

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

Институт математики, физики и информационных технологий  
Кафедра «Прикладная математика и информатика»

01.03.02 ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА

СИСТЕМНОЕ ПРОГРАММИРОВАНИЕ И КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ

### **БАКАЛАВРСКАЯ РАБОТА**

на тему: **Анализ экспериментальных данных контактной сварки с  
помощью алгоритмов машинного обучения**

Студент	_____	С. А. Петрушкин	_____
Руководитель	_____	В. С. Климов	_____
Консультант по аннотации	_____	Н. В. Яценко	_____

**Допустить к защите**  
Заведующий кафедрой А. В. Очеповский \_\_\_\_\_

« \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

Тольятти 2017

## **Аннотация**

Тема: «Анализ экспериментальных данных контактной сварки с помощью алгоритмов машинного обучения».

В данной выпускной квалификационной работе исследуются пути совершенствования методов диагностики сварки. В исследовании рассматривается задача прогнозирования качества сварки с использованием алгоритма ID3. При этом предложено наблюдать за изменением коэффициента мощности сварки и определять класс качества соединения. В работе представлены результаты апробации алгоритма на практике.

Структура ВКР представлена введением, тремя главами, заключением, списком литературы.

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

В первой главе проводится обзор методов анализа экспериментальных данных контактной сварки с помощью алгоритмов машинного обучения.

Во второй главе описывается проектирование программного обеспечения, в основе которого будет лежать алгоритм, анализировать данные контактной сварки с помощью алгоритмов машинного обучения.

В третьей главе будет проведена реализация и тестирование разработанного программного обеспечения, а также представлены результаты его работы.

В заключении представлены выводы по проделанной работе.

В работе использовано 12 таблиц, 24 рисунка, список литературы содержит 39 литературных источников. Общий объем выпускной квалификационной работы составляет 42 страниц.

## **ABSTRACT**

The title of the graduation work is “The analysis of experimental data of resistance welding by means of algorithms of machine learning”.

In this graduation work ways of improvement of methods of diagnostics of welding are investigated. In a research the problem of forecasting of quality of welding with use of an algorithm of "ID3" is considered. At the same time it is offered to monitoring change of power factor of welding and to define a class of quality of welded joint. In work results of approbation of an algorithm in practice are presented.

The structure of the graduation project is represented by an introduction, three chapters, a conclusion and a list of references. The introduction describes the relevance of the research conducted, the goal and the tasks.

In the first chapter, the review of methods of analysis of experimental data of resistance welding by means of algorithms of machine learning.

The second chapter describes the design of software, which is based on analysis of experimental data of resistance welding by means of algorithms of machine learning.

In the third chapter, the software was implemented and tested. Also, the results of the software were provided.

In conclusion conclusions on the work are presented.

The graduation work consists of 12 tables, 24 drawings and 4 applications, the list of references contains 37 sources. The total amount of final qualifying work is 78 pages.

## Оглавление

Введение.....	3
1 Анализ состояния вопроса об использовании алгоритмов машинного обучения при диагностике сварки.....	6
1.1 Проблемы диагностики контактной сварки.....	6
1.2. Применение искусственного интеллекта в задаче диагностики сварки.....	9
2 Разработка алгоритма диагностики сварки, основанного на машинном обучении.....	21
2.1 Алгоритм ID3 как метод построения модели классификации данных.....	21
2.2 Контактная сварка как объект диагностики.....	26
2.3 Алгоритм диагностики сварки с использованием ID3.....	30
2.4 Экспериментальная апробация предложенного алгоритма.....	31
3 Практическая реализация предложенных решений.....	35
3.1 Программная реализация алгоритма.....	35
3.2 Пример работы программы.....	40
Заключение.....	44
Список используемой литературы.....	46
Приложение А. UML диаграммы классов	<b>Ошибка! Закладка не определена.</b>

## Введение

Исследование любого технологического процесса начинается с составления математической его модели. Для управления и диагностики процессом требуется составления таких моделей, которые описывали бы связь входных управляющих сигналов и требуемых на выходе характеристик процесса. Однако существуют такие технологические процессы, для которых по объективным причинам не разработаны удовлетворительные математические модели, описывающих связь входных воздействий и получаемых характеристик. Примером того технологического процесса является – контактная точечная сварка.

Технология процесса контактной сварки подразумевает сжатие свариваемых деталей между двумя цилиндрическими электродами и пропускание через них нескольких импульсов тока. Ток производит нагрев и расплавление металла в зоне между электродами. При кристаллизации металла, образуется неразъемное соединение в виде сварной точки.

Сложность разработки математической модели связана с тем, что при сварке управление производится электрическими процессами (путем фазового регулирования), но при этом требуется обеспечить физико-механические характеристики соединения (например, силу разрушения соединения срез). Таким образом, математическая модель должна содержать в себе следующие компоненты:

- моделирование электрических процессов в сварочной машине (с учетом нелинейных процессов в сердечнике трансформатора);
- моделирование изменения плотности тока в сечении свариваемых деталей;
- моделирование изменения теплового поля в зоне сварки;
- моделирование процесса расплавления и кристаллизация металла в зоне сварки;

- моделирование разрушение соединения при статической и динамической нагрузке детали.

В теории, построение такой модели позволило бы, при диагностике сварки, прогнозировать прочность соединений (с необходимой точностью). При управлении такая модель позволила бы разрабатывать алгоритмы по обеспечению требуемой прочности соединений путем управления электрическими процессами.

С развитием алгоритмов машинного обучения появилась возможность автоматизировать построение математических моделей для сложных процессов. При этом алгоритмы машинного обучения самостоятельно анализируют экспериментальные данные и ищут в них закономерности.

Уже существуют исследования по применению различных типов нейронных сетей (как прямого распространения, так и рекуррентных) для прогнозирования прочностных характеристик соединения. Эти исследования проводят такие крупнейшие промышленные компании, как Weld Computer Corporation (США), Nissan Motor Company (Япония), Lockheed Martin Energy Research Corporation (США), General Motors Corporation (США) и др.

Однако не исследована возможность применения алгоритмов построения деревьев принятия решения для автоматического построения моделей классификации качества сварки.

Поэтому в данной исследовательской работе проверяется гипотеза о возможности использования алгоритма ID3 для автоматического построения классификатора оценка качества сварки. В процессе сварки измеряются электрические параметры, по которым с использованием классификатора прогнозируется соответствие/несоответствие соединения заданным прочностным характеристикам.

Целью работы является снижение трудоёмкости синтеза систем диагностики качества контактной сварки за счет применения алгоритмов построения деревьев принятия решений.

В ходе выполнения исследований на языке Java с использованием среды NetBeans была разработана программа, позволяющая по данным экспериментальных сварок с помощью алгоритмов машинного обучения строить модель классификации качества сварки в виде дерева принятия решений. Получаемые модели можно использовать для прогнозирования соответствия сварного соединения требуемым прочностным характеристикам.

# **1 Анализ состояния вопроса об использовании алгоритмов машинного обучения при диагностике сварки**

## **1.1 Проблемы диагностики контактной сварки**

Контактная сварка является самым высокопроизводительным процессом получения неразъемных соединений листовых металлических деталей во всех отраслях промышленности.



Рисунок 1.1 – Промышленный робот фирмы Nachi Robotic Systems Inc (США) для выполнения контактной сварки

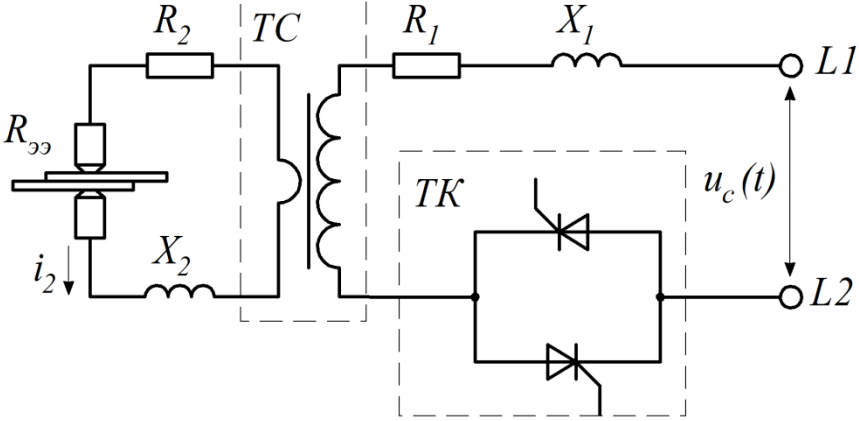
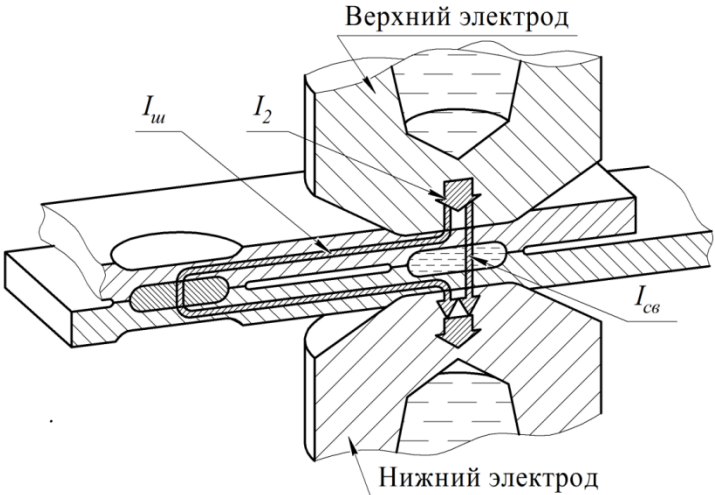
Технология процесса контактной сварки подразумевает сжатие свариваемых деталей между двумя цилиндрическими электродами и пропускание через них нескольких импульсов тока. Ток производит нагрев и расплавление металла в зоне между электродами. При кристаллизации металла, образуется неразъемное соединение в виде сварной точки.

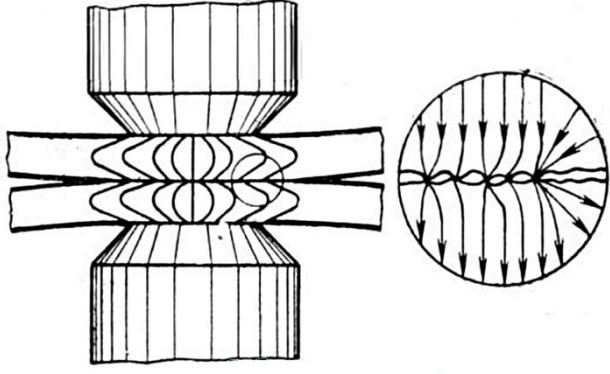
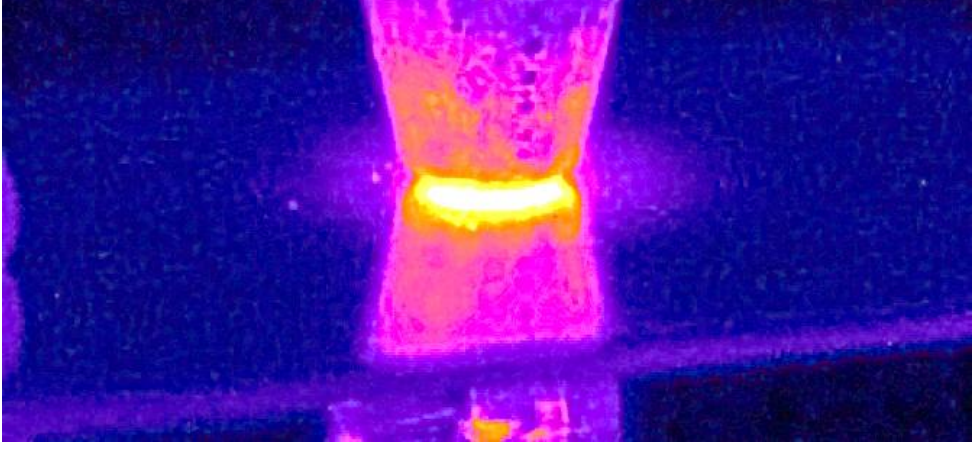
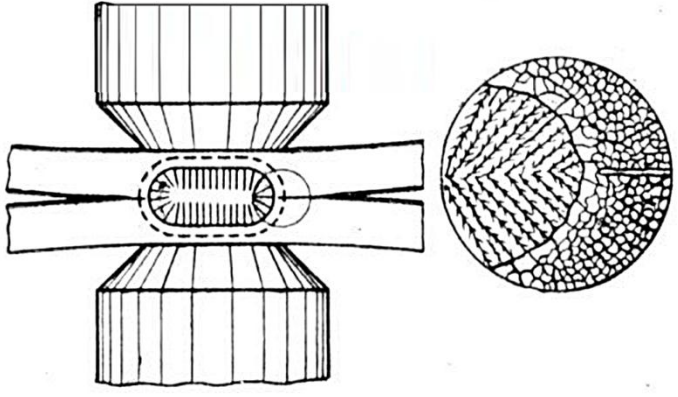
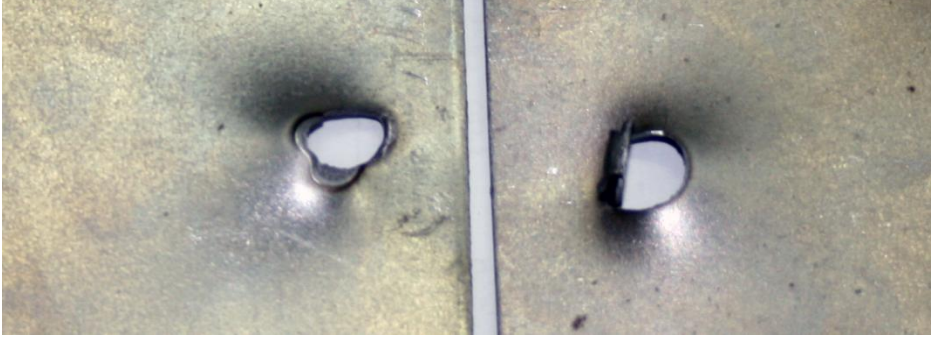
Основная проблема контактной сварки связана с нестабильностью качества получаемых соединений. При управлении сваркой регулируются электрические процессы (ток сварки), но на выходе требуется прочностные



характеристики (прочность соединений). Для прогнозирования точности с использованием классических подходов требуется составить математическую модель, связывающую электрические процессы в сварочном аппарате и прочностные характеристики соединения. Такая модель должна включать в себя компоненты, показанные в таблице 1.

