактной сварки, полученные в лабораторных условиях.

При выполнении экспериментальной сварки использовались пластины из низкоуглеродистой стали, располагаемые внахлест (друг над другом). Толщина каждой пластины составляла 0,8 мм, длительность сварки – 9 периодов сетевого напряжения (0,18 с).

Экспериментальная сварка проводилась:

- в условия отсутствия действия возмущений при рекомендуемых режимах сварки;
- в условия отсутствия действия возмущений при заниженных и завышенных параметрах режима сварки;
- в условиях действия возмущений процесса сварки в различных сочетаниях.

Такой спектр экспериментальных сварок позволил получить набор сварных соединений с различными прочностными характеристиками.

В процессе каждой сварки производилось измерение коэффициента мощности $\cos\varphi$ для каждого периода сетевого напряжения. В нашем случае каждую сварку характеризуют 9 значений: $\cos(\varphi_1), \cos(\varphi_2), \dots, \cos(\varphi_9)$.

Известно, что изменение коэффициента мощности $\cos\varphi$ в процессе сварки связано с металлургическими процессами, протекающими в зоне сварки (нагрев, размягчение и плавление металла). При этом характер протекания металлургических процессов напрямую связан с получаемыми прочностными характеристиками соединения.

Все полученные в результате экспериментальной сварки исследовались разрушающими методами контроля для оценки их качества. Критерием

качества сварки для нашей конфигурации сварного соединения является сила разрушения при испытаниях на срез не менее 7000 Н. Были выбраны следующие маркеры классов S : $S1$ – качественная сварка, $S2$ – некачественная сварка.

Диагностической ценностью измеряемых в процессе сварки параметров представляют не абсолютные значения $\cos(\varphi)$, а динамика (форма графика) изменения коэффициента мощности в процессе сварки. Поэтому при подготовке данных для анализа была проведена нормировка значения $\cos(\varphi_i)$ в рамках каждой отдельно взятой сварки путем линейного преобразования к диапазону $[0;100]$.

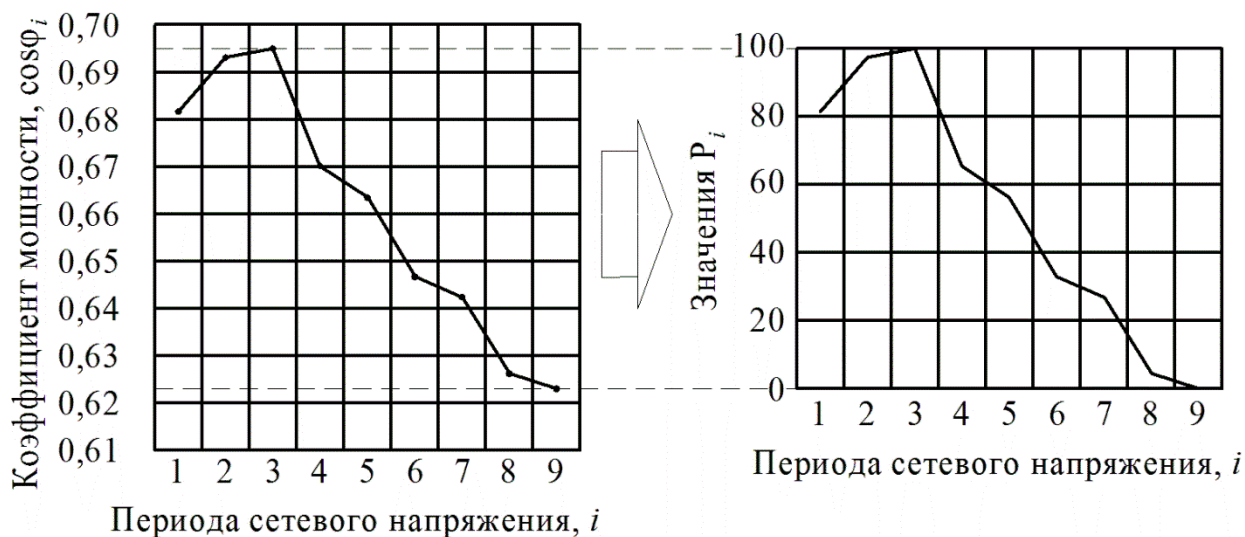


Рисунок 2.3 – Нормирования значений $\cos\varphi_i$ в масштабах одной сварки к диапазону $[0;100]$: P_i – нормированные значения

Описание параметров, содержащихся в обучающей выборке представлен в таблице 2.3.

Таблица 2.3 – Описание параметров обучающей выборки

Обозначение	Область определения	Описание	Тип параметра
P_1	[0; 100]	Нормирование значение коэффициента мощности в первом периоде	Входной параметр
P_2	[0; 100]	Нормирование значение коэффициента мощности в втором периоде	Входной параметр
P_3	[0; 100]	Нормирование значение коэффициента мощности в третьем периоде	Входной параметр
P_4	[0; 100]	Нормирование значение коэффициента мощности в четвертом периоде	Входной параметр
P_5	[0; 100]	Нормирование значение коэффициента мощности в пятом периоде	Входной параметр
P_6	[0; 100]	Нормирование значение коэффициента мощности в шестом периоде	Входной параметр
P_7	[0; 100]	Нормирование значение коэффициента мощности в седьмом периоде	Входной параметр
P_8	[0; 100]	Нормирование значение коэффициента мощности в восьмом периоде	Входной параметр
P_9	[0; 100]	Нормирование значение коэффициента мощности в девятом периоде	Входной параметр
C	{C1,C2}	Метка класса качества: C1 – качественная сварка (сила разрушения ≥ 7000 Н); C2 – некачественная сварка (сила разрушения < 7000 Н)	Выходной параметр

Результат подготовки данных экспериментальных сварок к анализу представлен в таблице 2.4.

Таблица 2.4 – Фрагмент экспериментальных данных, после нормировки значений $\cos \varphi_i$

№	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	$F_{пр}, Н$
1	61,83	85,40	100	100	32,53	16,67	0	0	0	C2
2	100	74,67	65,60	65,60	65,60	65,60	21,34	10,94	0	C2
3	83,82	17,59	83,82	100	100	100	100	17,59	0	C2
4	87,50	66,32	100	66,32	59,01	59,01	17,05	8,74	0	C2
5	62,17	93,74	100	68,66	93,74	62,17	15,98	8,19	0	C2
6	93,18	58,79	93,18	100	65,86	51,33	51,33	8,48	0	C2
7	76,26	71,37	100	80,95	76,26	42,13	36,56	30,72	0	C2
8	100	84,51	80,30	80,30	51,66	46,84	11,73	6,00	0	C2
9	94,50	100	100	100	64,33	58,32	14,61	7,47	0	C2
10	74,75	94,61	100	100	63,84	58,03	14,13	7,23	0	C2
11	88,99	100	100	100	63,84	20,71	7,23	0	0	C2
12	94,85	83,99	100	83,99	62,35	36,40	29,22	4,33	0	C2
13	100	100	100	75,18	45,80	36,94	5,30	0	0	C1
14	80,55	97,20	100	60,78	50,59	25,79	21,64	0	12,73	C2
15	93,74	100	100	100	75,74	47,47	13,83	4,81	0	C1
16	93,19	100	82,19	79,84	58,92	37,22	11,27	3,92	0	C1
17	84,35	100	88,481	70,73	49,77	27,44	21,11	17,73	0	C1
...					...					
100	100	76,61	33,39	36,36	22,71	19,49	5,624	1,94	0	C1

Теперь, когда определен формат обучающей выборки, можно переходить к адаптации технологии аффинитивного анализа под решаемую задачу классификации качества сварки.

