

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»

02.03.03 МАТЕМАТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ И АДМИНИСТРИРОВАНИЕ  
ИНФОРМАЦИОННЫХ СИСТЕМ

ТЕХНОЛОГИЯ ПРОГРАММИРОВАНИЯ

### **БАКАЛАВРСКАЯ РАБОТА**

на тему « Разработка регрессионной модели прогнозирования прочности  
сварных соединений с использованием деревьев решений и алгоритмов  
машинного обучения »

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

**Допустить к защите**  
Заведующий кафедрой к.тех.н., доцент, А.В. Очеповский \_\_\_\_\_

« \_\_\_\_\_ » \_\_\_\_\_ 2017 г.

Тольятти 2017

## **Аннотация**

Тема: «Разработка регрессионной модели прогнозирования прочности сварных соединений с использованием деревьев решений и алгоритмов машинного обучения».

В данной выпускной квалификационной работе исследуются пути совершенствования методов диагностики сварки. В исследовании рассматривается задача прогнозирования качества сварки с использованием алгоритма CART. При этом предложено наблюдать за изменением коэффициента мощности сварки и прогнозировать силу разрушения 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 математический аппарат алгоритма CART. Затем разрабатывается алгоритм диагностики сварки на основе данного алгоритма.

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

В заключении представлены выводы по проделанной работе.

Данная бакалаврская работа состоит из пояснительной записки на 55 стр., включая 24 рисунка, 1 таблицу, списка 22 источника, в том числе 5 источников на иностранном языке и 1 приложения.

## **ABSTRACT**

The title of the graduation work is “Development of a Regression Forecasting Model of Durability of Welded Connections with Use of Decision Trees and 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 "CART" is considered. At the same time it is offered to monitoring change of power factor of welding and to predict connection corrupting force of welded joint. In work results of approbation of an algorithm in practice are presented.

The structure of the graduation work is represented by an introduction, three parts, a conclusion and a list of references.

The introduction describes the relevance of the conducted research, the goal and the tasks.

In the first part review of methods of analysis of experimental data of resistance welding by means of algorithms of machine learning is provided.

The second part the mathematical apparatus of an algorithm of CART is described. Then the algorithm of diagnostics of welding on the basis of this algorithm is developed.

In the third part, 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 an explanatory note on 55 pages, including 24 figures, 1 tables, the list of 22 references including 5 foreign sources and 1 appendix.

## Оглавление

Введение.....	3
1 Анализ состояния вопроса .....	6
1.1 Обзор предметной области .....	6
1.2 Обзор исследований по применению алгоритмов машинного обучения для диагностики качества сварки .....	9
2 Разработка алгоритма диагностики сварки, основанного на методе CART .....	22
2.1 Алгоритм CART как метод построения регрессионной модели.....	22
2.2 Основные параметры контактной сварки.....	26
2.3 Алгоритм диагностики сварки с использованием алгоритма CART .....	30
3 Практическая реализация предложенных решений .....	31
3.1 Моделирование построения CART в среде Matlab .....	31
3.2 Практическая апробация алгоритма и исследование точности полученной регрессионной модели .....	32
Заключение .....	42
Список используемой литературы .....	44
Приложение А. Регрессионная модель, полученная с помощью алгоритма CART .....	47

## Введение

Контактная сварка используется в промышленности как в качестве основного технологического процесса соединения листовых деталей. Такой тип сварки распространен в автомобилестроении, авиационной промышленности, вагоностроении, судостроении.

Преимуществами такого типа сварки является высокая производительность, необходимая для массового и крупносерийного производства. Контактная точечная сварка представлена на рисунке 1.1.