Таблица 1 – Необходимые компоненты математической модели сварочного процесса для прогнозирования качества сварки.

Часть модели	Иллюстрация
Моделирование электрических процессов	
Моделирование протекания тока по детали	

<p>Моделирование изменения плотности тока в зоне сварки</p>	 <p>The image shows a 3D schematic of a cross-joint weld on the left, with lines representing current density distribution. On the right is a circular cross-section showing a vector field of current density with arrows pointing towards the center of the weld.</p>
<p>моделирование изменения теплового поля в зоне сварки</p>	 <p>A false-color thermal image showing a bright yellow and orange weld zone in the center, surrounded by a purple and blue heat-affected zone.</p>
<p>моделирование процесса расплавления и кристаллизация металла в зоне сварки</p>	 <p>The image shows a 3D schematic of a cross-joint weld on the left, with a central area indicating melting and crystallization. On the right is a circular cross-section showing a detailed microstructure of the metal.</p>
<p>Моделирование разрушения соединения</p>	 <p>Two photographs showing the failure of a weld joint. The left image shows a weld with a large, irregular hole, and the right image shows a weld with a smaller, more circular hole.</p>
<p>Моделирование влияние на процесс возмущающих</p>	<p>Износ электродов, колбания напряжения питающей сети; колебания давления в пневмосистеме привода сжатия электродов; несоотность электродов; загрязненность поверхности свариваемых деталей и др.</p>

Составления такой модели с точностью, необходимой для диагностики сварки – на данный момент остается нерешенной задачей.

Однако развитие алгоритмов искусственного интеллекта позволило разработчикам систем управления и диагностики процессом взглянуть на проблему с другой стороны.

## **1.2. Применение искусственного интеллекта в задаче диагностики сварки**

С эволюцией алгоритмов индуктивного машинного обучения начал формироваться отдельный класс систем диагностики качества сварки, способных точно предсказывать прочностные характеристики соединения по измеряемым в процессе сварки электрическим параметрам. Данные системы диагностики выгодно отличаются от других способов контроля тем, что результаты оценки качества доступны сразу же после окончания сварки для их получения не требуется проведения дополнительных испытаний текущего сварного соединения. Все исследования в данной области построены на гипотезе о том, что с помощью алгоритмов машинного обучения можно строить математические модели процесса, основанные не на физике процесса, а на анализе экспериментальных данных (частные случаи).

Так в одной из работ зарубежные ученые предлагают диагностику качества сварки привести к решению задачи классификации. При этом оценка качества проводится путем прогнозирования принадлежности полученного диаметра сварного ядра к одному из нескольких диапазонов с помощью Байесовской сети доверия. Входными параметрами модели классификации являются – кривые напряжения на электродах, тока сварки и силы сжатия (рис. 1.1) [32]. В процессе сварки проводится измерение электрических параметров с частотой 70кГц. Затем для полученных в результате измерения гистограмм вычисляют нижний и

верхний квартиль, а также медиану. С помощью этих характеристик Байесовская сеть доверия определяет принадлежность получившегося диаметра сварного ядра к одному из нескольких заранее определенных диапазонов (классов) (рис. 1.2). Если сеть определила принадлежность диаметра литого ядра к диапазону, соответствующему качественным соединениям, то текущее соединения считают качественным. В работе представлены результаты успешного использования двух конфигураций Байесовских сетей с различными входными параметрами, полученных из гистограмм: для первой сети входными параметрами являются – медиана напряжения на электродах, верхний квартиль сварочного тока и верхний квартиль силы сжатия, а для второй – медиана напряжения, медиана силы сжатия и верхний квартиль силы сжатия. Авторы отмечают, что при использовании двух сетей одновременно повышается достоверность оценки качества сварного соединения.

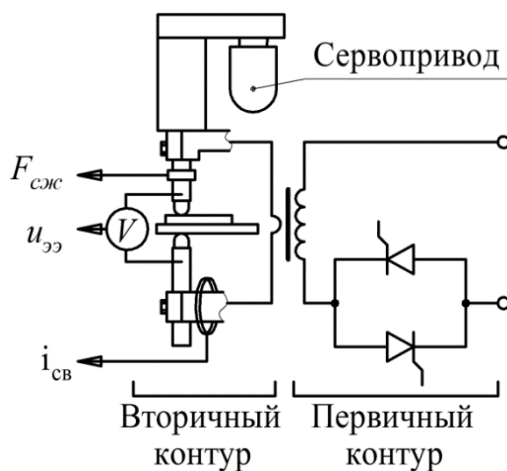


Рис. 1.1. Схема измерения входных параметров системы диагностики контактной сварки на основе Байесовской сети доверия, где  $F_{сж}$  – сила сжатия электродов,  $u_{ээ}$  – напряжение на участке «электрод-электрод»,  $i_{св}$  – ток сварки

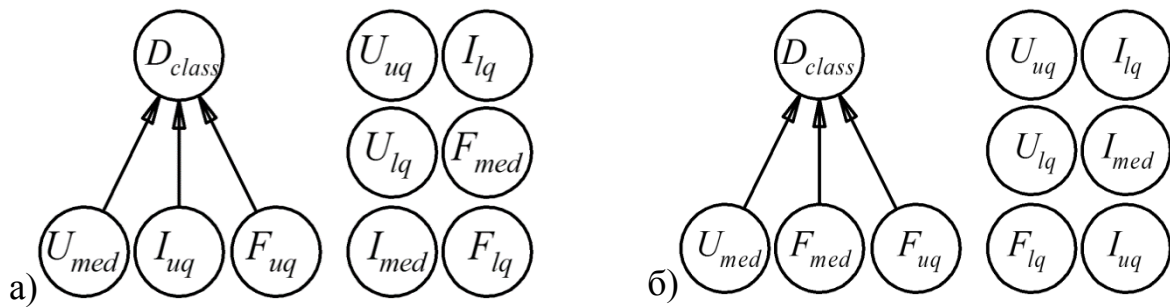


Рис. 1.2. Байесовские сети доверия для прогнозирования типоразмера  $D_{class}$  диаметра литого ядра: а) сеть, в которой входными параметрами являются медиана  $U_{med}$  напряжения на электродах, верхний квартиль  $I_{uq}$  сварочного тока и верхний квартиль  $F_{uq}$  силы сжатия электродов; б) сеть, в которой входными параметрами являются медиана  $U_{med}$  напряжения на электродах, верхний квартиль  $F_{uq}$  и медиана  $F_{med}$  силы сжатия электродов

Предложенный способ диагностики качества обладает рядом существенных недостатков при реализации его в производственной среде. Целая серия недостатков связана с необходимостью использования дополнительных датчиков на сварочной машине для измерения тока, напряжения на электродах и силы сжатия. К ним относятся усложнение и удорожание аппаратной реализации данного способа из-за необходимости обеспечения измерения всех основных параметров сварки с выбранным быстродействием (7000-8000 сигналов за один полупериод сварки). Кроме того, загромождение рабочего пространства сварочной машины датчиками приводит к усложнению доступа крупногабаритных заготовок к сварочным электродам. К недостаткам также можно отнести необходимость глубокой модернизации сварочной машины для установки нужных датчиков.

Кроме того, сами авторы отмечают, что полученные в лабораторных условиях измерения электрических параметров сварки из-за электромагнитных помех нуждаются в дополнительной обработке для извлечения полезного сигнала. Это может отрицательно сказываться на точности работы модели в условиях производственной среды из-за более сильного влияния электромагнитных помех.

Другим серьезным недостатком является необходимость проведения обширной обучающей выборки для обучения модели классификатора в виде Байесовской сети доверия. Причем, чем больше предварительных экспериментов будет произведено, тем точнее будет функционировать система оценки качества сварки [33-35]. Авторы данной работы отмечают, что требуется проведение не менее 200 экспериментов.

В другой зарубежной работе авторы предлагают оценку качества сварки привести к задаче регрессии. При этом прогнозируется сила разрушения соединения с помощью двухслойной сети прямого распространения [19]. В качестве входных параметров модели используются значения сопротивления участка электрод-электрод, которые рассчитываются по измеренным значениям сварочного тока и напряжения на электродах (рис. 1.3 а). Эти рассчитанные значения подаются во входной слой обученной нейронной сети, в котором количество нейронов равно количеству полупериодов сварки (рис. 1.3 б). Через выходной слой, состоящий из одного нейрона, сеть выдает предполагаемое значение разрушающего усилия соединения. Для обучения такой двухслойной нейронной сети авторы исследования предлагают использовать метод обратного распространения ошибки, предполагающего обучение по уже известным примерам сварки с экспериментально определенной силой разрушения соединений. Допустимая погрешность определения нейронной сетью силы разрушения сварного соединения учитывается в процессе обучения, которое продолжается до тех пор, пока не будет достигнута требуемая точность. В процессе функционирования система диагностики качества считает сварное соединение качественным, если его предсказанное значение разрушающего усилия выше заранее известного предельного значения.

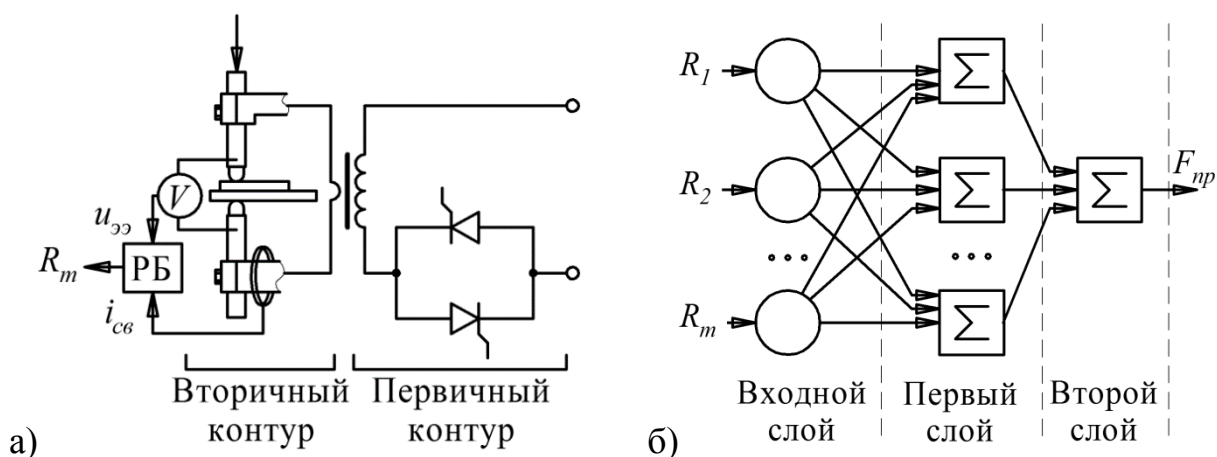


Рис. 1.3. Система диагностики контактной сварки, где  $u_{\text{ээ}}$  – мгновенное значение напряжения между электродами,  $i_{\text{св}}$  – мгновенное значение тока сварки,  $R_m$  – сопротивление на участке «электрод-электрод» в сварочном периоде с номером  $m$ : а – схема измерения параметров сварки, где РБ – расчетный блок; б – структура нейронной сети по предсказыванию разрушающего усилия соединения

Данный способ оценки качества сварки позволяет успешно предсказывать разрушающее усилие с требуемой точностью по динамике изменения сопротивления участка электрод-электрод, однако оно обладает недостатками, которые ограничивают его применение в производственной среде. Ряд недостатков связан с необходимостью использования дополнительных датчиков на сварочной машине для измерения сварочного тока, напряжения на электродах. Таких, как усложнение и удорожание аппаратной реализации данного способа из-за необходимости обеспечения в процессе сварки съема осциллограмм сварочного тока и напряжения. Загромождение рабочего пространства сварочной машины датчиками напряжения, тока и силы сжатия может стать непреодолимым препятствием при использовании данной системы диагностики качества, например, на сварочных роботах и при сварке крупногабаритных заготовок. Необходимость глубокой модернизации сварочной машины для установки

нужных датчиков связана со сложностью измерения напряжения на торцевой части электродов.

Также на измерение электрических параметров сварки оказывают влияния электромагнитные помехи, из-за которых осциллограммы напряжения и тока нуждаются в очистке от нежелательных сигналов. Кроме того, влияние электромагнитных помех в производственной среде сильнее, чем в лабораторных условиях, что отрицательно скажется на точности работы данной системы диагностики качества.

С точки зрения выбранного типа нейронной сети и алгоритма ее обучения, данный способ оценки качества обладает следующими недостатками, затрудняющими его использование в производственной среде. Длительность обучения нейронной сети носит неопределённый характер, т.е. для того, чтобы обучить систему производить оценку качества с заданной точностью, могут потребоваться дни или даже недели, нейронная сеть может и вообще не обучиться. Это связано с такими проблемами сети во время обучения, как паралич сети, попадание ее в локальный минимум [36-38]. Кроме того, настройка нейронной сети осуществляется по массиву данных, полученных в результате проведения многочисленных экспериментов по сварке, причем, чем больше этот массив данных, тем больше вероятности того, что сеть будет работать с заданной точностью [21, 22, 39]. Однако на практике вполне может оказаться, что нейронная сеть работает правильно только на массиве примеров, а на реальных сварных соединениях она совершает грубые ошибки. В этом случае сеть нуждается в переобучении [36, 40,41]. Такая исходная неопределенность обучения ставит под сомнение использования данной структуры нейронной сети с целью оценки качества сварки.

Существует и отечественные исследования по диагностики сварки с помощью алгоритмов машинного обучения. В одной из работ диагностика сварки приводится к задаче регрессии, которая решается с помощью двухслойной нейронной сети прямого распространения [42]. Выходной



параметр полученной модели - диаметра ядра сварной точки. Авторами предложено производить в процессе сварки измерение мгновенных значений сварочного тока и напряжения между электродами с интервалом от 50 мкс до 200 мкс (рис. 1.4 а). Затем полученные гистограммы предлагается разбить на четыре одинаковых по времени участка и для каждого интервала рассчитать эффективные значения тока напряжения. Рассчитанные значения подаются во входной слой обученной нейронной сети, а через выходной слой, состоящий из одного нейрона, сеть выдает предполагаемое значение диаметра литого ядра текущего сварного соединения (рис. 1.4 б). Авторы предлагают для обучения нейронной сети использовать метод обратного распространения ошибки, предполагающего обучение по уже известным примерам сварки с определенными значениями диаметра литого ядра сварных точек.