2.3 Алгоритм синтеза классификатора на основе данных аффинитивного анализа

Разработанный алгоритм синтеза классификатора качества сварки, состоит из следующих этапов:

1. На основе анализа меток класса и областей определения входных параметров формируется множество элементарных событий.
2. Преобразование каждого элемента обучающей выборки в отдельную транзакцию.

3. Поиск с помощью алгоритма Apriori ассоциативных правил, удовлетворяющих минимальным значениям поддержки *Supp* и достоверности *Conf*.

4. Отбор из полученных ассоциативных правил тех, у которых в следствии присутствует метка класса.

5. Объединение отобранных правил в классификатор.

Рассмотрим каждый этап алгоритма подробнее. На первом этапе необходимо выполнить проанализировать выборку экспериментальных данных для формирования множество возможных элементарных событий. Для этого предложено область определения каждого входного параметра P_1, P_2, \dots, P_9 разбить на отрезки одинаковой длины. Тогда элементарным событием будет являться факт попадания определенного значения рассматриваемого параметра в определенную область.

Количество отрезков, на которые разбивается область определения входных параметров, следует выбирать опытным путем. В нашем случае приемлемым по точности работы (конечного классификатора) является разбиение области определения каждого входного параметра на 10 равных частей.

Возможным значениям выходного параметра (меткам класса) также необходимо сопоставить элементарные события. Для простоты восприятия событию связанному с получением определенного класса (качества сварки) будет дано такое же обозначение, как у метки класса.

Таким образом, на данном этапе формируется множество элементарных событий и формулируются условия их активации.

Графически выполнение первого этапа представлено на рисунке 2.4.

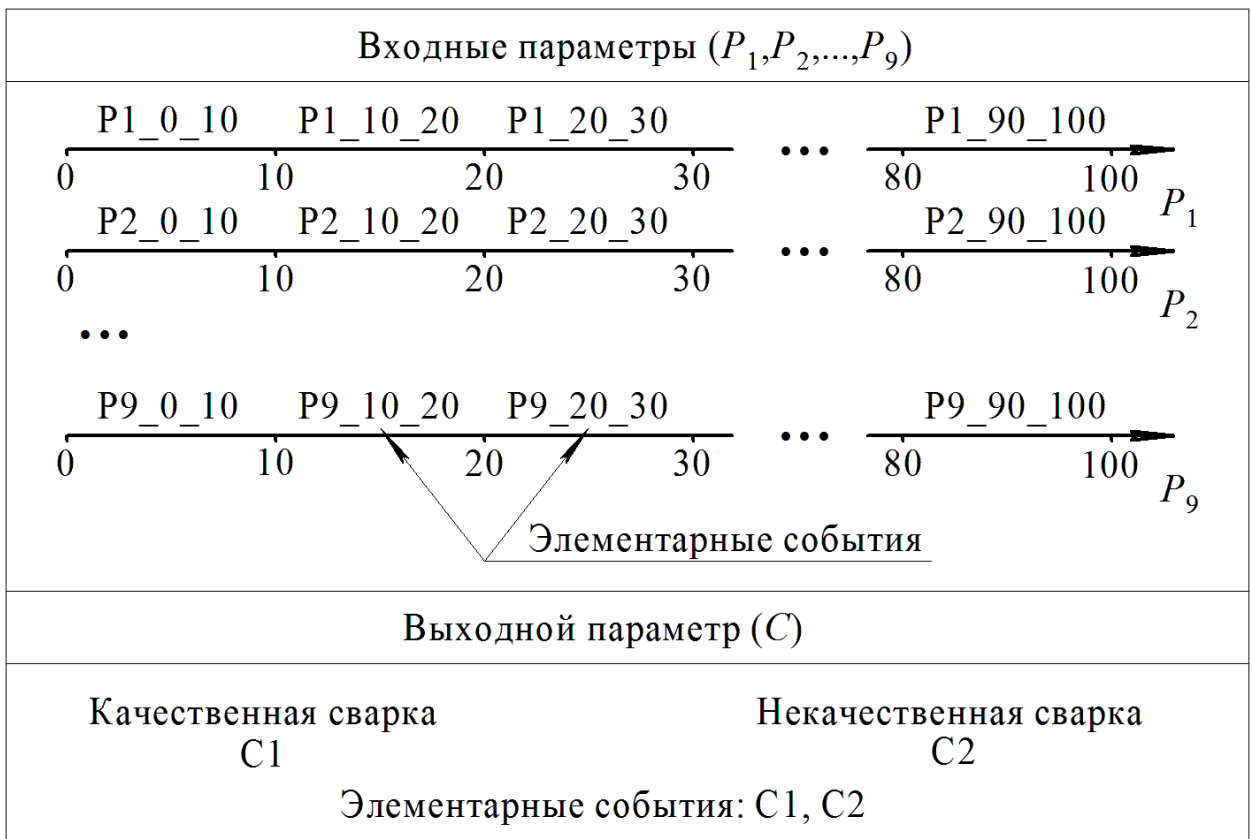


Рисунок 2.4 – Формирование множества элементарных событий

На втором этапе алгоритма осуществляется преобразование каждого элемента обучающей выборки (с учетом сформулированного множества событий) в отдельную транзакцию. Т.е. результатом выполнения данного этапа будет являться множество транзакций, подходящих для выполнения аффинитивного анализа. Количество транзакций равно количеству элементов выборки экспериментальных данных сварки.

Пример формирования транзакции на основе элемента выборки экспериментальных данных представлен на рисунке 2.5.

На третьем этапе с использованием алгоритма Apriori осуществляется поиск ассоциативных правил, удовлетворяющих минимальным значениям поддержки *Supp* и достоверности *Conf*.

Выборка экспериментальных данных контактной сварки

№	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	C
...										
3	93,18	58,79	93,18	100	65,87	51,34	51,34	8,49	0	C1
...										



Множество транзакций

T	Транзакции
...	
t_3	{P1_90_100, P2_50_60, P3_90_100, P4_90_100, P5_60_70, P6_50_60, P7_50_60, P8_0_10, P9_0_10, C1}
...	