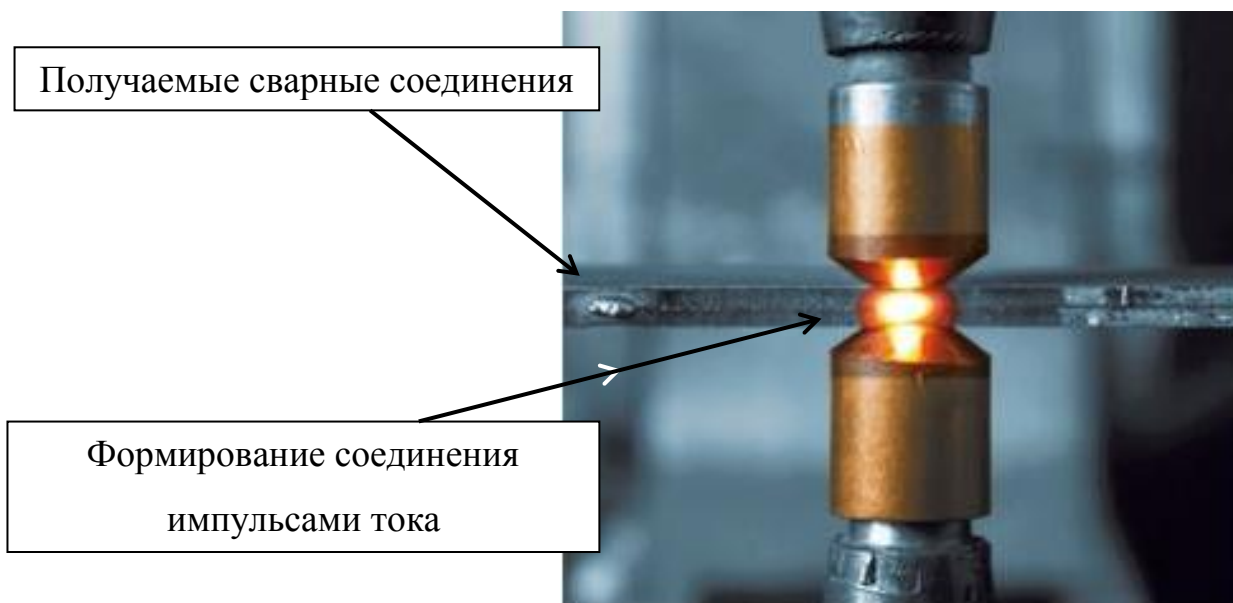


Рисунок 1.1 – Контактная точечная сварка

Главной проблемой контактной сварки является низкая стабильность качества получаемых соединений. Критерием качества является соответствие соединения прочностным характеристикам.

Данная проблема связана с влиянием на процесс сварки большого количества возмущающих воздействий. К ним можно отнести – загрязнённость поверхности свариваемых деталей, деградацию электродов, колебание напряжения в сети питания, несносность сварочных электродов.

Так как данные возмущения носят случайный характер, то составление математической модели их влияния на прочностные характеристики соединения затруднено.

Так как данные возмущения влияют на протекания процесса сварки, то предполагается, что их влияние (количественно) можно оценить путем анализа изменения электрических параметров (осциллограммы напряжения тока) и механических параметров (график сдвига электродов в результате теплового расширения металла).

Для того чтобы связать изменение электрических и механических параметров сварки с получаемыми характеристиками соединения необходимо решить задачу регрессионного анализа.

Зарубежные разработчики систем диагностики решают данную задачу (поиска зависимости между изменением электрических параметров процесса сварки и получаемыми прочностными характеристиками) традиционно с помощью нейронных сетей. Для этого они путем проведения экспериментальных сварок накапливают базу примеров, которые впоследствии анализируют с помощью LVQ, рекуррентных и нейронных сетей прямого распространения с целью поиска скрытых закономерностей.

В процессе анализа данных из полученной базы примеров нейронные сети обучаются предсказывать прочность получаемых сварных соединений (с заданной точностью) путём наблюдения за осциллограммами параметров процесса сварки.

Нейронные сети относятся к индуктивным алгоритмам машинного обучения (induction learning), к подобласти «обучение с учителем» (supervised learning).

Но у нейронных сетей есть существенные недостатки, которые усложняют процесс синтеза систем диагностики сваркой, к ним относятся:

- Стохастичный процесс обучения. Это означает, что по одним и тем же данным нейронная сеть может обучаться с разной точностью. А с учетом таких проблем обучения сети, как паралич сети и попадание в локальный минимум сеть может и вовсе не обучиться.