В работе представлены успешные результаты обучения нейронной сети для предсказания диаметра литого ядра сварного соединения низкоуглеродистой стали толщиной  $\delta=0,8+0,8$  мм с точностью 0,2 мм. Однако, данный способ оценки качества сварного соединения обладает рядом недостатков, связанных с необходимостью использования дополнительных датчиков на сварочной машине для измерения сварочного тока и напряжения на электродах. Эти недостатки аналогичны тем, которые приведены для предыдущего аналога. Кроме того, влияние электромагнитных помех в производственной среде сильнее, чем в лабораторных условиях, поэтому и погрешность работы данной системы диагностики качества будет значительно выше, чем описанная авторами. С точки зрения выбранного типа нейронной сети и алгоритма ее обучения, данный способ оценки качества не отличается от описанного выше аналога и наследует те же недостатки. Исключением является применение адаптивного алгоритма выбора шага коррекции весов, позволяющего выходить из локальных минимумов при обучении сети.

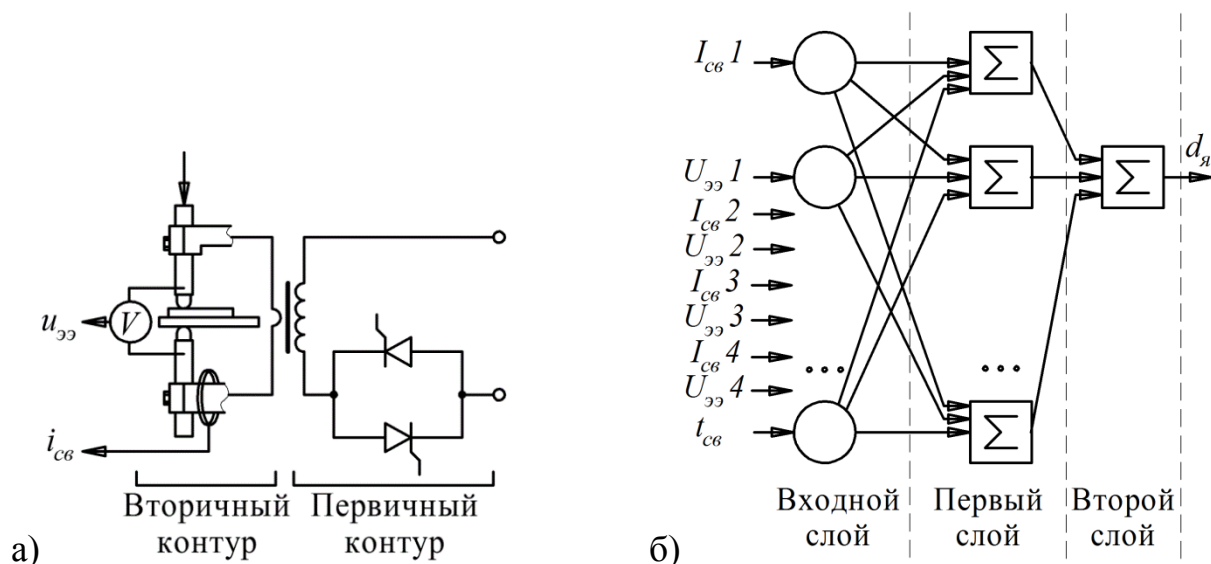


Рис. 1.4. Система диагностики контактной сварки, где  $d_я$  – диаметр литого ядра,  $u_{ээ}$  и  $U_{ээ}$  – мгновенное и действующее значение напряжения между электродами,  $i_{св}$  и  $I_{св}$  – мгновенное и действующее значение сварочного тока,  $t_{св}$  – длительность сварки: а – схема измерения параметров сварки, б – структура нейронной сети по предсказыванию диаметра литого ядра сварной точки

В другой зарубежной работе рассказывается о попытке применения конкурентной нейронной сети LVQ для определения принадлежности сварного соединения к одному из трех определенных классов (диагностика сварки в этом случае рассматривается как задача классификации)[20]. При этом в процессе сварки предлагается производить измерения мгновенных значений сварочного тока со стороны первичного контура и напряжения на электродах со стороны вторичного контура сварочной машины (рис. 1.5 а). Эти измерения используются для расчета динамики изменения сопротивления участка электрод-электрод, по которой обученная конкурентная LVQ нейронная сеть определяет принадлежность текущего соединения к одному из трех типов: “coldwelds” (дефект типа “склейка”), “normalwelds”, “expulsionwelds” (выплеск во время сварки) (рис. 1.5 б). Текущую сварку считают качественной, если нейронная сеть классифицировала ее с группой “normalwelds”. Для обучения такой

нейронной сети требуется проведение предварительных экспериментов с целью получения кривых изменения сопротивления участка электрод-электрод, соответствующих сварным соединениям с различным качеством сварки. В отличие от остальных аналогов, нейронная сеть данной системы диагностики качества сварки способна обучиться на гораздо меньшем количестве экспериментально полученных кривых сопротивления [43]. Авторы работы сообщают, что им удалось обучить нейронную сеть системы диагностики качества работать с заданной точностью на основе данных всего 14 сварок, что на порядок меньше, чем у описанных выше аналогов. Способность обучаться на малом количестве примеров означает, что при использовании такой системы диагностики качества на производстве, ее можно оперативно перенастроить под используемые параметры сварки.

Основные недостатки данного способа диагностики качества сварного соединения связаны с необходимостью использования дополнительных датчиков тока и напряжения на электродах. Эти недостатки аналогичны тем, которые приведены у представленных выше аналогов. Кроме того, влияние электромагнитных помех в производственной среде сильнее, чем в лабораторных условиях, поэтому и погрешность работы данной системы диагностики качества будет значительно выше, чем описанная авторами. Предложенный алгоритм использования конкурентной нейронной LVQ не позволяет получать численный результат оценки качества соединения: с ее помощью возможно диагностировать только такие дефекты сварки, как склейка и выплеск.

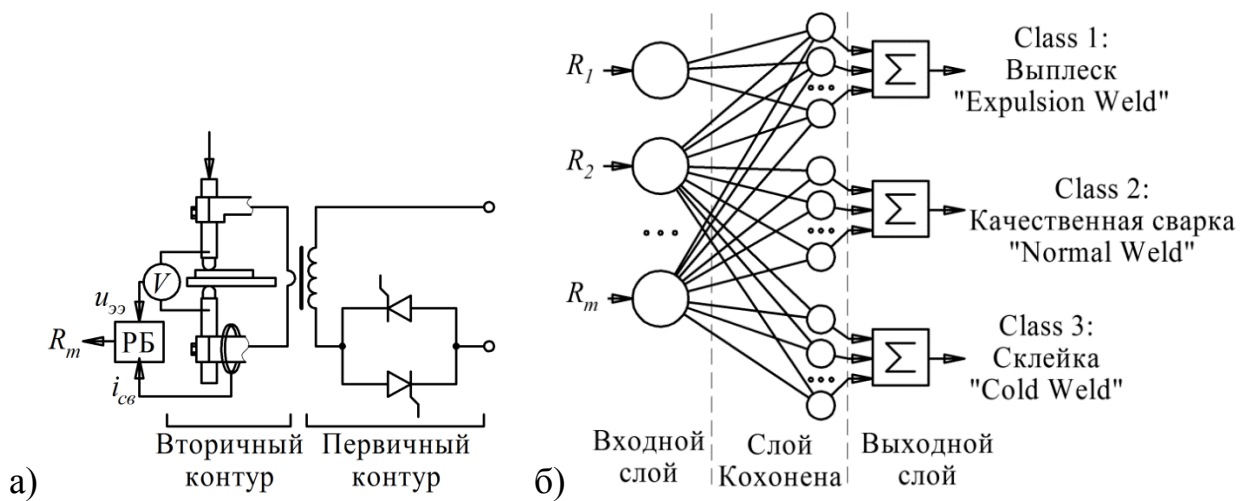


Рис. 1.5. Система диагностики контактной сварки, где  $u_{ээ}$  – мгновенное значение напряжения между электродами,  $i_{св}$  – мгновенное значение тока сварки,  $R_m$  – сопротивление на участке «электрод-электрод» в сварочном периоде с номером  $m$ : а – схема измерения параметров сварки, где РБ – расчетный блок; б – структура нейронной сети по определению дефектов контактной сварки выплеск и склейка

Авторы другой зарубежной работы также изучают возможность оценки качества сварки с помощью конкурентной нейронной сети LVQ [44]. Для этого предлагается измерять усилие сжатия электродов, которое в процессе сварки изменяется за счет теплового расширения металла (рис. 1.6 а). По полученной в результате измерения гистограммеобученная конкурентная нейронная сеть определяет принадлежность текущего соединения к одному из пяти типов: “insufficientwelding”, “poorwelding”, “goodwelding”, “richwelding”, “excessivewelding” (рис. 1.6 б). Данная работа во многом схожа предыдущим аналогом, при этом отказ от измерения во время сварки электрических параметров позволяет избежать влияния электромагнитных помех на измерения. Однако из-за выбранного для измерения параметра данный способ диагностики качества можно использовать только на стационарных сварочных машинах, при этом точность диагностики сварки становится сильно зависима от состояния подвижных частей сварочной машины [45, 46]. Использование конкурентной нейронной сети LVQ с

одной стороны позволяет снизить трудоемкость обучения сети, а с другой стороны исключает возможность получения численного результата оценки качества соединения [47].

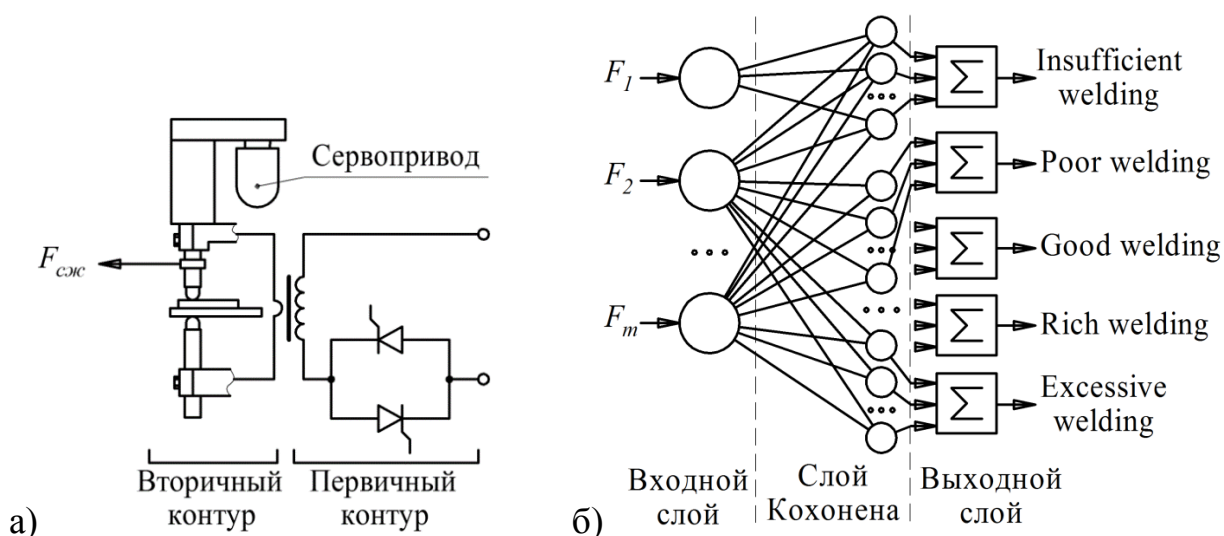


Рис. 1.6. Система диагностики контактной сварки, где

$F_{сж}$  – осциллограмма изменения силы сжатия электродов за сварку,  
 $F_m$  – сила сжатия электродов в сварочном периоде с номером  $m$ : а – схема измерения силы сжатия электродов, б – структура нейронной сети по определению типа сварного соединения

Таким образом, существующие на сегодняшний день работы по применению алгоритмов машинного обучения для диагностики качества сварки можно представить в виде сравнительной таблицы.

Недостатки применяемых в данных исследованиях алгоритмов машинного обучения заключаются в следующем:

1) Отсутствие повторяемости получаемых моделей, которая связана со стохастическим процессом их обучения. Даже при использовании одной и той же обучающей выборки получаемые модели всегда отличаются по точности работы.

2) Сложность анализа и интерпретации получаемых моделей. Использование приведенных выше алгоритмов машинного обучения заключается фактически в нахождении такого вектора параметров, при котором итоговая модель давала бы приемлемую точность.

Эти недостатки приводят к повышению трудоёмкости синтеза систем диагностики качества контактной сварки с использованием нейронных сетей и байесовских сетей доверия.

Эти недостатки отсутствуют алгоритма ID3, предназначенного для построения модели классификации данных. Однако не исследована возможность применения данного алгоритма для решения задач диагностики сварки. Поэтому в данной исследовательской работе проверяется гипотеза о возможности использования алгоритма ID3 для автоматического построения классификатора оценка качества сварки.

Таким образом, целью работы является снижение трудоёмкости синтеза систем диагностики качества контактной сварки за счет применения алгоритмов построения деревьев принятия решений.

Поставленная цель достигается путем последовательного решения ряда задач:

1. Анализ состояния вопроса об использовании алгоритмов машинного обучения в системах диагностики сварки.
2. Исследование возможности применения алгоритма ID3 для решения задач классификации качества сварки.
3. Разработка алгоритма диагностики сварки с использованием алгоритма ID3.
4. Проектирование, разработка, апробация программного обеспечения для реализации предложенных решений.

## 2 Разработка алгоритма диагностики сварки, основанного на машинном обучении

### 2.1 Алгоритм ID3 как метод построения модели классификации данных

Формально задача классификации описывается следующим образом. Существует множество входных переменных  $X$  – описаний объектов, а также множество  $Y$  (конечное) обозначений классов. В задаче классификации существует неизвестная целевая зависимость – отображение  $y^*: X \rightarrow Y$ , значения которой известны только на объектах из обучающей выборки  $X^m$ :

$$X^m = (x_1, y_1), \dots, (x_m, y_m) , \quad (2.1)$$

где  $m$  – размер обучающей выборки,  $x_i$  – вектор значений входных переменных для  $i$ -го объекта из обучающей выборки,  $y_i$  – значение метки класса для  $i$ -го объекта из обучающей выборки. При этом вектор  $x_i$  описывается набором значений атрибутов  $A$ , а  $y_i$  может принимать одно значений меток класса  $C_1, C_2, \dots, C_k$

$$\begin{cases} x_i = (A_1, A_2, \dots, A_n) \\ y_i \in C_1, C_2, \dots, C_k \end{cases} , \quad (2.2)$$

где,  $n$  – количество атрибутов, описывающих объект из обучающей выборки,  $k$  – количество меток класса в обучающей выборке.