Рисунок 2.5 – Пример формирования транзакции на примера элемента выборки экспериментальных данных контактной сварки

Найденные ассоциативные правила показывают зависимость возникновения одних событий от других. Полученный на третьем этапе набор ассоциативных правил сам по себе не является классификатором. Из полученного набора правил необходимо вычлениить только те правила, в следствии которых есть события связанные с метками классов (C1, C2).

Поэтому на четвертом этапе работы алгоритма осуществляется отбор из полученных ассоциативных правил тех, у которых в следствии присутствует метка класса.

Те ассоциативные правила, которые не содержат в следствии события связанного с меткой класса исключаются. Если ассоциативные правила в следствии содержат помимо событий связанных с меткой класса еще и события связанные со значениями входных параметров, то у таких правил модифицируется следствие. Остальные правила оставляются без изменений.

Примеры обработки найденных ассоциативных правил показаны на рисунке 2.6.

Пример ассоциативного правила	Результат
$\{p2=P2_90_100, p7=P7_0_10\} \Rightarrow \{p8=P8_0_10\}$	исключение правила, т.к. в следствии отсутствует указание на метку класса
$\{p2=P2_90_100, p6=P6_20_30\} \Rightarrow \{p8=P8_0_10, C=C1\}$	модифицирование правила $\{p2=P2_90_100, p6=P6_20_30\} \Rightarrow \{C=C1\}$
$\{p5=P5_60_70\} \Rightarrow \{C=C2\}$	сохранение без изменений $\{p5=P5_60_70\} \Rightarrow \{C=C2\}$

Рисунок 2.6 – Примеры обработки ассоциативных правил

После такой обработки все ассоциативные правила в следствии содержат единственное событие – ссылку на метку класса. Это дает на возможность объединить такие правила в классификатор. На пятом этапе работы полученные правило объединяются в классификатор.

В нашем случае классификатор состоит из 27 правил, представленных в таблице 3.1.

Полученный классификатор работает следующим образом:

1. На вход классификатора подается вектор входных параметров **P** исследуемого объекта (данные одной сварки).

2. Затем для каждой метки класса (сначала для C1 – «качественная сварка», затем для C2 – «некачественная сварка») рассчитывается количество сработавших ассоциативных правил. Правило срабатывает, если имеют место быть те события, которые находятся в условии ассоциативного правила.

3. Исследуемому событию присевается та метка класса, для которой сработало наибольшее количество правил.

4. В тех исключительных ситуациях, когда количество сработавших правил, указывающих на различные метки класса одинаково, исследуемому объекту присевается метка исходя из принципа ожидания наихудшего исхода (некачественная сварка).

Графическая интерпретация алгоритма классификации объектов представлена на рисунке 2.7.

1. Подача вектора входных значений

P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	C
95,51	100	92,13	75,18	45,81	36,95	5,3	5,3	0	?

2. Превод числовых значений в множество событий
 $\{P1_90_100; P2_90_100; P3_90_100; P4_70_80; P5_40_50; P6_30_40; P7_0_10; P8_0_10; P9_0_10\}$

3. Подсчет сработавших правил классификатора для каждого класса (всего в классификаторе 27 правил)

C1	C2
$\{P2_90_100\} \Rightarrow \{C1\}$	$\{P3_90_100\} \Rightarrow \{C2\}$
$\{P7_0_10\} \Rightarrow \{C1\}$	$\{P2_90_100\} \Rightarrow \{C2\}$
$\{P5_40_50\} \Rightarrow \{C1\}$	$\{P2_90_100, P3_90_100\} \Rightarrow \{C2\}$
$\{P2_90_100, P5_40_50\} \Rightarrow \{C1\}$	
4 ссылки на {C1}	3 ссылки на {C2}

4. Вывод о классе исследуемого объекта

P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	C
95,51	100	92,13	75,18	45,81	36,95	5,3	5,3	0	C1

Рисунок 2.7 – Алгоритм классификации объектов

Таким образом, был разработан алгоритм построения классификатора на основе аффинитивного анализа экспериментальных данных. Показан принцип работы классификатора на реальных данных

3 ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ПРЕДЛОЖЕННЫХ РЕШЕНИЙ

3.1 Апробация алгоритма на данных контактной сварки

Апробация алгоритма синтеза классификатора (на основе результатов аффинитивного анализа) проводилась на данных процесса контактной сварки. Экспериментальные данные процесса контактной сварки представлены в таблице 2.4. Для построения классификатора качества использовались данные 100 сварок. Количество классов 2: С1 – качественная сварка, С2 – некачественная сварка.

В полученном с помощью разработанного алгоритма классификаторе содержится 27 правил, представленных в таблице 3.1.

Таблица 3.1 – Сгенерированные с помощью алгоритма правила классификации

№	Под- держка supp	Досто- верность conf	Лифт lift	Левередж levr	Условие	Следствие
1	0,46	0,64	1,19	0,07	p2=P2_90_100	C=C1
2	0,34	0,74	1,62	0,13	p3=P3_90_100	C=C2
3	0,29	0,74	1,36	0,08	p7=P7_0_10	C=C1
4	0,25	0,36	0,78	-0,07	p2=P2_90_100	C=C2
5	0,24	0,78	1,43	0,07	p2=P2_90_100, p7=P7_0_10	C=C1
6	0,22	0,65	1,42	0,07	p2=P2_90_100, p3=P3_90_100	C=C2
7	0,22	0,81	1,5	0,07	p3=P3_80_90	C=C1
8	0,22	0,93	1,71	0,09	p2=P2_90_100, p3=P3_80_90	C=C1
9	0,22	0,87	1,6	0,08	p6=P6_20_30	C=C1
10	0,19	0,85	1,56	0,07	p2=P2_90_100, p6=P6_20_30	C=C1
11	0,19	0,92	2	0,09	p5=P5_60_70	C=C2

12	0,17	0,5	0,92	-0,01	p4=P4_60_70	C=C1
13	0,17	0,5	1,09	0,01	p4=P4_60_70	C=C2
14	0,17	0,53	1,15	0,02	p7=P7_10_20	C=C2
15	0,15	0,6	1,11	0,01	p2=P2_90_100, p4=P4_60_70	C=C1
16	0,15	0,47	0,87	-0,02	p7=P7_10_20	C=C1
17	0,15	1	2,19	0,08	p3=P3_90_100, p5=P5_60_70	C=C2
18	0,15	0,9	1,97	0,07	p4=P4_90_100	C=C2
19	0,14	1	1,84	0,06	p3=P3_80_90, p6=P6_20_30	C=C1
20	0,14	1	1,84	0,06	p2=P2_90_100, p3=P3_80_90, p6=P6_20_30	C=C1
21	0,14	0,67	1,23	0,03	p6=P6_10_20	C=C1
22	0,14	0,8	1,75	0,06	p6=P6_40_50	C=C2
23	0,14	0,89	1,94	0,07	p3=P3_90_100, p4=P4_90_100	C=C2
24	0,14	0,8	1,75	0,06	p5=P5_50_60	C=C2
25	0,14	0,89	1,64	0,05	p5=P5_30_40	C=C1
26	0,14	0,89	1,64	0,05	p5=P5_40_50	C=C1
27	0,14	1	1,84	0,06	p2=P2_90_100, p5=P5_40_50	C=C1

Описание работы классификатора представлено на речонке 2.7. Проверка точности работы полученного классификатора исследовалась на тестовой выборке, состоящей из данных 59 сваркой. Результаты тестирования классификатора качества сварки представлены в таблице 3.2.