- Нейронная сеть воспринимает значимость всех входных сигналов как одинаковую. Это означает, что в процессе обучения существует вероятность, что нейронная сеть найдет закономерность там, где ее в действительности нет (например, связь между курсом доллара и расположением звезд).

Развитием машинного обучения появились и другие алгоритмы способные решать задачи регрессионного анализа, но лишенные описанных выше недостатков. Например, алгоритм построения деревьев принятия решений CART (полное название алгоритма – «Classification and Regression Tree»).

Данная выпускная квалификационная работа направлена на исследование возможности использования алгоритма CART для синтеза систем диагностики контактной точечной сварки. Возможность использования алгоритма CART для этой цели является гипотезой исследования.

Целью данной выпускной квалификационной работы обеспечение качества сварной продукции за счет разработки алгоритма диагностики сварки основанного на методе CART.

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

Получен акт внедрения исследований на предприятии ПАО «АвтоВАЗ».

## 1 Анализ состояния вопроса

### 1.1 Обзор предметной области

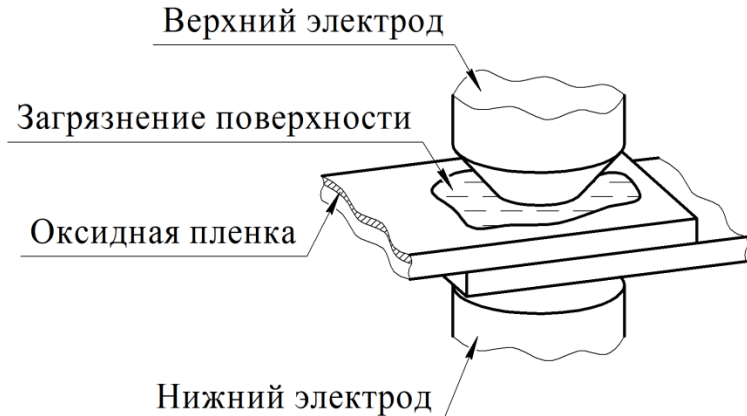
Контактная сварка (по сравнению с другими типами сварки) обладает такими важными преимуществами как возможность обеспечения высокой производительности получения неразъемных соединений, простотой автоматизации управления вспомогательными операциями (подвод электродов к месту сварки, сжатие электродов)

Она используется на большинстве крупных машиностроительных предприятиях самарской области. Так, например, на предприятии ПАО «АвтоВАЗ» контактная сварка применяется при изготовлении составных деталей автомобилей и при производстве автомобильных кузовов. При этом кузов автомобиля содержит порядка 2000 точечно сварных соединений[2].

Главной проблемой контактной сварки является низкая стабильность качества получаемых соединений. Критерием качества является соответствие соединения прочностным характеристикам.

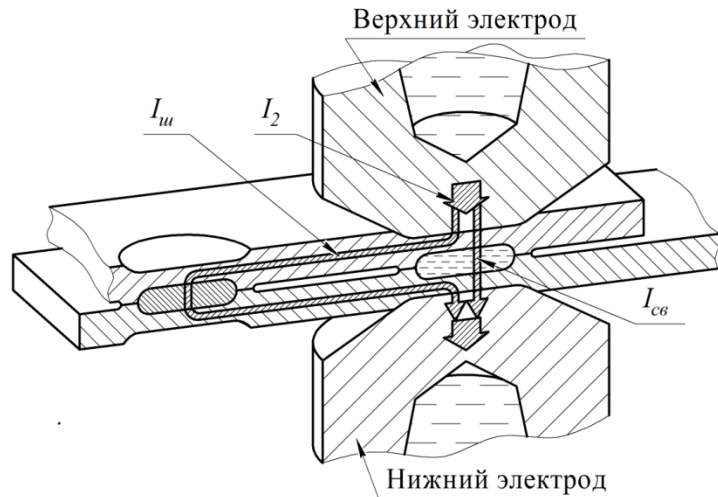
Данная проблема связана с влиянием на процесс сварки большого количества возмущающих воздействий. К ним можно отнести – загрязнённость поверхности свариваемых деталей, деградацию электродов, колебание напряжения в сети питания, несносность сварочных электродов, ток шунтирования[5].