Требуется построить алгоритм  $\alpha: X \rightarrow Y$ , способный классифицировать объект описываемый вектором входных значений  $x$ , при  $x \in X$ .

Одним из возможных представлений модели классификации данных является дерево принятия решений. Дерево решений – это связный ациклический граф, представляющий правила классификации в иерархической последовательной структуре. Дерево состоит узлов и листов. В каждом узле расположено проверка по одному из атрибутов вектора  $x_i$ . В каждом листе расположено одно из значений меток класса  $C_1, C_2, \dots, C_k$ . Чтобы классифицировать любой из объектов обучающей выборки надо спуститься по дереву до листа и выдать значение метки класса листа. В

зависимости от значений атрибутов классифицируемого объекта спускаясь по дереву можно прийти к различным листам.

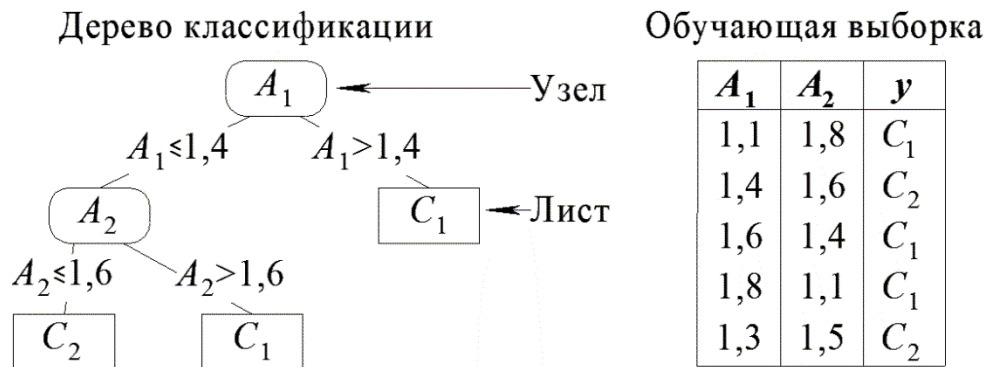


Рисунок 2.1 – Пример обучающей выборки и соответствующее ей дерево принятия решений

Общий принцип построению деревьев решений, заключается в рекурсивном разбиении множества объектов  $T$  из обучающей выборки на подмножества, содержащие объекты, относящиеся к одинаковым классам.

Относительно обучающей выборки  $T$  и множества классов  $C_1, C_2, \dots, C_k$  возможны три ситуации:

- множество  $T$  содержит один или более объектов, относящихся к одному классу  $C_i$ . Тогда дерево решений для  $T$  - это лист, определяющий класс  $C_i$ ;
- множество  $T$  не содержит ни одного объекта (пустое множество). Тогда данное множество  $T$  – это лист, и класс, ассоциированный с листом, выбирается из другого множества, отличного от  $T$ , например, из множества ассоциированного с родителем;
- множество  $T$  содержит объекты, относящиеся к разным классам. В этом случае следует разбить множество  $T$  на некоторые подмножества. Для этого выбирается один из атрибутов  $A_1, A_2, \dots, A_k$  и  $T$  разбивается на подмножества  $T_1, T_2, \dots, T_j$ , по выбранному атрибуту и соответствующему ему условию.



Эти шаги повторяются рекурсивно для всех получающихся подмножеств до тех пор, пока все подмножества не будут объявлены листьями.

Все алгоритмы автоматического построения деревьев принятия решений относятся к индуктивным алгоритмам машинного обучения (и к подобласти “обучения с учителем”).

Одним из важных вопросов построения дерева принятия решений является выбор таких условий в узлах, которые обеспечат наилучшее разбиение обучающей выборки на подмножества. Это необходимо, чтобы получившееся дерево было, во-первых компактным (содержало наименьшее количество узлов), а во-вторых обеспечивало максимальную точность классификации.

Алгоритм ID3 для построения дерева принятия решения выбирает из всех возможных вариантов такие проверки в узле, которые обеспечивают максимальное снижение информационной двоичной энтропии.

Рассмотрим этот вопрос подробнее на множестве  $T$  объектов, содержащих  $k$  меток класса –  $C_1, C_2, \dots, C_k$ . Если перемешать объекты из множества  $T$  и расположить их в ряд, то мы реализуем операцию под названием "перестановка". При этом считается, что объекты с одинаковыми метками класса идентичны. Тогда количество уникальных перестановок для множества рассчитывается, как:

$$W = \frac{N!}{N_1! N_2! \dots} = \frac{N!}{\prod_{i=1}^k N_i!}, \quad (2.3)$$

где  $N_i$  – количество объектов с меткой класса  $C_i$  в множестве  $T$ ,  $N$  – количество объектов в обучающей выборке.

Если все уникальные перестановки пронумеровать числами от 0 до  $W - 1$ , то количество бит необходимое для кодирования каждого уникального варианта перестановки равно  $\log_2(W)$ . Среднее количество бит на каждую перестановку называется комбинаторной энтропией и рассчитывается как:

$$S = \frac{\log_2(W)}{N} = \frac{1}{N} \cdot \log_2\left(\frac{N!}{\prod_{i=1}^k N_i!}\right) \quad (2.4)$$

Важное комбинаторной энтропии, которое используется при построении дерева принятия решения – чем меньше энтропия, тем однороднее множество. Самые однородные множества в дереве принятия решений – это листы не содержащие примесей других классов. Самое неоднородное множество – начальная обучающая выборка.

Для удобства расчета энтропии выражение, приведенное выше, можно упростить, воспользовавшись формулой Стирлинга:

$$\ln N! = N \cdot \ln N - N + O(\ln N) \approx N \cdot \ln N - N \quad (2.5)$$

Тогда энтропия для множества  $T$  с учетом упрощений можно рассчитать так:

$$S = -\sum_{i=1}^k \left(\frac{N_i}{N} \cdot \log_2 \frac{N_i}{N}\right) \quad (2.6)$$

С учетом того, что значение энтропии  $S$  зависит от рассматриваемого множества  $T$ , выразим ее как функцию:

$$S(T) = -\sum_{i=1}^k \left(\frac{\text{freq}(T, C_i)}{|T|} \cdot \log_2 \frac{\text{freq}(T, C_i)}{|T|}\right), \quad (2.7)$$

где  $\text{freq}(T, C_i)$  – функция, которая возвращает количество объектов сметкой класса  $C_i$  в множестве  $T$ .

При построении дерева принятия решения в каждый узел помещается условие, которое разбивает исходное множество  $T$  на несколько  $j$  подмножеств:

$$T = T_1 \cup T_2 \cup \dots \cup T_j \quad (2.8)$$

Тогда для каждого подмножества  $T_1, T_2, \dots, T_j$  по формуле 2.7 можно рассчитать энтропию.

Таким образом энтропию множества  $T$  после разбиения на подмножества рассчитывается как:

$$S_0(T, T_1, T_2, \dots, T_j) = \sum_{i=1}^j \left( \frac{|T_i|}{|T|} \cdot S(T_i) \right) \quad (2.9)$$

Если из энтропии  $S(T)$  вычесть энтропию разбиения  $S_0(T, T_1, T_2, \dots, T_j)$ , то можно рассчитать прирост информации  $G(T, T_1, T_2, \dots, T_j)$ , обеспечивающийся данным условием в узле:

$$G(T, T_1, T_2, \dots, T_j) = S(T) - \sum_{i=1}^j \left( \frac{|T_i|}{|T|} \cdot S(T_i) \right) \quad (2.10)$$

Чем однородней получаются подмножества в результате разбиения, тем больший прирост информации обеспечивает данное разбиение.

Таким образом, алгоритм ID3 на каждом шаге выполнения выбирает такую проверку в узле, которая обеспечивает максимальный прирост информации:

$$G(T, T_1, T_2, \dots, T_j) \rightarrow \max \quad (2.11)$$

Энтропия будет равна нулю, если в результате построения дерева принятия решения удалось достигнуть листов без содержания примесей посторонних классов.

Для каждого категориального атрибута существует только один вариант разбиения – когда исходное множество  $T$  делится на подмножества в количестве равном значений данного атрибута.

Для каждого числового атрибута существует несколько вариантов разбиений, количество которых равно мощности множества порогов. Множество порогов находится путем записи всех уникальных значений данного числового атрибута в обучающей выборке в порядке возрастания без повторов с отбрасыванием наибольшего значения. Затем каждое значение порога выступает в роли точки деления исходного множества  $T$ , состоящего из объектов  $t$ , на подмножества  $T_1$  и  $T_2$  (2.12):

$$\begin{cases} T_1 = \{ t \in T \mid a(t) \leq z \\ T_2 = \{ t \in T \mid a(t) > z \} \end{cases}, \quad (2.12)$$

где,  $a$  – значение атрибута, по которому производится разбиение.

В конечном итоге из всех возможных вариантов разбиения выбирается то, которое соответствует критерию (2.11). Как было сказано выше, построение дерева ведется до тех пор, пока не будут найдены все листья.

## 2.2 Контактная сварка как объект диагностики

Электрическая схема сварочной машины (рисунок 2.2), состоит из активного  $R_1$  и индуктивного  $X_1$  сопротивлений первичного контура, активного  $R_2$  и индуктивного  $X_2$  сопротивлений вторичного контура, тиристорного контактора  $TK$  в первичном контуре, сварочного трансформатора  $TC$  и сопротивления  $R_{\text{ээ}}$  свариваемых деталей, через которое протекает ток сварки  $i_2$ .

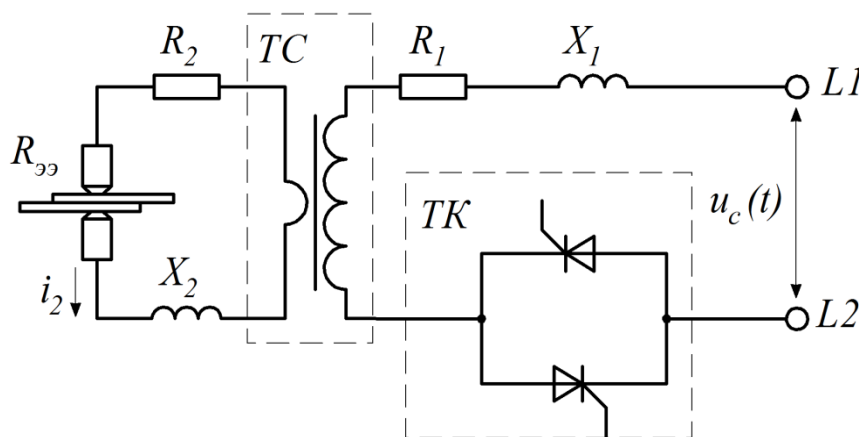


Рисунок 2.2 – Электрическая схема контактной сварочной машины

Сопротивление  $R_{\text{ээ}}$  участка “электрод-электрод”, как уже было отмечено авторами других работ, является одним из наиболее информативных диагностических признаков контактной сварки, т.к. позволяет определять на каком этапе находится процесс формирования литого ядра соединения [87].

Механизм формирования литого ядра соединения был связан с изменением сопротивления  $R_{\text{ээ}}$  (рисунок 2.3): так на коротком этапе I увеличивается площадь контакта заготовок, что вызывает уменьшение сопротивления; на этапе II и III происходит увеличение температуры и как результат – увеличение удельного сопротивления (динамическое сопротивление начинает возрастать); на IV этапе происходит зарождение

литого ядра соединения; после этого сокращающийся путь протекания тока ускоряет снижение сопротивления  $R_{ээ}$ .

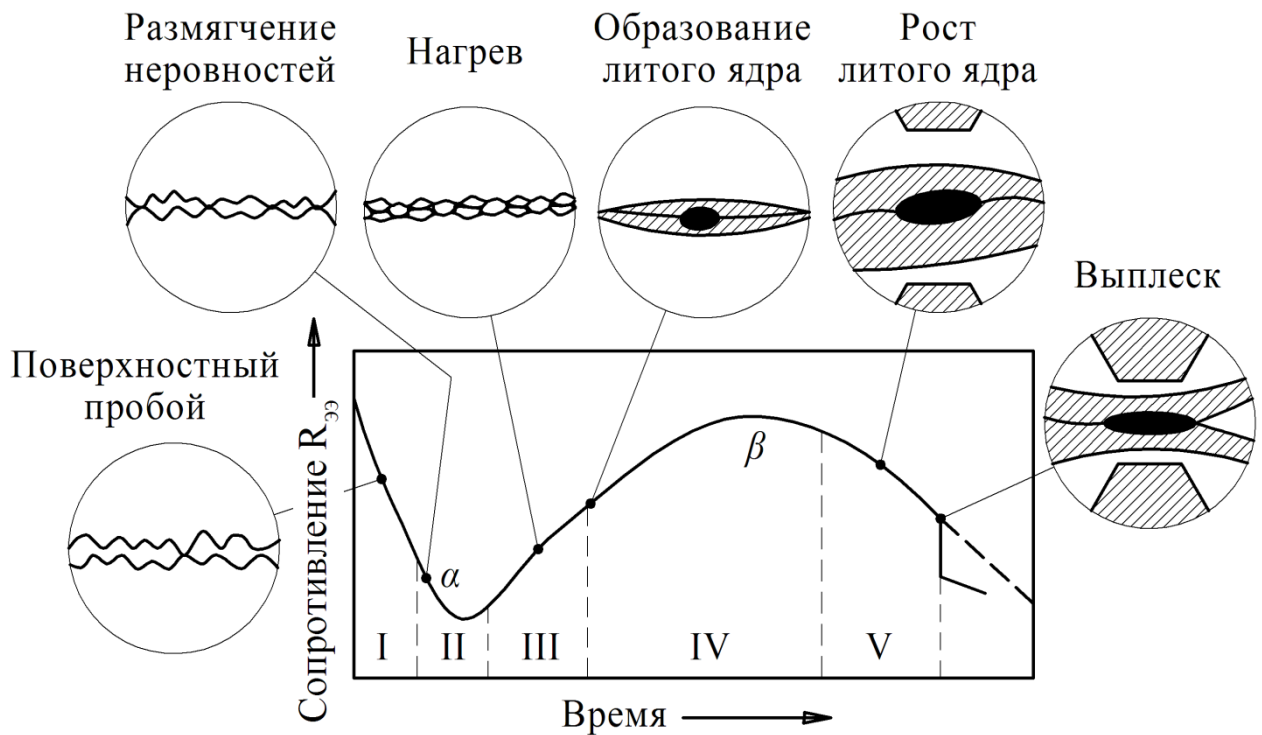


Рисунок 2.3 – Теоретическая кривая динамического сопротивления

Любые негативные (возмущающие) факторы, действующие на процесс контактной сварки (загрязнение поверхности заготовок, ток шунтирования, износ электродов) приводят к замедлению процесса формирования литого ядра (рисунок 2.3). Что эквивалентно сварки на меньших значениях сварочного тока  $I_2$ , чем необходимое значение.