Таблица 3.2 – Результаты тестирования полученного классификатора на тестовой выборке данных

№ объекта выборки	Номера правил (таблица 3.1), указывающих на класс С1	Номера правил (таблица 3.1), указывающих на класс С2	Количество указаний на С1	Количество указаний на С2	Прогноз класса	Фактический класс
1	3, 25	2, 18, 23	2	3	С2	С2
2	12	11, 13	1	2	С2	С2
3	7	18	1	1	С2	С2
4	12, 16	2, 13, 14, 24	2	4	С2	С2
5	1, 12, 15, 16	2, 4, 6, 13, 14	4	5	С2	С2
6	-	2, 11, 17, 18, 23	0	5	С2	С2
7	-	2	0	1	С2	С2
8	7, 16	14, 24	2	2	С2	С2
9	1, 16	2, 4, 6, 11, 14, 17, 18, 23	2	8	С2	С2
10	1, 16	2, 4, 6, 11, 14, 17, 18, 23	2	8	С2	С2
11	1, 3	2, 4, 6, 11, 17, 18, 23	2	7	С2	С2
12	-	2, 11, 17	0	3	С2	С2
13	1, 3, 26, 27	2, 4, 6	4	3	С1	С1
14	1, 12, 15	2, 4, 6, 13, 24	3	5	С2	С2
15	1, 16	2, 4, 6, 14, 18, 23	2	6	С2	С1
16	1, 7, 8, 16	4, 14, 24	4	3	С1	С1
17	1, 7, 8, 26, 27	4	5	1	С1	С1
18	1, 3, 25	4	3	1	С1	С1
19	1, 3, 25	2, 4, 6	3	3	С2	С1
20	1, 3, 12, 15, 25	2, 4, 6, 13	5	4	С1	С1
21	1, 7, 8, 26, 27	4	5	1	С1	С1
22	1, 7, 8, 25	4	4	1	С1	С1
23	1, 3, 7, 8	4	4	1	С1	С1
24	1, 16	4, 14	2	2	С2	С2
25	1, 16	4, 14	2	2	С2	С1
26	1, 7, 8, 12, 15, 16	4, 13, 14, 24	6	4	С1	С1
27	1, 7, 8, 26, 27	4	5	1	С1	С1
28	3, 12	2, 13, 24	2	3	С2	С2
29	26	-	1	0	С1	С2
30	-	2, 18, 23, 24	0	4	С2	С2

31	12, 16	11, 13, 14	2	3	C2	C2
32	1, 3, 12, 15	2, 4, 6, 13, 24	4	5	C2	C2
33	1, 12, 15	2, 4, 6, 13, 24	3	5	C2	C2
34	1, 16	2, 4, 6, 11, 14, 17, 18, 23	2	8	C2	C2
35	1, 16	2, 4, 6, 11, 14, 17, 18, 23	2	8	C2	C2
36	1	2, 4, 6, 11, 17	1	5	C2	C2
37	1, 16	2, 4, 6, 11, 14, 17	2	6	C2	C2
38	1, 12, 15	2, 4, 6, 13, 24	3	5	C2	C2
39	1, 7, 8	4, 11	3	2	C1	C1
40	1, 16, 26, 27	2, 4, 6, 14	4	4	C2	C1
41	1, 7, 8, 12, 15, 26, 27	4, 13	7	2	C1	C1
42	1, 7, 8, 12, 15, 16, 26, 27	4, 13, 14	8	3	C1	C1
43	1, 3, 7, 8, 12, 15, 25	4, 13	7	2	C1	C1
44	1, 3, 7, 8, 12, 15, 25	4, 13	7	2	C1	C1
45	1, 3, 12, 15, 25	2, 4, 6, 13	5	4	C1	C1
46	1, 3, 12, 15	2, 4, 6, 13	4	4	C2	C1
47	1, 7, 8, 16	4, 14	4	2	C1	C1
48	1, 3, 7, 8, 12, 15	4, 13	6	2	C1	C2
49	1, 3	4	2	1	C1	C1
50	1, 3	4	2	1	C1	C1
51	1, 3	2, 4, 6	2	3	C2	C2
52	1, 3	4	2	1	C1	C1
53	3, 12	13	2	1	C1	C1
54	1, 3, 12, 15	4, 13	4	2	C1	C1
55	3	-	1	0	C1	C1
56	1, 3, 26, 27	4	4	1	C1	C1
57	16	14	1	1	C2	C1
58	3	-	1	0	C1	C1
59	16, 25	14	2	1	C1	C1

Результаты тестирования полученного классификатора качества сварных соединений показывает, что в 51 из 59 случаев класс качества был определен правильно. Таким образом точность работы классификатора составляет 86,44%.

3.2 Программная реализация предложенных решений

При выполнении бакалаврской работы для реализации предложенных решений было разработано программное обеспечение, обладающее следующими характеристиками:

- наличие графического интерфейса;
- наличие программная реализация алгоритма построения классификатора на основе анализа экспериментальных данных контактной сварки;
- программное моделирование работы классификатора;
- импорт экспериментальных данных (обучающей выборки) и тестовой выборки из Excel;
- экспорт результатов тестирования классификатора на тестовой выборки в Excel файл для дальнейшего анализа точности его работы;
- сохранение правил классификации в отдельный файл.

Работа с программой осуществляется следующим образом. При запуске приложения откроется главное окно программы, внешний вид которого представлен на рисунке 3.1 (при запуске программы таблицы «Исходные данные», «Транзакции», «Правила классификации» и «Тестовая выборка» будут пустыми).

Первый шагом необходимо импортировать в программу данные экспериментальных сварок. Для этого необходимо выбрать Excel файл, содержащий в себе электронную таблицу с соответствующими данными. Это делается путем выбора в пункте «Файл» подпункта «Загрузить обучающую выборку». Содержимое пункта «Файл» показано на рисунке 3.2.