Таблица 1 – Возмущения процесса сварки

Возмущение	Иллюстрация влияния возмущений
Загрязнение поверхности	

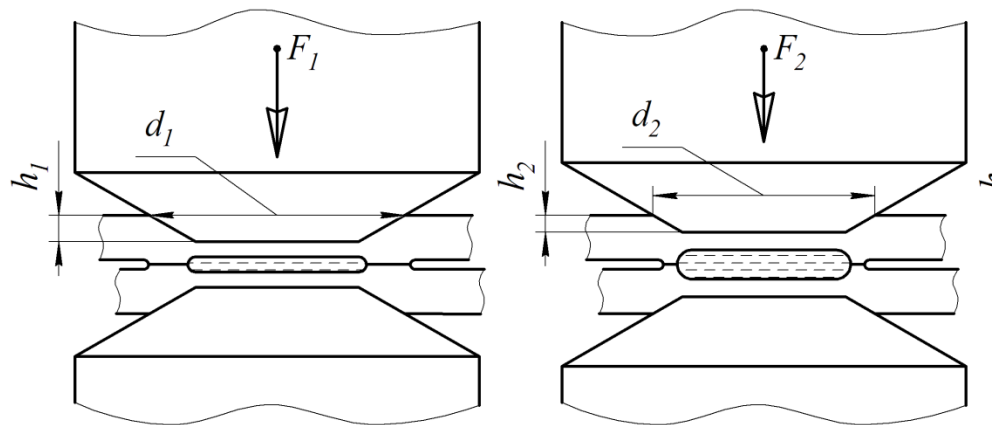


Ток шунтирования

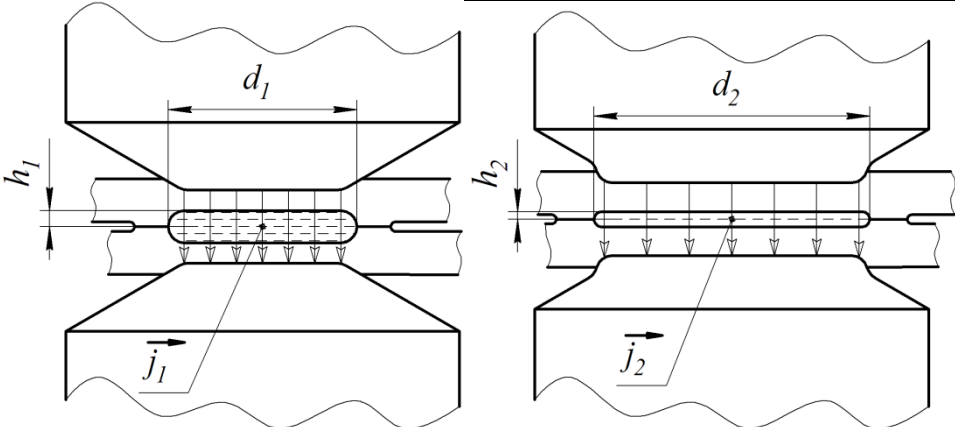
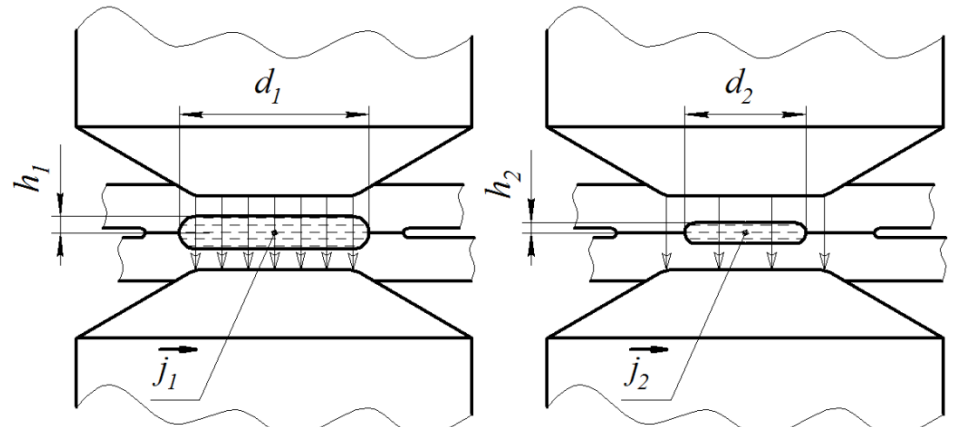