Определять сопротивление  $R_{ээ}$  как путем измерения тока  $I_2$  и напряжения на электродах  $U_{ээ}$  не рационально, так как это:

- приводит к захламлению рабочего пространства сварочных клещей датчиками
- сложность фильтрации измеренных осциллограмм от электромагнитных помехи как следствие – сложная аппаратная реализация такого метода.

Однако из упрощенной математической модели электрических процессов в сварочной машине известно, что коэффициент мощности сварки вычисляется так:

$$\cos \varphi = \frac{R_{\text{эз}} + R_{2\kappa}}{\sqrt{(R_{\text{эз}} + R_{2\kappa})^2 + (X_{2\kappa})^2}}$$

В течении одной сварки электрические параметры сварочной машины ( $R_{2\kappa}$ ,  $X_{2\kappa}$ ) меняются не значительно, следовательно  $\cos \varphi$  в процессе сварки связано с изменением сопротивления  $R_{\text{эз}}$ .

Известно, что угол  $\varphi$  можно рассчитать путем решения равенства:

$$\sin(\lambda + \alpha - \varphi) / \sin(\alpha - \varphi) = e^{-\lambda \cdot \text{ctg} \varphi},$$

где  $\alpha$  – угол фазовой отсечки (в градусах),  $\lambda$  – длительность сварочного импульса (в градусах) (рисунок 2.4).

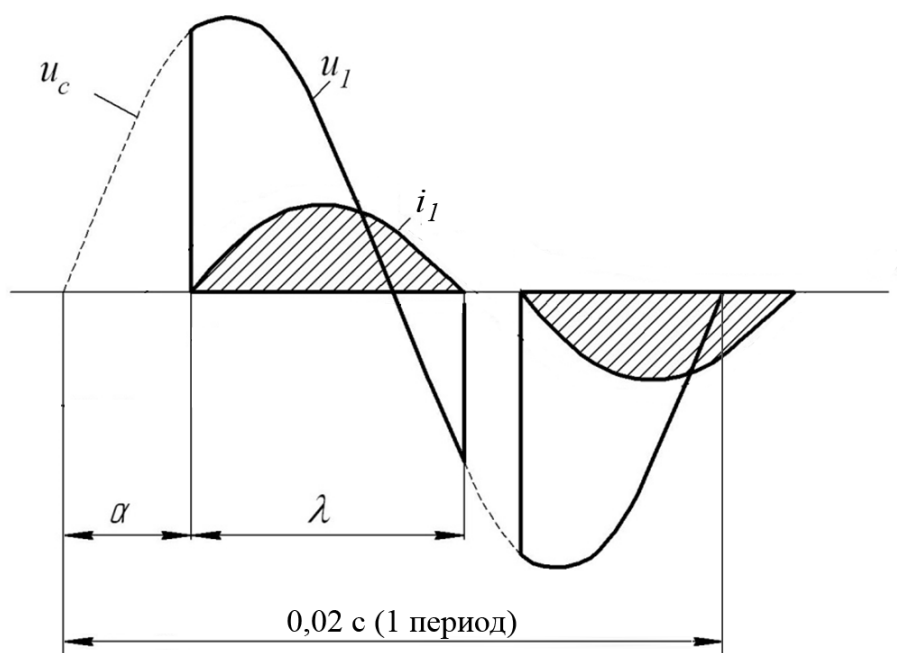


Рисунок 2.4 – Осциллограммы тока и напряжения со стороны первичной обмотки трансформатора

В процессе сварки можно управлять углом  $\alpha$ , но длительность  $\lambda$  – не регулируется, она зависит от сопротивления  $R_{\text{эз}}$ .

В качестве диагностического признака предлагается использовать  $\cos \varphi$ , который можно определить для каждого периода сварки, измеряя  $\alpha$  и  $\lambda$ .

стороны первичной обмотки. Причем нет необходимости измерять осциллограммы тока и напряжения.

Таким образом, каждую сварку будет описывать вектор значений  $\cos\varphi$ , количество компонентов которого равно количеству периодов сварки (рисунок 2.5).

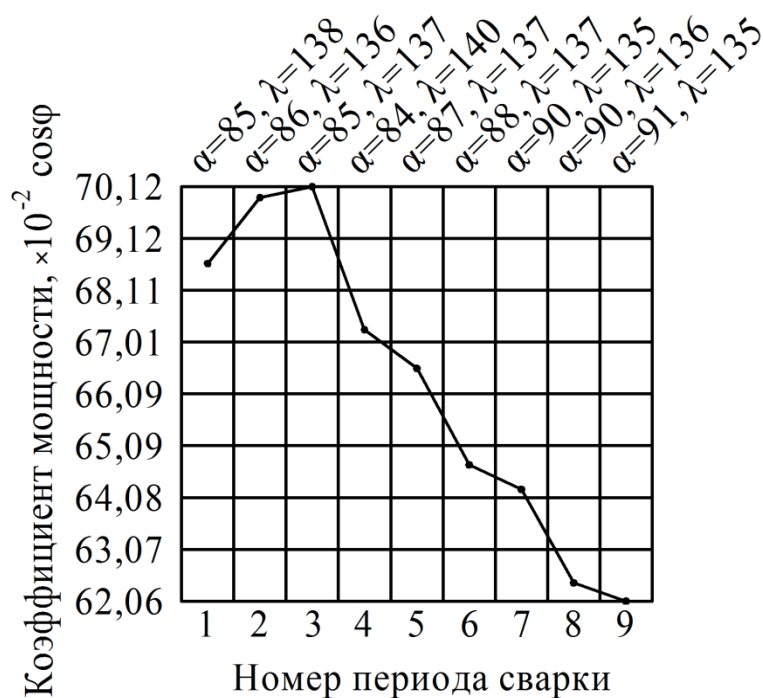


Рисунок 2.5 – Динамика изменения коэффициента мощности при сварке пакета заготовок из стали 08кп толщиной 0,8+0,8 мм

Так как нам для диагностики интересны не абсолютные значения коэффициента мощности, а динамика его изменения в процессе сварки, то вектор значений  $\cos\varphi$  для каждой сварки предложено нормировать к диапазону от 0 до 1 (от 0% до 100%) (рисунок 2.6).

Изучение особенностей полученных диаграмм в зависимости от влияния различных возмущений на процесс сварки и получаемых прочностных характеристик – слишком трудоемко, чтобы заниматься этим “вручную”. Поэтому принято решение возложить анализ получаемых диаграмм и поиск в них закономерностей на алгоритм ID3, строящий на выходе дерево принятия решений.

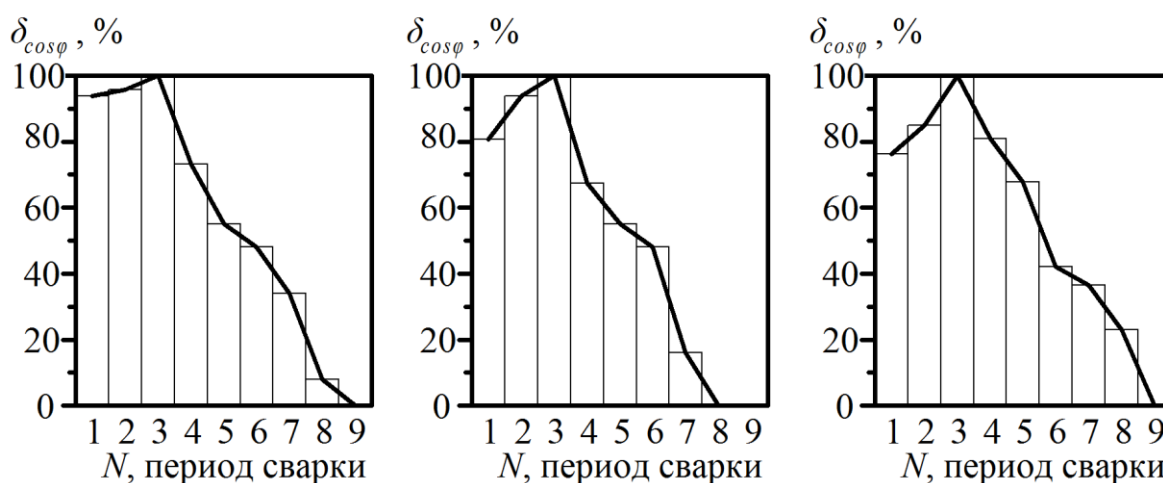


Рисунок 2.6 – Несколько примеров диаграмм изменения динамики  $\cos\varphi$  при сварке заготовок из стали 08кп толщиной 0,8+0,8 мм

### 2.3 Алгоритм диагностики сварки с использованием ID3

Для построения деревьев принятия решения проводится экспериментальная сварка заготовок в режиме стабилизации тока на различных значениях сварочного тока в диапазоне  $\pm 50\%$  от рекомендуемого значения тока сварки. При этом для каждого периода сварки измеряются характеристики импульса тока - фазовая задержка подачи импульса и длительность его протекания, по которым рассчитывают коэффициент мощности сварки  $\cos\varphi$ . Таким образом, для каждой экспериментальной сварки получают вектор значений  $\cos\varphi$  (количество компонентов вектора соответствует количеству периодов сетевого напряжения, при котором подавались импульсы тока). Вектор  $\cos\varphi$  от каждой экспериментальной сварки нормируется путем линейного преобразования к диапазону от 0% до 100% (где 0% соответствует наименьшему значению компоненты в векторе  $\cos\varphi$ , а 100% - соответствует наибольшему значению). Затем полученные в результате экспериментальной сварки соединения исследуются методами контроля, необходимыми для определения соответствия сварки выбранным критериям качества. После исследования всех экспериментальных сварок получают массив данных, состоящий для каждой экспериментальной сварки из: вектора нормированных значений  $\cos\varphi$  вида  $(P_1, P_2, \dots, P_n)$ , где



количество компонентов  $n$  равно количеству периодов сварки и метки, указывающей на соответствие или не соответствие данной сварки критериям качества (например, метка «С<sub>1</sub>» для некачественных сварок и метка «С<sub>2</sub>» для качественных сварок).

Затем полученный массив данных (также называемый обучающей выборкой) анализируется с помощью алгоритма машинного обучения ID3, для автоматического определения такого сочетания проверок по значениям компонентов  $P_1, P_2, \dots, P_n$ , которое позволяет отделить сварки с метками «С<sub>1</sub>» от сварок метками «С<sub>2</sub>». Результат работы данных алгоритмов можно представить в виде дерева принятия решений, который является классификатором качества сварок в зависимости от значений компонентов  $P_1, P_2, \dots, P_n$ .

#### **2.4 Экспериментальная апробация предложенного алгоритма**

Экспериментальная сварка проводилась на пакете заготовок из стали 08кп толщиной 0,8+0,8 мм. Ток сварки выбирался из диапазона 5,5 кА до 11 кА с шагом 0,5 кА. Длительность сварки – 0,18 с. Сварка в режиме стабилизации тока. Сварка проводилась под действием различных возмущений. Всего экспериментальных сварок – 60 штук. Критерием качества сварных соединений было выбрано выполнение следующих условий: отсутствие выплесков, диаметр литого ядра не менее 5,2 мм, сила разрушения на срез при статических испытаниях не менее 7 кН.

По полученным данным были построено дерево классификации качества сварки с использованием алгоритма ID3 (рисунок 2.7).

## Критерий качества:

- сила разрушения на срез не менее 7 кН
- отсутствие выплесков

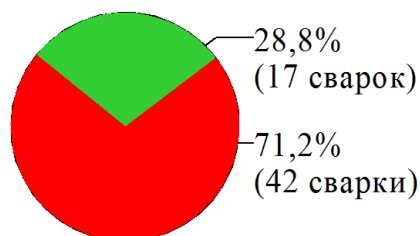
## Данные о процессе сварки:



1. Вектор значений  $(P_1, P_2, \dots, P_9)$ .
2. Метка качества сварки  $C_1$  или  $C_2$ .

## Состав данных:

- $C_1$  - качественная сварка
- $C_2$  - некачественная сварка



## Алгоритм ID3

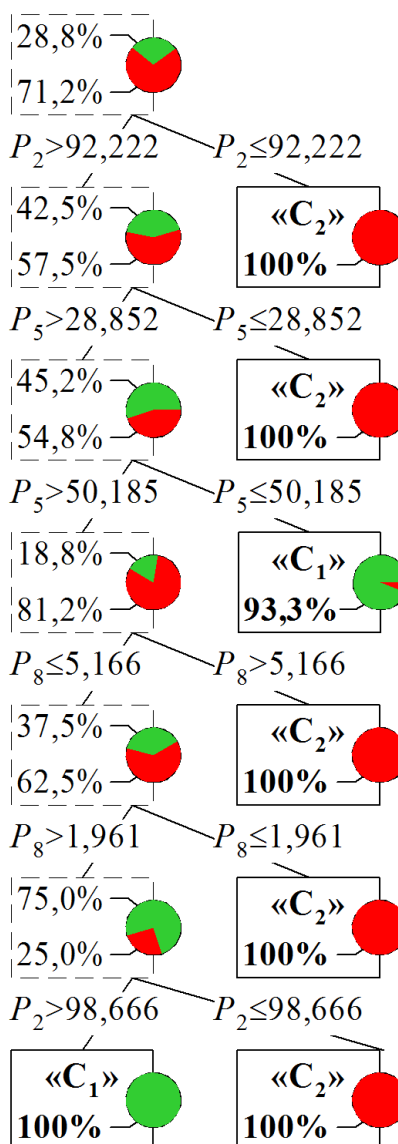


Рисунок 2.7 – Состав обучающей выборки и полученный классификатор качества

Точность работы полученного классификатора проверялась следующими методами: «cross-validation» (с разделением обучающей выборки на 5 частей), «cross-validation» (с разделением обучающей выборки на 6 частей), «leave-one-out», «randomsampling» (30 циклов),

«testontraindata». Результаты проверки точности дерева принятия решений представлены на рисунке 2.8.