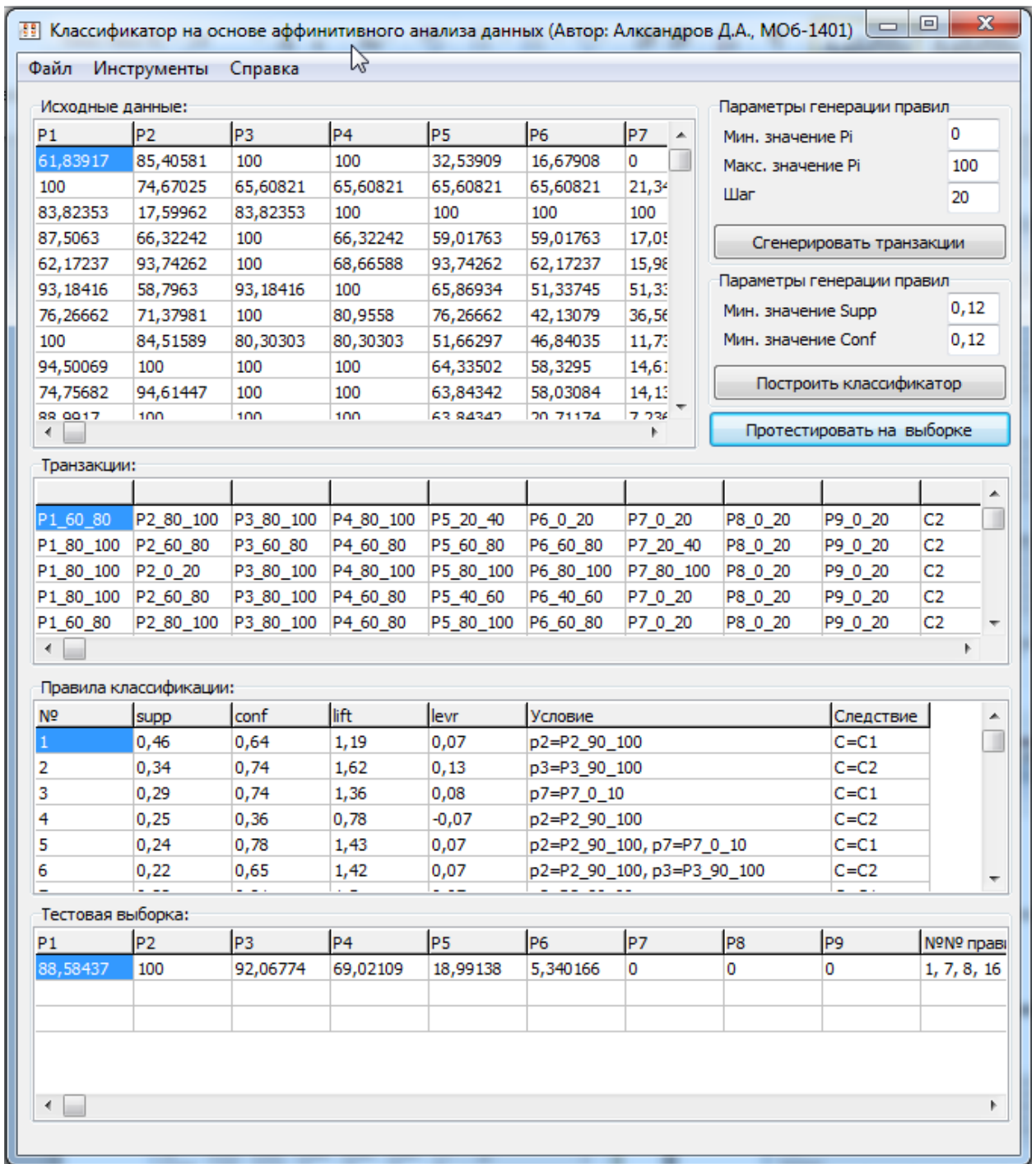


Рисунок 3.1 – Главное окно программы

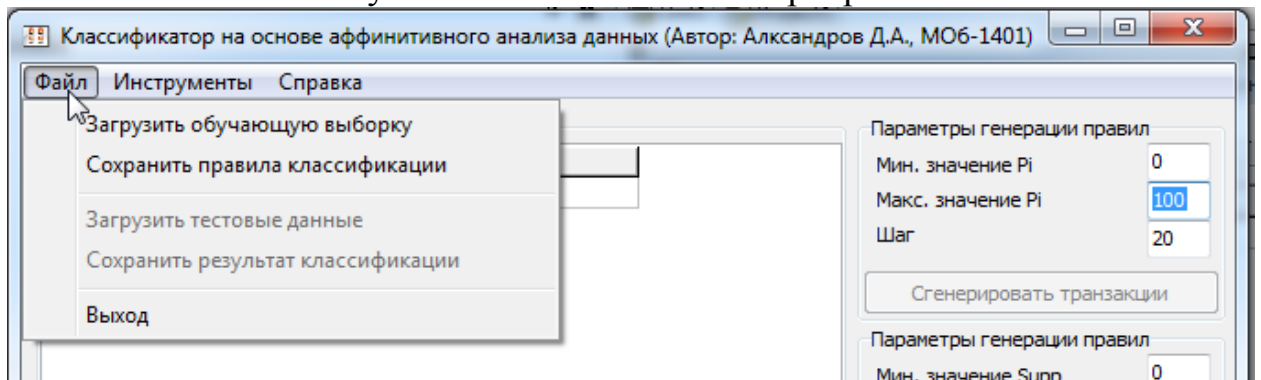


Рисунок 3.2 – Содержимое пункта меню «Файл»

При выполнении импорта данных будет заполнена таблица «Исходные данные», занимающая верхнюю часть главного окна программы. Результат выполнения импорта исходных данных приложен на рисунке 3.3.