Поступающий через электроды в зону сварки ток  $I_2$  разделяется на “полезный” ток сварки  $I_{св}$  и “вредный” ток, проходящий через соседние сварные точки  $I_u$

Колебание усилия сжатия



Увеличение усилия сжатия  $F$  электродов приводит к увеличению диаметра  $d$  отпечатка электродов и глубины  $h$  промятия заготовок, что влияет на форму литого ядра и прочностные характеристики соединения

<p>Деграация электродов</p>	 <p>Деграация электродов заключается в расплющивании рабочей части электрода, что приводит при том же токе к снижению плотности <math>j</math> сварочного тока. В результате у сварного соединения получается другая глубина проплавления <math>h</math> и диаметр <math>d</math>.</p>
<p>Колебание напряжения сети</p>	 <p>Снижение напряжения питающей сети, приводит к снижению плотности <math>j</math> сварочного тока. В результате у сварного соединения получается глубина проплавления <math>h</math> и диаметр <math>d</math> получается меньше, чем это необходимо для обеспечения требуемых прочностных характеристик</p>

Так как данные возмущения влияют на протекания процесса сварки, то предполагается, что их влияние (количественно) можно оценить путем

анализа изменения электрических параметров (осциллограммы напряжения, тока) и механических параметров (график сдвига электродов в результате теплового расширения металла)[1].

Для того чтобы связать изменение электрических и механических параметров сварки с получаемыми характеристиками соединения необходимо решить задачу регрессионного анализа. Если же необходимо только оценить, к какому типу относиться полученное соединение (к качественной сварке или к некачественной), то решается задача классификации[6].

## **1.2 Обзор исследований по применению алгоритмов машинного обучения для диагностики качества сварки**

В настоящее время перспективным является использование алгоритмов машинного обучения при синтезе систем диагностики качества контактной сварки.

Данные системы диагностики выгодно отличаются от других способов контроля тем, что результаты оценки качества доступны сразу же после окончания сварки и для получения результатов оценки качества не требуется проведения никаких дополнительных действий. Так, до появления интеллектуальных систем диагностики сварки, применялся ультразвуковой контроль, радиографический контроль, металлографический контроль и другие трудоемкие способы контроля[9].

Проведем обзор исследований по применению новых подходов в системах диагностики качества сварки.

Один из новых подходов был предложен P. Laurinen, H. Junno, L. Tuovinen, J. Rönning в статье Studying the Quality of Resistance Spot Welding Joints Using Bayesian Networks журнала Proceedings of Artificial Intelligence and Applicatio [4].

В этом исследовании синтез системы диагностики сварки предложено рассматривать, как решение задачи классификации.

В качестве математического метода для решения данной задачи, в этом исследовании применяется Байесовские сети доверия (Bayesian network или как их еще называют - belief network).

Оценка качества в этом исследовании проводится путем прогнозирования принадлежности получившегося диаметра сварного ядра к одному из нескольких диапазонов (классов).

В процессе сварки, исследователями предложено измерять осциллограммы: сварочного тока  $i_{св}$  (ток в рабочем контуре сварочной машины), напряжения между электродами  $u_{св}$  и силы сжатия  $f_{сж}$  электродов.

При этом оценка качества формируется путем анализа таких значений, как:

- $U_{uq}, U_{lq}, U_{med}$  – верхний, нижний квартиль и медиана напряжения на электродах,

- $I_{uq}, I_{lq}, I_{med}$  – верхний, нижний квартиль и медиана сварочного тока,

- $U_{uq}, U_{lq}, U_{med}$  – верхний, нижний квартиль и медиана силы сжатия электродов[10].