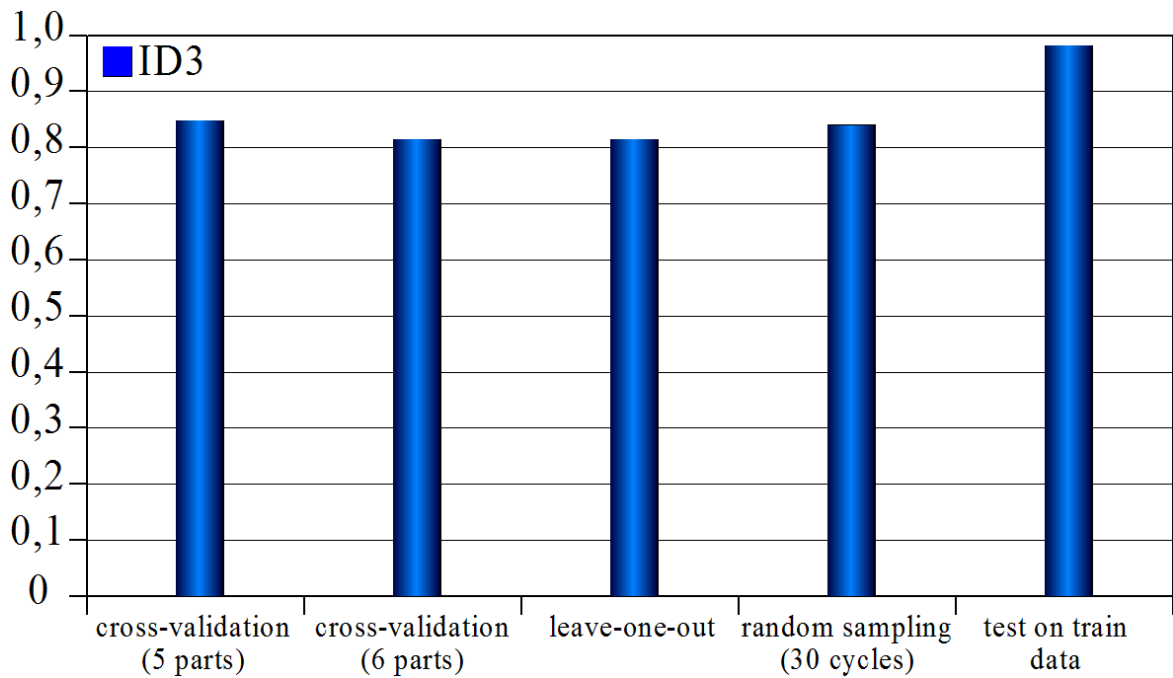


Рисунок 2.8 – Проверка точности работы классификатора

Хронология используемых проверок в дереве представлена на рисунке 2.9. Примечательно, что алгоритм ID3 самостоятельно определил, что наиболее важные с точки зрения диагностики атрибутами являются  $P_2$ ,  $P_5$ ,  $P_8$ .

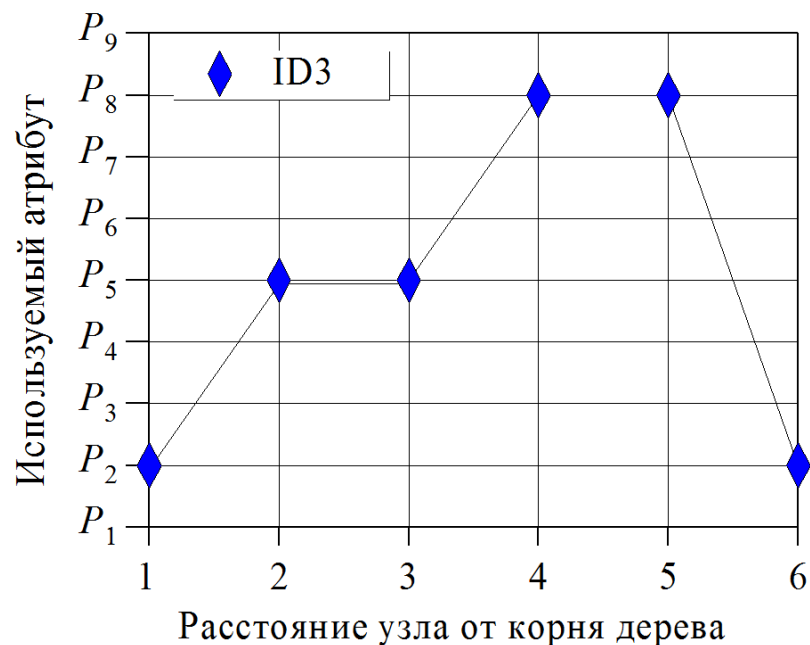


Рисунок 2.9 – Хронология проверок в узлах дерева

Рекомендации по увеличению точности работы деревьев принятия решений:

1) Увеличение количества экспериментальные сварок в значительной степени влияет на точность классификации получаемого дерева принятия решений.

2) При экспериментальной сварке выбор значений тока с меньшим шагом позволяет увеличить разнообразность обучающей выборки, тем самым повышая точность классификации получаемого дерева принятия решений.

3) Увеличение длительности сварки приводит к увеличению количества компонент вектора нормированных значений  $\cos\varphi$ . Это приводит к увеличению количества вариантов проверок в дереве принятия решений, что приводит к увеличению точности классификации.

### 3 Практическая реализация предложенных решений

#### 3.1 Программная реализация алгоритма

В ходе выполнения исследований по выпускной квалификационной работе в интегрированной среде разработки NetBeans на языке программирования Java была спроектирована, разработана и спроектирована программа для построения моделей классификации со следующими характеристиками:

- наличие графического интерфейса пользователя у программы;
- использование алгоритма машинного обучения ID3 для автоматизированного построения классификатора качества сварки;
- графическое отображение построенной модели в виде дерева принятия решения в виде связного ациклического графа (дерева);
- возможность использования в обучающей выборке до 255 входных числовых параметров;
- возможность использование в обучающей выборке до 25 различных классов;
- функция автоматического импорта обучающей выборки из Excel-файла (с определением количества атрибутов и классов);
- функция редактирование обучающей выборки в окне программы;
- возможность просмотра характеристик классификатора на каждой итерации его обучения по алгоритму ID3: контроль распределения объектов обучающей выборки по узлам, просмотр рассчитанных значений прироста информации в каждом узле дерева.

Файловая структура проекта в среде NetBeans показана на рисунке 3.1.

UML диаграммы классов программы представлены в приложении А.

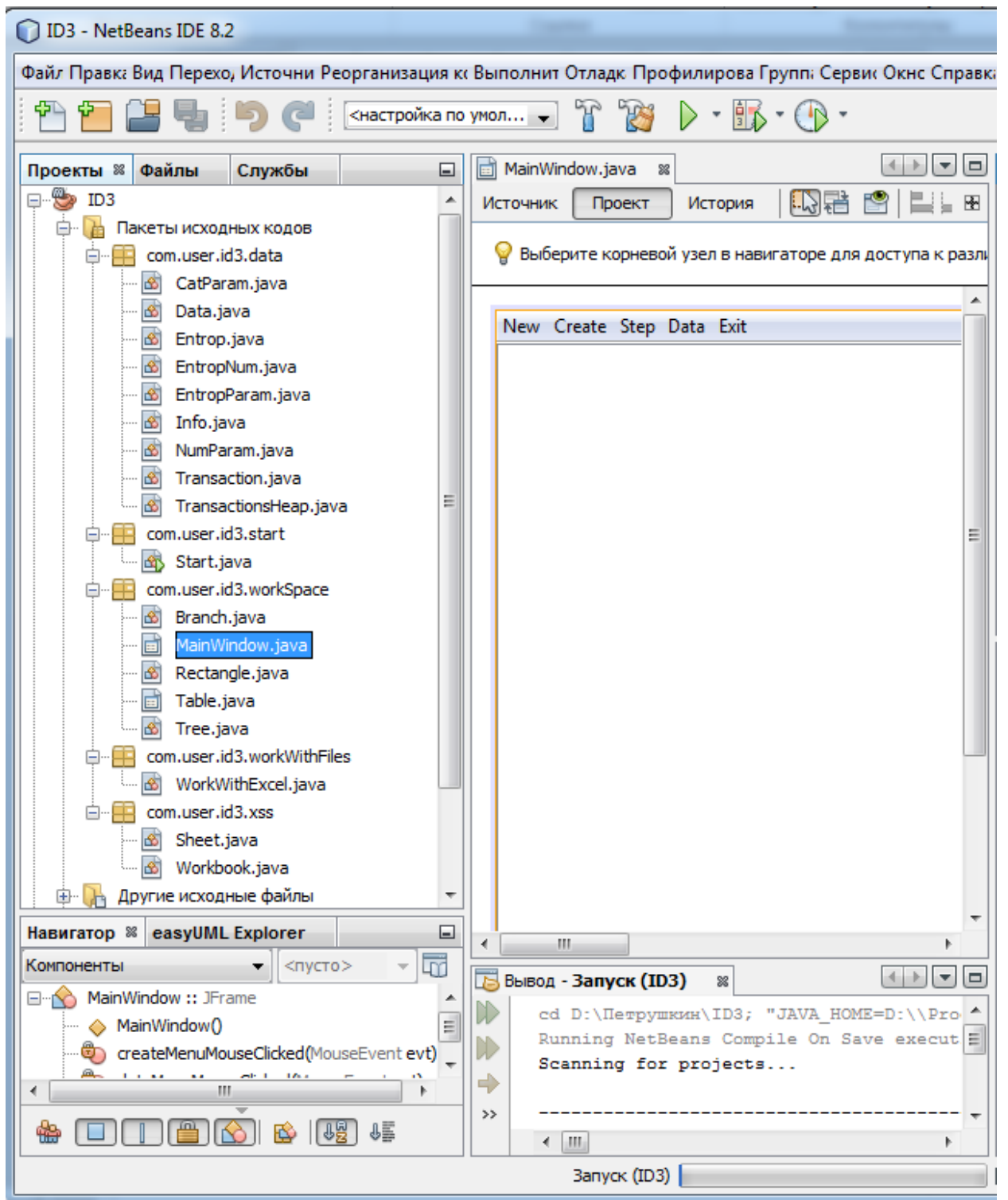


Рисунок 3.1 – Проект приложения в среде NetBeans

При запуске приложения пользователь увидит главную форму программы. Оно состоит из двух основных интерактивных элементов. В верхней части находится меню для управления программой и данными, в

нижней части находится область построения модели классификации и ее отображения в виде графа (дерева) (рисунок 3.2).

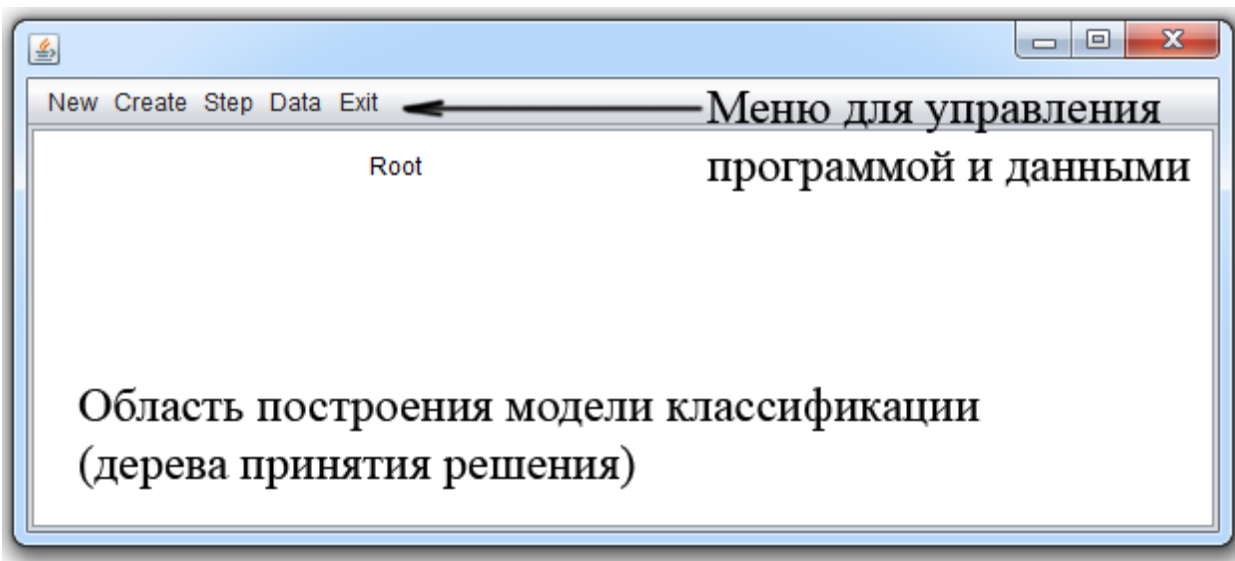


Рисунок 3.2 – Главное окно программы

Пункт меню «New» предназначен для очистки всех переменных хранящихся в памяти программы, а также для очистки области содержащей графическое представление дерева принятия решений.

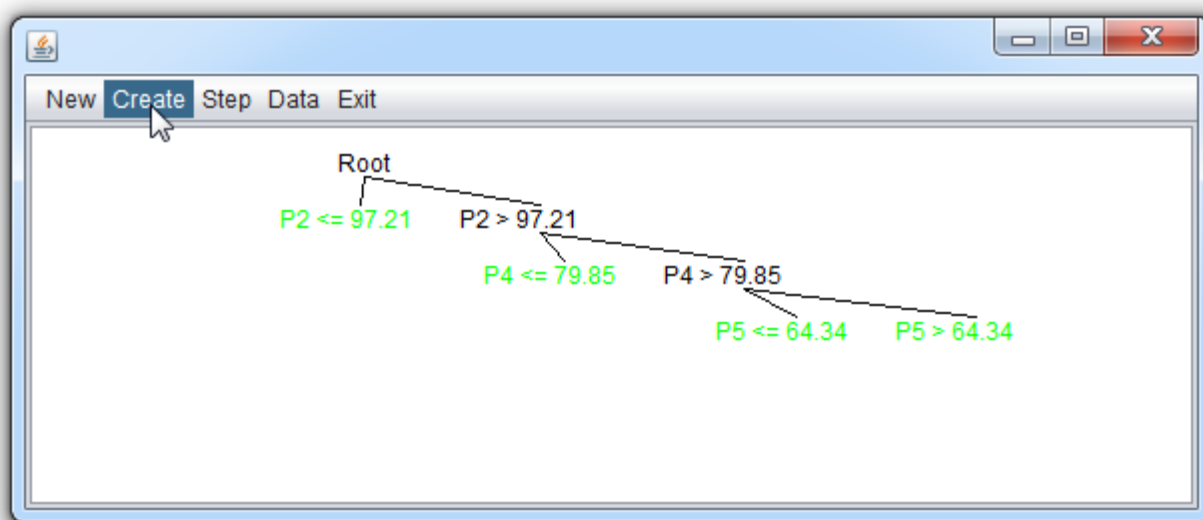


Рисунок 3.3 – Дерево решений, построенное по обучающей выборке

Пункт меню «Create» предназначен для построения полного дерева принятия решений в соответствии с алгоритмом ID3 и загруженной в программу обучающей выборки. При этом зеленым цветом показаны листья

(множества, не содержащие примесей других меток класса). Пример показан на рисунке 3.3.