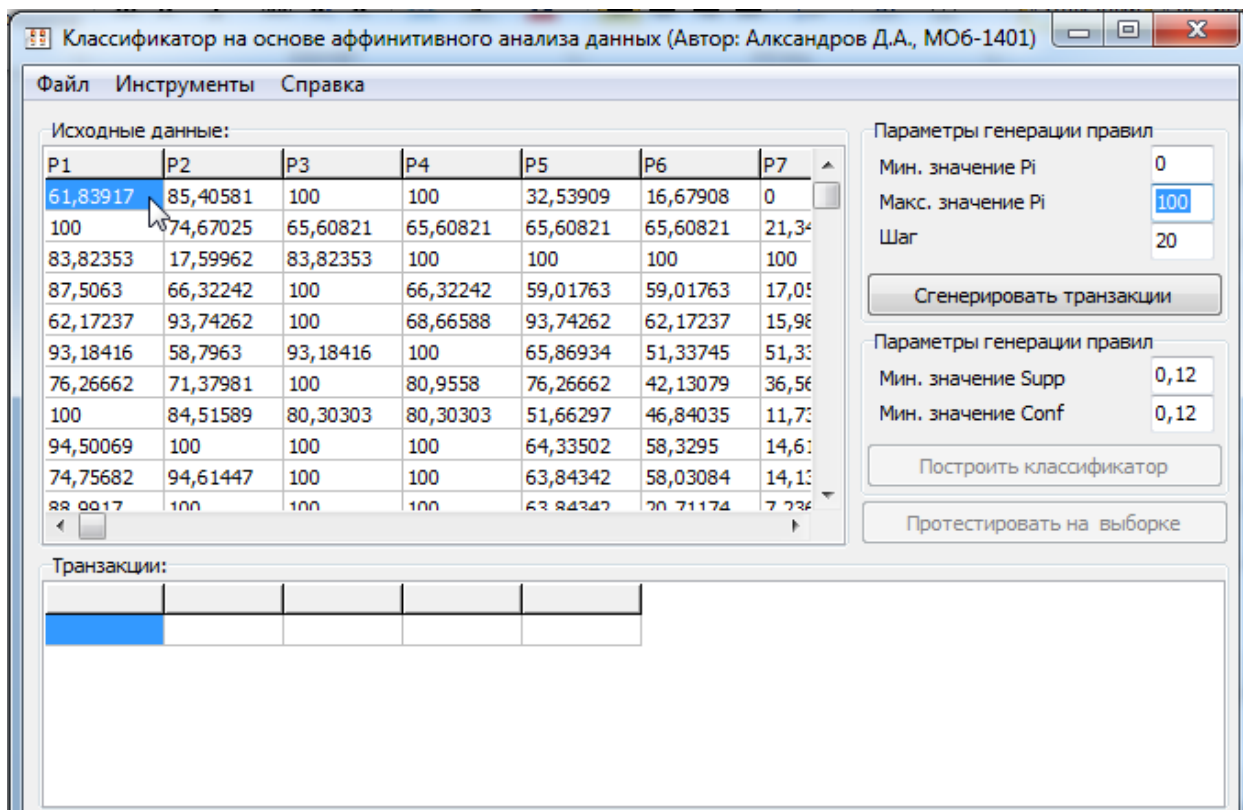


Рисунок 3.3 – Загрузка исходной выборки данных

Как только как только в программу загружены экспериментальные данные процесса сварки, можно приступать к формированию множества элементарных событий и формированию на их основе множества транзакций.

Для этого необходимо в правой верхней части программы задать параметры генерации правил и нажать на кнопку «Сгенерировать транзакции».

Результат выполнения данного этапа представлен на рисунке 3.4.

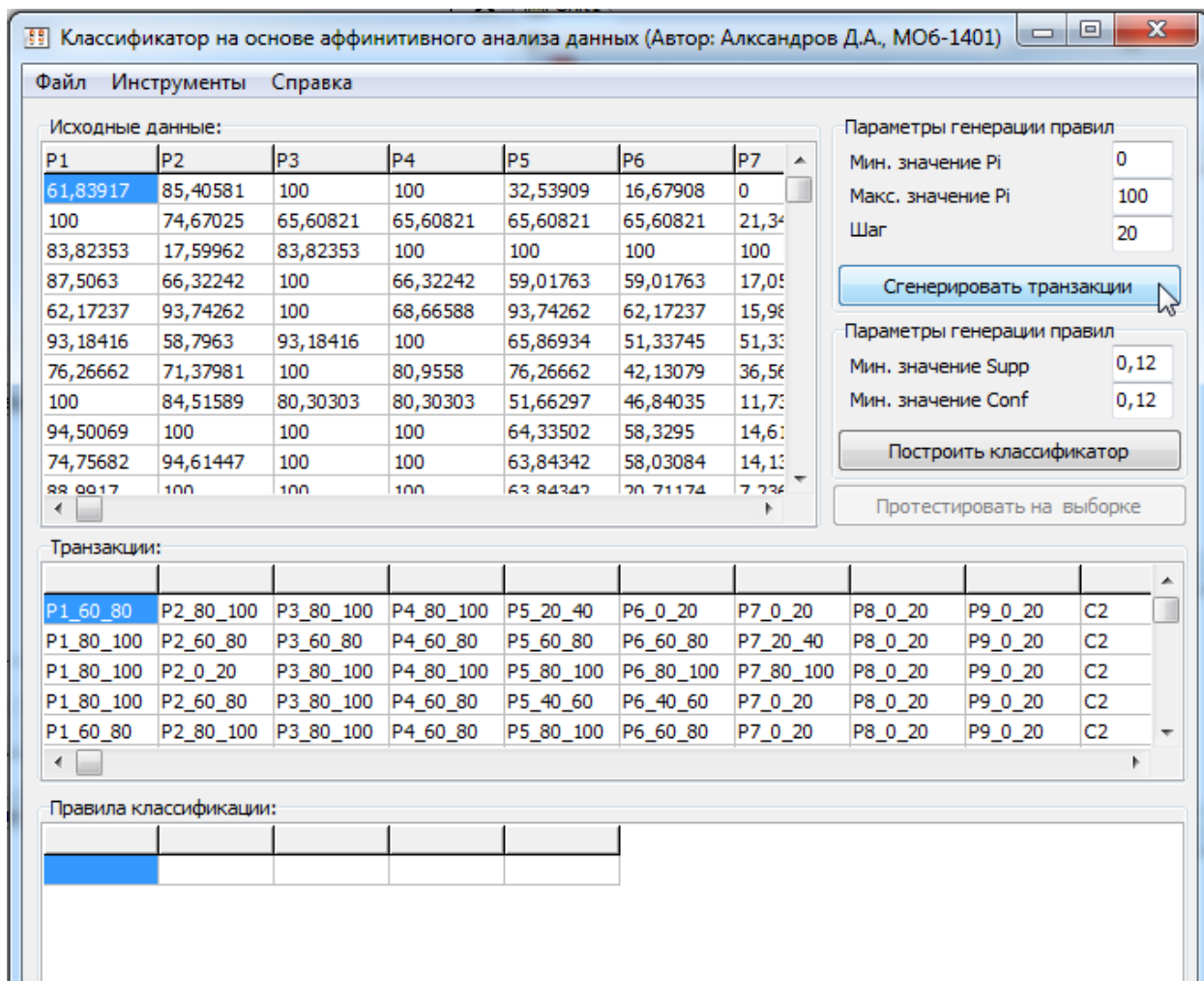


Рисунок 3.4 – Генерирование множества транзакций

Как только выполнено преобразование экспериментальных данных контактной сварки в множество транзакций можно переходить к аффинитивному анализу данных с целью поиска правил классификации объектов.

Для этого необходимо в правой части минимальные значения для поддержки и достоверности ассоциативных правил (правила со значениями ниже заданных включаться в классификатор не будут). Затем для запуска алгоритма нажать на кнопку «Построить классификатор». Найденные ассоциативные правила будут добавлены в таблицу «Правила классификации».

Результат выполнения данного этапа представлен на рисунке 3.5.