Пункт меню «Step» предназначен для пошагового просмотра работы алгоритма ID3 с обучающей выборкой.

Пункт меню «Data» предназначен для просмотра текущей обучающей выборки и для ее редактирования. При выборе данного пункта откроется окно такого вида, как это показано на рисунке 3.4.

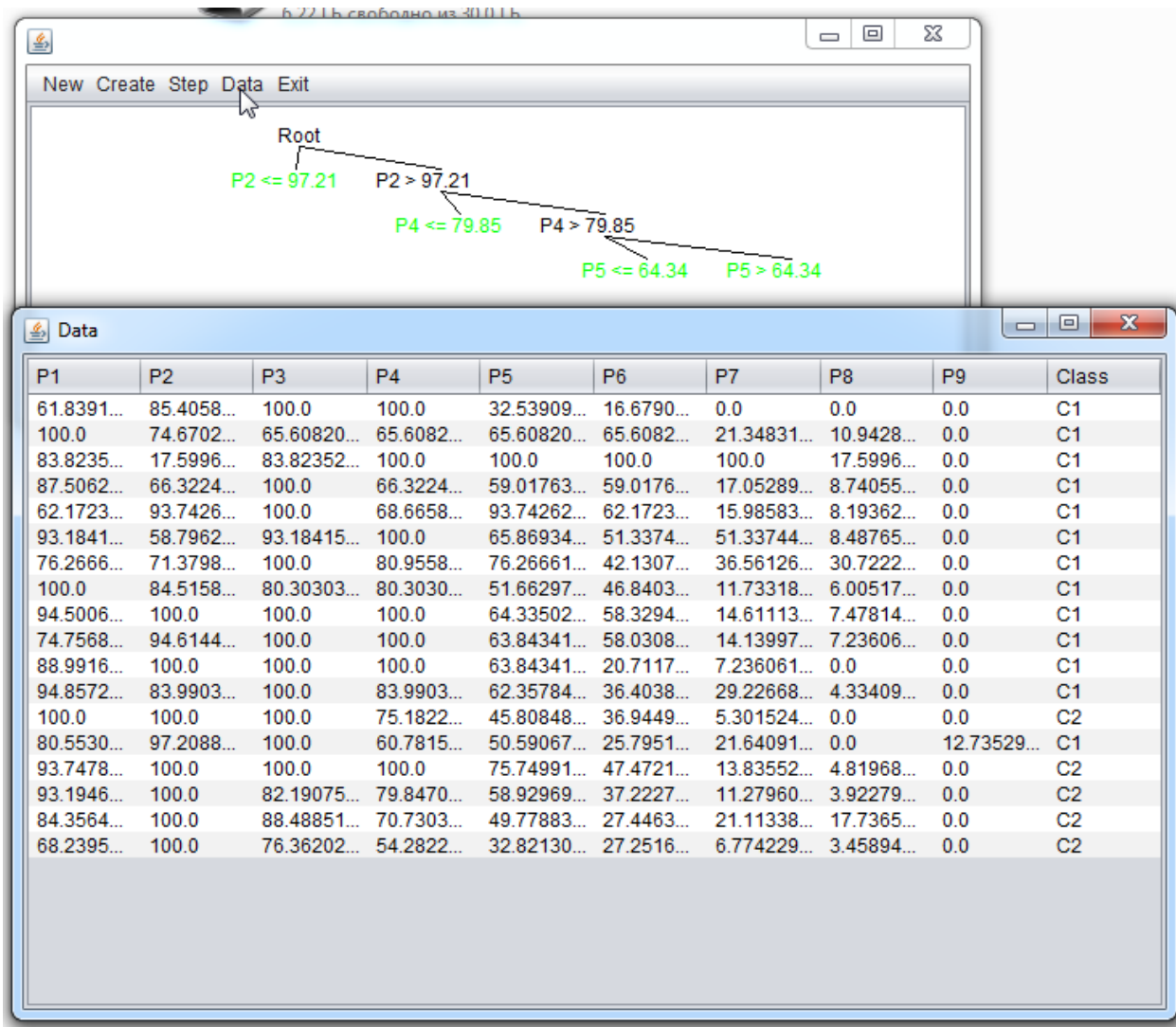


Рисунок 3.4 – Просмотр и редактирование обучающей выборки

Если рядом с jar-файлом программы будет расположен excel-файл с именем ID3.xlsx, то обучающая выборка будет автоматически импортирована от туда. При этом определение количества атрибутов и классов происходит автоматически (рисунок 3.5).



Имя	Дата изменения	Тип	Размер
lib	24.03.2017 15:14	Папка с файлами	
src	24.03.2017 15:14	Папка с файлами	
target	24.03.2017 15:14	Папка с файлами	
ID3.xlsx	02.04.2017 23:11	Лист Microsoft Ex...	11 КБ
ID3-4.xlsx	31.03.2017 17:48	Лист Microsoft Ex...	15 КБ
nbactions.xml	24.03.2017 15:20	Документ XML	2 КБ
nbactions-release-profile.xml	13.02.2017 9:29	Документ XML	2 КБ
pom.xml	24.03.2017 15:20	Документ XML	2 КБ

Рисунок 3.5 – Расположение.xlsx-файла с обучающей выборкой

Пункт меню «Exit» предназначен для выхода из программы.

Когда классификатор в виде дерева принятия решений построен, имеется возможность просмотра его ключевых характеристик. Так для просмотра распределения объектов обучающей выборки по узлам дерева необходимо кликнуть по интересующему узлу левой кнопкой мыши. При этом откроется такое окно, как это показано на рисунке 3.6.

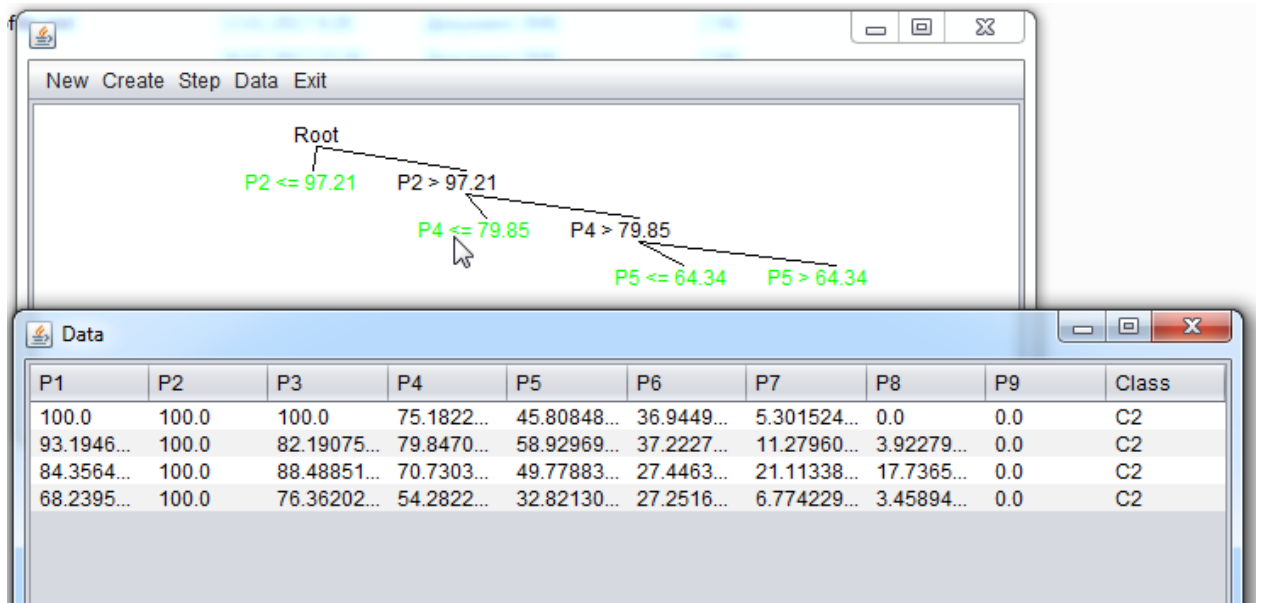


Рисунок 3.6 – Просмотр объектов обучающей выборки в определенном узле

Для просмотра рассчитанных значений прироста информации в каждом узле дерева необходимо нажать по нужному узлу правой кнопкой мыши (рисунке 3.7).

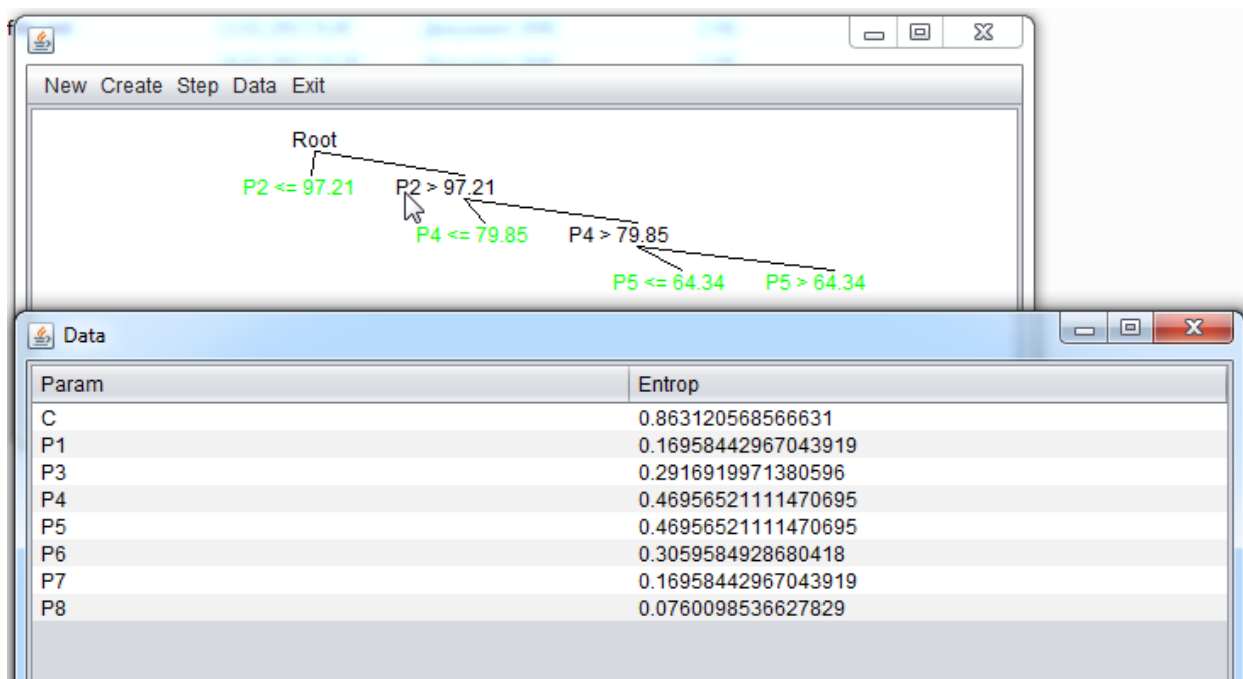


Рисунок 3.7 – Просмотр рассчитанных значений прироста информации в каждом узле

### 3.2 Пример работы программы

Рассмотрим пример работы программы.

Создадим обучающую выборку для построения классификатора качества сварки. Для этого занесем экспериментально полученные данные от 60 сварок пакета заготовок пластин толщиной 0,8 + 0,8 мм из стали 08кп в Excel-файл. Критериями качества выберем следующие характеристики соединения:

- сила разрушения соединения на срезе менее 7кН;
- отсутствие выплесков.

Для соединений удовлетворяющим критериям выставим метку класса «С1», для соединений не удовлетворяющим критериям – «С2». Тогда Excel-файл для 60 сварок будет выглядеть так, как это показано на рисунке 3.8. Значения атрибутов  $P_1, P_2, \dots, P_9$  объяснены во второй главе.

	A	B	C	D	E	F	G	H	I	J	K
1	№	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$	$P_7$	$P_8$	$P_9$	Class
2	1	61,839166	85,405808	100	100	32,539092	16,679077	0	0	0	C1
3	2	100	74,670249	65,608207	65,608207	65,608207	65,608207	21,348315	10,94284	0	C1
4	3	83,823529	17,59962	83,823529	100	100	100	100	17,99962	0	C1
5	4	87,506297	66,322418	100	66,322418	59,017632	59,017632	17,052897	8,740554	0	C1
6	5	62,172373	93,742621	100	68,66588	93,742621	62,172373	15,985832	8,193625	0	C1
7	6	93,184156	58,796296	93,184156	100	65,869342	51,337449	51,337449	8,487654	0	C1
8	7	76,266619	71,379806	100	80,955803	76,266619	42,130794	36,561265	30,72224	0	C1
9	8	100	84,515891	80,30303	80,30303	51,662971	46,840355	11,733186	6,005174	0	C1
10	9	94,50069	100	100	100	64,335021	58,329498	14,611137	7,478141	0	C1
11	10	74,756821	94,614472	100	100	63,843416	58,030842	14,139976	7,236062	0	C1
12	11	88,991696	100	100	100	63,843416	20,711744	7,2360617	0	0	C1
13	12	94,857215	83,990397	100	83,990397	62,357847	36,403841	29,226687	4,334091	0	C1
14	13	100	100	100	75,18224	45,808482	36,944997	5,3015242	0	0	C2
15	14	80,553031	97,20888	100	60,781514	50,590679	25,795145	21,640919	0	12,7353	C1
16	15	93,747893	100	100	100	75,749916	47,472194	13,835524	4,819683	0	C2
17	16	93,194636	100	82,190751	79,847099	58,92969	37,22271	11,279609	3,922797	0	C2
18	17	84,356457	100	88,48851	70,730392	49,778833	27,446326	21,113389	17,73654	0	C2
19	18	68,239584	100	76,362029	54,282241	32,821306	27,25163	6,7742292	3,458946	0	C2
20	19	100	95,294259	95,294259	78,553376	39,559514	28,595499	6,9081719	3,526297	0	C2
21	20	93,33677	100	100	63,577781	36,731538	30,622503	7,3978605	3,77626	0	C2
22	21	84,149856	100	88,184438	70,294835	49,323875	27,089337	20,982044	17,73443	0	C2
23	22	94,933717	100	85,75686	54,773405	30,531292	25,218374	22,392354	6,381667	0	C2
24	23	93,777874	100	86,892201	53,714783	29,794409	24,921136	5,8631567	0	0	C2
25	24	83,878973	100	76,768159	17,285482	14,870669	14,870669	14,870669	14,87067	0	C1
26	25	89,310138	100	22,027594	17,780444	17,780444	17,780444	17,780444	17,78044	0	C1
27	26	93,963824	100	83,679967	60,592101	55,091118	13,088544	10,825825	10,82582	0	C1
28	27	100	95,511991	81,998209	50,980197	42,899791	25,922977	20,897602	5,761767	0	C1

Рисунок 3.8 – Excel-файл с обучающей выборкой

Загрузим обучающую выборку в программу, используя автоматический импорт данных из Excel. Проверим правильность импорта, кликнув по пункту Data главного меню (рисунок 3.9).

P1	P2	P3	P4	P5	P6	P7	P8	P9	Class
61.8391...	85.4058...	100.0	100.0	32.5390...	16.6790...	0.0	0.0	0.0	C1
100.0	74.6702...	65.6082...	65.6082...	65.6082...	65.6082...	21.3483...	10.9428...	0.0	C1
83.8235...	17.5996...	83.8235...	100.0	100.0	100.0	100.0	17.5996...	0.0	C1
87.5062...	66.3224...	100.0	66.3224...	59.0176...	59.0176...	17.0528...	8.74055...	0.0	C1
62.1723...	93.7426...	100.0	68.6658...	93.7426...	62.1723...	15.9858...	8.19362...	0.0	C1
93.1841...	58.7962...	93.1841...	100.0	65.8693...	51.3374...	51.3374...	8.48765...	0.0	C1
76.2666...	71.3798...	100.0	80.9558...	76.2666...	42.1307...	36.5612...	30.7222...	0.0	C1
100.0	84.5158...	80.3030...	80.3030...	51.6629...	46.8403...	11.7331...	6.00517...	0.0	C1
94.5006...	100.0	100.0	100.0	64.3350...	58.3294...	14.6111...	7.47814...	0.0	C1
74.7568...	94.6144...	100.0	100.0	63.8434...	58.0308...	14.1399...	7.23606...	0.0	C1
88.9916...	100.0	100.0	100.0	63.8434...	20.7117...	7.23606...	0.0	0.0	C1
94.8572...	83.9903...	100.0	83.9903...	62.3578...	36.4038...	29.2266...	4.33409...	0.0	C1
100.0	100.0	100.0	75.1822...	45.8084...	36.9449...	5.30152...	0.0	0.0	C2
80.5530...	97.2088...	100.0	60.7815...	50.5906...	25.7951...	21.6409...	0.0	12.7352...	C1
93.7478...	100.0	100.0	100.0	75.7499...	47.4721...	13.8355...	4.81968...	0.0	C2
93.1946...	100.0	82.1907...	79.8470...	58.9296...	37.2227...	11.2796...	3.92279...	0.0	C2
84.3564...	100.0	88.4885...	70.7303...	49.7788...	27.4463...	21.1133...	17.7365...	0.0	C2
68.2395...	100.0	76.3620...	54.2822...	32.8213...	27.2516...	6.77422...	3.45894...	0.0	C2
100.0	95.2942...	95.2942...	78.5533...	39.5595...	28.5954...	6.90817...	3.52629...	0.0	C2
93.3367...	100.0	100.0	63.5777...	36.7315...	30.6225...	7.39786...	3.77625...	0.0	C2
84.1498...	100.0	88.1844...	70.2948...	49.3238...	27.0893...	20.9820...	17.7344...	0.0	C2
94.9337...	100.0	85.7568...	54.7734...	30.5312...	25.2183...	22.3923...	6.38166...	0.0	C2
93.7778...	100.0	86.8922...	53.7147...	29.7944...	24.9211...	5.86315...	0.0	0.0	C2
83.8789...	100.0	76.7681...	17.2854...	14.8706...	14.8706...	14.8706...	14.8706...	0.0	C1
89.3101...	100.0	22.0275...	17.7804...	17.7804...	17.7804...	17.7804...	17.7804...	0.0	C1
93.9638...	100.0	83.6799...	60.5921...	55.0911...	13.0885...	10.8258...	10.8258...	0.0	C1
100.0	95.5119...	81.9982...	50.9801...	42.8997...	25.9229...	20.8976...	5.76176...	0.0	C1
93.3750...	65.5972...	100.0	65.5972...	58.6854...	16.2493...	8.32029...	0.0	0.0	C1
100.0	81.2002...	76.5225...	76.5225...	47.8022...	42.4260...	26.8040...	6.44307...	0.0	C1

Рисунок 3.9 – Данные импортированные из Excel-файла

Построим классификатор качества сварки по импортированной обучающей выборке с использованием алгоритма ID3. Полученное в данном случае дерево принятия решения представлено на рисунке 3.10.

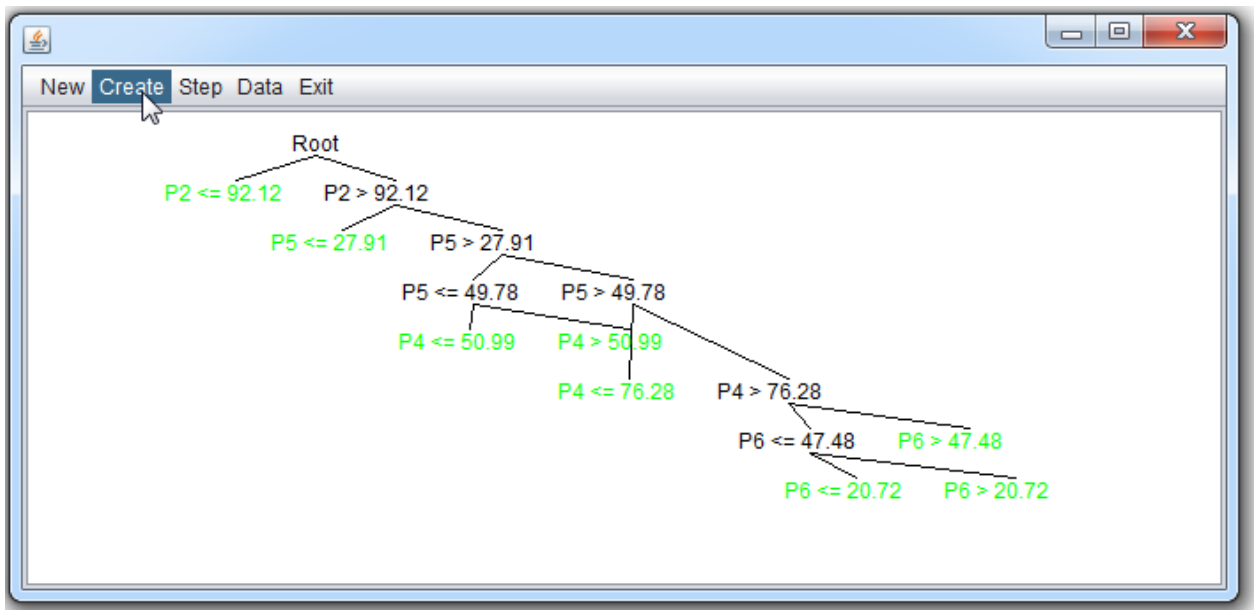


Рисунок 3.10 – Дерево принятия решений, построенное с использованием импортированной обучающей выборки

При необходимости можно просмотреть наиболее важные характеристики классификатора: распределения объектов обучающей выборки по узлам, просмотр рассчитанных значений прироста информации в каждом узле дерева.

Построенный классификатор можно использовать для прогнозирования качества сварок, не входящих в обучающую выборку. Исследование точности полученного классификатора представлено во второй главе.

## Заключение

Основные выводы по проведенным в ВКР исследованиям:

1. На основе изучения научных статей по теме исследования установлено, что совершенствование систем диагностики качества контактной сварки возможно за счет применения в них алгоритмов машинного обучения. Использование машинного обучения позволяет автоматизировать процесс разработки алгоритма диагностики, который строится на основе анализа набора частных случаев сварки.

2. Показано, что диагностику сварки можно привести к задаче классификации. При этом для простоты модели достаточно ввести два целевых класса: “качественная сварка” – куда будут относиться сварные соединения с необходимыми прочностными характеристиками “некачественная сварка” для соединений с неудовлетворительными характеристиками.

3. Доказано, что модель классификации качества сварки может быть построена автоматически с использованием алгоритмов построения деревьев принятия решений (в работе рассмотрен пример использования именно алгоритма ID3). Пример полученной модели (виде дерева принятия решений) приведен во второй главе. При этом входными параметрами модели являются нормированные значения  $P_i$  коэффициента мощности сварки  $\cos\varphi$ , который рассчитывается для каждого  $i$ -го периода сварки. Таким образом, каждая сварка описывается вектором  $P = (P_1, P_2, \dots, P_i)$ . Выходным параметром модели номер одного из двух классов  $y \in C_1, C_2$ .

4. С использованием технологического оборудования лаборатории “Контактная сварка” института машиностроения ТГУ было проведено около 100 экспериментальных сварок пакета заготовок из стали 08кп. Часть данных была использована в роли обучающей выборки для построения модели классификации качества сварки. Другая часть использовалась для оценки точности полученной модели. Вопрос точности модели подробно изучался во

второй главе. Полученные результаты: точность классификации по методу cross-validation с разделением на 5 частей равна 88%, по методу leave-one-out – 82%, по методу randomsampling при 30 циклах 83%. При тестировании модели на собственной обучающей выборке точность классификации – 98%

5. С использованием свободно-распространяемой среды разработки NetBeans на языке программирования Java разработано программное обеспечение, на практике реализующее предложенные подходы по построению моделей классификации качества сварки на основе анализа обучающей выборки.

6. Исследования, проведенные в рамках данной работы, будут использованы при выполнении гранта РФФИ на тему “Управление металлургическими процессами с использованием искусственного интеллекта”

## Список используемой литературы

1. Sammut, C. Encyclopedia of Machine Learning / Claude Sammut, Geoffrey I. Webb. – Springer US, 2010. – 1031 p.
2. Laurinen, P. Studying the Quality of Resistance Spot Welding Joints Using Bayesian Networks / P. Laurinen, H. Junno, L. Tuovinen, J. Röning // Proceedings of Artificial Intelligence and Applications. – 2004. – С. 705–711.
3. Aravinthan, A. A neural network system for spot weld strength prediction / A. Aravinthan, K. Sivayoganathan, D. Al-Dabass, V. Balendran // UKSIM2001: Conference Proceedings of the UK Simulation Society. - 2001. - P. 156-160
4. Патон, Б.Е. Оценка качества контактной точечной сварки с помощью нейронных сетей / Б.Е. Патон, Н.В. Подола, В.С. Гавриш // Автоматическая сварка. – 1998. – № 12. – С. 3–10.
5. Скачков, И.О. К вопросу применения нейронных сетей для контроля качества сварных соединений при подводной сварке / И.О. Скачков, А.Е. Пирумов, С.Ю. Максимов, Е.А. Прилипко // Автоматическая сварка. - 2006. - №6. - С. 27-31.
6. Park, Y.J. Quality evaluation by classification of electrode force patterns in the resistance spot welding process using neural networks / Y.J. Park, H. Cho // Proceedings of the Institution of Mechanical Engineers B, Journal of Engineering Manufacture. – 2004. – С. 151–162.
7. Huang, D. Intelligent Computing in Signal Processing and Pattern Recognition / De-Shuang Huang, Kang Li, George William Irwin. – Springer Berlin Heidelberg, 2006 – 1182 p.
8. Singal, P.K. Incorporation of Fuzziness in ID3 and Generation of Network Architecture Neural Computing & Applications / Pawan K. Singal, Sushmita Mitra, Sankar K. Pal // Neural Computing & Applications, 2010. – №10(2). – pp. 155-164
9. Generation of User Interest Ontology Using ID3 Algorithm in the Social Web / Jong-Soo Sohn, Qing Wang, In-Jeong Chung // IT Convergence and Security 2012. Lecture Notes in Electrical Engineering, 2012. - pp 1067-1074



10. Wang, J. Investigation on AQ11, ID3 and the principle of discernibility matrix / J. Wang , J. Cui, K. J. Comput Zhao // Journal of Computer Science and Technology, 2001. – №16(1). – pp. 1-12
11. Rathore, S.S A decision tree logic based recommendation system to select software fault prediction techniques / S.S Rathore, S. Kumar // Computing, 2017. – №99(3). – pp. 255-285
12. Tanha, J. Semi-supervised self-training for decision tree classifiers / J. Tanha, M. Someren, H. Afsarmanesh // International Journal of Machine Learning and Cybernetics, 2017. – №8(1). – pp. 355-370
13. Deepa, N. Multi-class classification using hybrid soft decision model for agriculture crop selection / N. Deepa, K. Ganesa // Neural Computing and Applications, 2016. – pp. 1-14
14. Liu, H. Rule-based systems: a granular computing perspective / H. Liu, A. Gegov, M. Cocea // Granular Computing, 2016. – №1(4). – pp. 259-274
15. Liu, H. Granular computing-based approach for classification towards reduction of bias in ensemble learning / H. Liu, M. Cocea // Granular Computing, 2016. – pp. 1-9
16. Chang, M. Mining unexpected patterns using decision trees and interestingness measures: a case study of endometriosis / M.-Y. Chang, R.-D. Chiang, S.-J. Wu, Ch.-H. Chan // Soft Computing, 2016. – №20(10). – pp. 3991-4003.
17. Wu, Ch.-Ch. Decision tree induction with a constrained number of leaf nodes // Chia-Chi Wu, Yen-Liang Chen, Yi-Hung Liu, Xiang-Yu Yang // Applied Intelligence, 2016. – №45(3). – pp. 673-685
18. Saloot, M.A. Hadith data mining and classification: a comparative analysis / M.A. Saloot, N. Idris, R. Mahmud [et al.] // Artificial Intelligence Review, 2016. – №46(1). – pp. 113-128
19. Caiuta, R. Meta-learning based selection of software reliability models / R. Caiuta, A. Pozo, S.R. Vergilio // Automated Software Engineering, 2016. – pp. 1-28

20. Chen, N. Financial credit risk assessment: a recent review / N. Chen, B. Ribeiro, A. Chen // *Artificial Intelligence Review*, 2016. – №45(1). – pp. 1-23