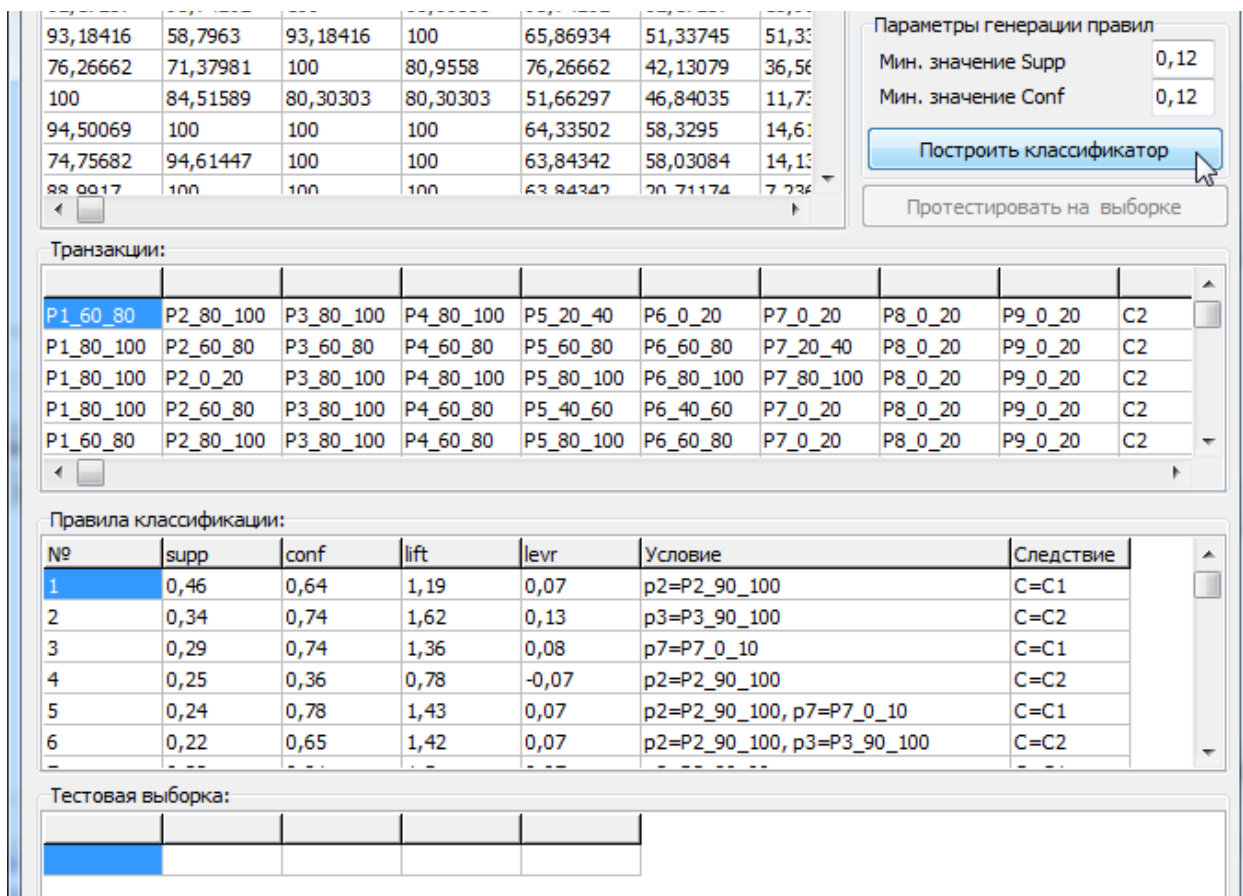


Рисунок 3.5 – Генерирование правил классификации

После того, как был сформирован классификатор качества сварки, можно протестировать его работу.

Чтобы это сделать, необходимо импортировать тестовую выборку данных из Excel файла. Для этого необходимо выбрать подпункт «Загрузить тестовые данные» в пункте «Файл» (рисунок 3.2). При этом загруженные данные отобразятся в таблице «Тестовая выборка», расположенной в нижней части главного окна программы.

Теперь, нажав на кнопку «Протестировать на выборке» можно запустить процесс классификации объектов. Результат классификации каждого объект будет отображен в столбце «Прогноз» таблицы. Номера правил, сработавших при классификации будут отображены в столбцах «№№ правил C1» и «№№ правил C2».

Транзакции:									
P1_60_80	P2_80_100	P3_80_100	P4_80_100	P5_20_40	P6_0_20	P7_0_20	P8_0_20	P9_0_20	C2
P1_80_100	P2_60_80	P3_60_80	P4_60_80	P5_60_80	P6_60_80	P7_20_40	P8_0_20	P9_0_20	C2
P1_80_100	P2_0_20	P3_80_100	P4_80_100	P5_80_100	P6_80_100	P7_80_100	P8_0_20	P9_0_20	C2
P1_80_100	P2_60_80	P3_80_100	P4_60_80	P5_40_60	P6_40_60	P7_0_20	P8_0_20	P9_0_20	C2
P1_60_80	P2_80_100	P3_80_100	P4_60_80	P5_80_100	P6_60_80	P7_0_20	P8_0_20	P9_0_20	C2

Правила классификации:						
№	supp	conf	lift	levr	Условие	Следствие
1	0,46	0,64	1,19	0,07	p2=P2_90_100	C=C1
2	0,34	0,74	1,62	0,13	p3=P3_90_100	C=C2
3	0,29	0,74	1,36	0,08	p7=P7_0_10	C=C1
4	0,25	0,36	0,78	-0,07	p2=P2_90_100	C=C2
5	0,24	0,78	1,43	0,07	p2=P2_90_100, p7=P7_0_10	C=C1
6	0,22	0,65	1,42	0,07	p2=P2_90_100, p3=P3_90_100	C=C2

Тестовая выборка:									
P6	P7	P8	P9	№№ правил	№№ правил	Количество	Количество	Факт	Прогноз
5,340166	0	0	0	1, 7, 8, 16	4, 14	4	2	C1	C1

Рисунок 3.5 – Классификация объектов тестовой выборки

Результат классификации объектов тестовой выборки можно экспортировать в Excel файл для дальнейшего анализа результатов. Для этого необходимо выбрать подпункт «Сохранить результат классификации» в пункте «Файл» главного меню программы (рисунок 3.2). При вместе с результатом классификации будут экспортированы все данные, содержащиеся в таблице «Тестовая выборка».

ЗАКЛЮЧЕНИЕ

На основе проведенных в бакалаврской работе исследований были сделаны следующие выводы:

1. Анализ литературных источников показал, что использование методов машинного обучения позволяет автоматизировать задачу синтеза алгоритмов управления и диагностики различных технологических процессов (в том числе и контактной сварки).

2. Алгоритмы аффинитивного анализа являются относительно молодыми (алгоритм Apriori был предложен 90-х годах). Поэтому актуальными являются исследования направленные на изучение возможностей аффинитивного анализа в том числе в задачах классификации данных.

3. В результате исследований был разработан алгоритм синтеза классификатора данных, основанных на результатах аффинитивного анализа исходных данных. Показан пример практического использования алгоритма. Также описаны особенности использования получаемых с помощью предложенного алгоритма классификаторов.

4. С использованием предложенного алгоритма были проанализированы экспериментальные данные 100 контактных точечных сварок и получен классификатор качества сварных соединений. Построенный классификатор качества позволяет на основе измеряемых в процессе сварки параметров прогнозировать качество соединения. Тестирование полученного классификатора на данных 59 сварок позволило оценить точность его работы, которая составляет 86,44%.

5. В результате выполнения бакалаврской работы было реализовано программное обеспечение реализующее предложенные подходы синтеза классификаторов на основе аффинитивного анализа данных на практике. Алгоритм использования реализованного программного обеспечения описан в разделе 3.2 бакалаврской работы.

В ходе выполнения бакалаврской работы было опубликовано четыре статьи на тему практического применения алгоритмов машинного обучения [19-22].

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. Djenouri, Y. GA-Apriori: Combining Apriori Heuristic and Genetic Algorithms for Solving the Frequent Itemsets Mining Problem [Text] / Youcef Djenouri, Marco Comuzzi // Pacific-Asia Conference on Knowledge Discovery and Data Mining: Trends and Applications in Knowledge Discovery and Data Mining (PAKDD 2017). – Springer International Publishing AG, 2017. – pp. 138-148
3. Arora, P. Design and Performance Analysis of Distributed Implementation of Apriori Algorithm in Grid Environment [Text] / Priyanka Arora, Sarbjeet Singh // ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India. – Springer International Publishing Switzerland, 2014. – pp. 653-661
4. Zhao, X. An Improved Apriori Algorithm with Prejudging and Screening [Text] / Xuejian Zhao, Dongjun Li, Yuan Yuan, Zhixin Sun, Yong Chen // Advances in Computer and Computational Sciences: Proceedings of ICCCS 2016. – Springer Nature Singapore Pte Ltd., 2017. – pp. 649-657
5. Liao, B. An Improved Algorithm of Apriori [Text] / Binhua Liao // International Symposium on Intelligence Computation and Applications: Computational Intelligence and Intelligent Systems (ISICA 2009). – Springer-Verlag Berlin Heidelberg, 2009. – pp. 427-432
6. Benhamouda, N.C. Meta-Apriori: A New Algorithm for Frequent Pattern Detection [Text] / Neyla Cherifa Benhamouda, Habiba Drias, Célia Hirèche // Asian Conference on Intelligent Information and Database Systems: Intelligent Information and Database Systems (ACIIDS 2016). – Springer-Verlag Berlin Heidelberg, 2016 – pp. 277-285

7. Dhanya, S. An Enhancement of the MapReduce Apriori Algorithm Using Vertical Data Layout and Set Theory Concept of Intersection [Text] / S. Dhanya, M. Vysaakan, A. S. Mahesh // Intelligent Systems Technologies and Applications. – Springer International Publishing Switzerland, 2016. – pp. 225-233
8. Fernández-Baizán, M.C. Using the Apriori Algorithm to Improve Rough Sets Results [Text] / María C. Fernández-Baizán, Menasalvas Ruiz, José M. Peña Sánchez, Juan Francisco Martínez Sarrías, Socorro Millán // International Conference on Rough Sets and Current Trends in Computing (RSCTC 2000). – Springer-Verlag Berlin Heidelberg, 2001. – pp. 291-295
9. Liang, X. Improved Apriori Algorithm for Mining Association Rules of Many Diseases [Text] / Xu Liang, Caixia Xue, Ming Huang // International Symposium on Intelligence Computation and Applications: Computational Intelligence and Intelligent Systems (ISICA 2010). – Springer-Verlag Berlin Heidelberg, 2010. – pp. 272-279
10. Xiang, L. Simulation System of Examination Score Analysis Based on an Improved Apriori Algorithm [Text] / Li Xiang // Advances in Computer Science and Engineering. – Springer-Verlag Berlin Heidelberg, 2012. – pp. 19-25
11. Sharma, N.K. Study and Analysis of Incremental Apriori Algorithm [Text] / Neeraj Kumar Sharma, N. K. Nagwani // International Conference on High Performance Architecture and Grid Computing (HPAGC 2011). – Springer-Verlag Berlin Heidelberg, 2011. – pp. 470-472
12. Agarwal, J. Mining Frequent Quality Factors of Software System Using Apriori Algorithm [Text] / Jyoti Agarwal, Sanjay Kumar Dubey, Rajdev Tiwari // Proceedings of the International Conference on Data Engineering and Communication Technology (ICDECT 2016). – Springer Science+Business Media Singapore, 2017. – pp. 481-490
13. Choo, Y.H. A Rough-Apriori Technique in Mining Linguistic Association Rules / Yun-Huoy Choo, Azuraliza Abu Bakar, Abdul Razak Hamdan // International Conference on Advanced Data Mining and Applications (ADMA 2008). – Springer-Verlag Berlin Heidelberg, 2008. – pp. 548-555

14. Dahbi, A. Using Multiple Minimum Support to Auto-adjust the Threshold of Support in Apriori Algorithm / Azzeddine Dahbi, Youssef Balouki, Taoufiq Gadi // Proceedings of the Ninth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2017). – Springer International Publishing AG 2018. – pp. 111-119
15. Jovanoski, V. Classification Rule Learning with APRIORI-C / Viktor Jovanoski, Nada Lavrač // Portuguese Conference on Artificial Intelligence (EPIA 2001). – Springer-Verlag Berlin Heidelberg, 2001. – pp. 44-51
16. Inokuchi, A. An Apriori-Based Algorithm for Mining Frequent Substructures from Graph Data / Akihiro Inokuchi, Takashi Washio, Hiroshi Motoda // European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2000). – Springer-Verlag Berlin Heidelberg, 2000. – pp. 13-23
17. Pommerenke, C. A Modified Apriori Algorithm for Analysing High-Dimensional Gene Data / Claudia Pommerenke, Benedikt Friedrich, Thorsten Johl, Lothar Jänsch, Susanne Häussler, Frank Klawonn // International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2011). – Springer-Verlag Berlin Heidelberg, 2011. – pp. 236-243
18. Child, Ch. The Apriori Stochastic Dependency Detection (ASDD) Algorithm for Learning Stochastic Logic Rules / Christopher Child, Kostas Stathis // International Workshop on Computational Logic in Multi-Agent Systems (CLIMA 2004). – Springer-Verlag Berlin Heidelberg, 2004. – pp. 234-249
19. Александров, Д.А. Применение нейронных сетей прямого распространения для прогнозирования прочности сварных соединений [Текст] / Д.А. Александров, В.А. Костылев // 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. Александров, Д.А. Построение регрессионной модели для прогнозирования прочности сварного соединения [Текст] / В.А. Костылев, 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.

21. Александров, Д.А. Сегментация изображения с помощью алгоритма кластеризации / Д.А. Александров, С.Р. Никитин // 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.

22. Александров, Д.А. Модификация алгоритма для интеллектуальных систем распознавания образов / 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.