Схема работы данного подхода представлена на рисунке 1.2

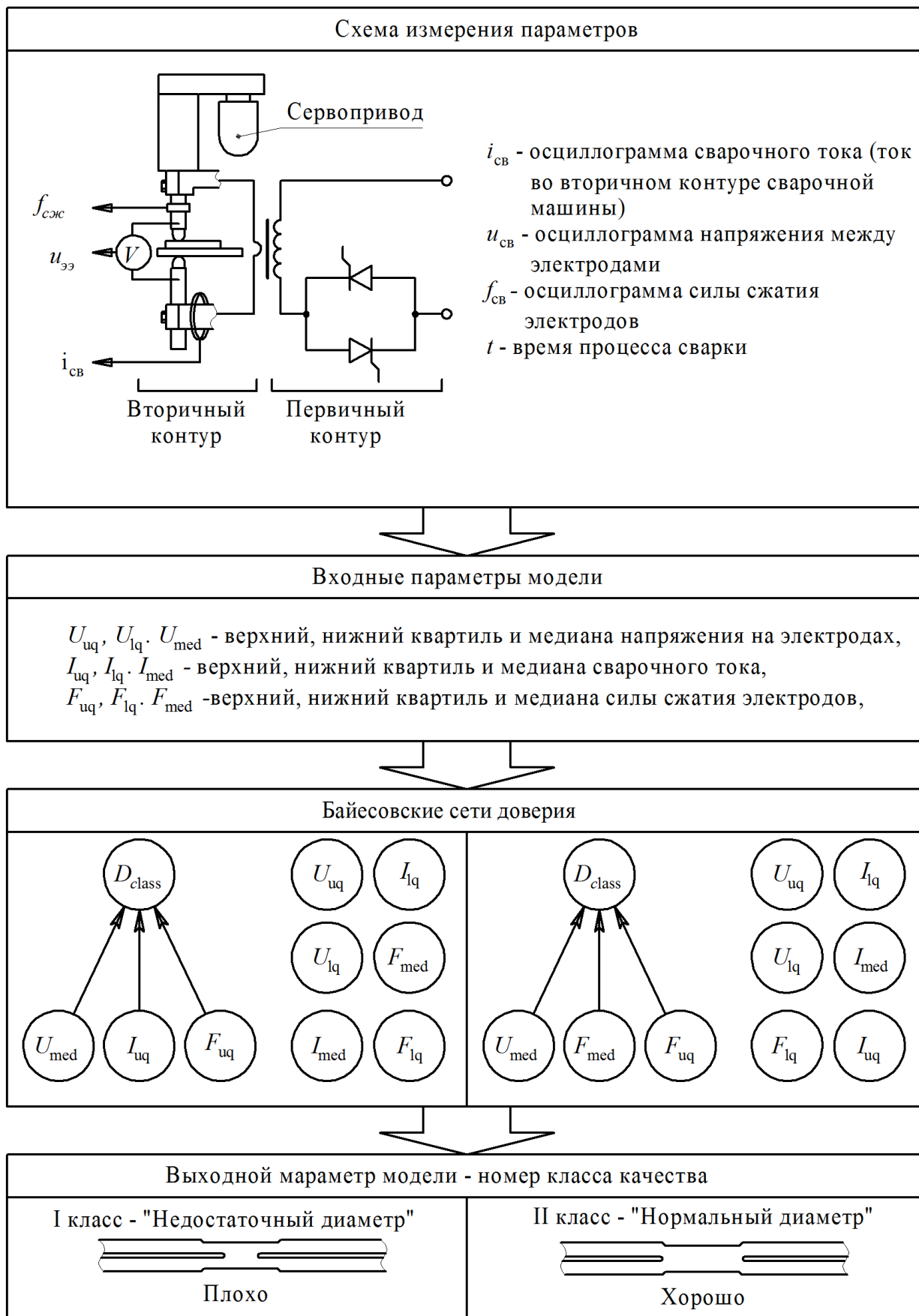


Рисунок 1.2 – Модель диагностики Р. Laurinen, Н. Junno, L. Tuovinen

Подход к диагностике сварки был предложен А. Aravinthan, К. Sivayoganathan, D. Al-Dabass, V. Balendran в статье «A neural network system for spot weld strength prediction» журнала посвященному конференции «Conference Proceedings of the UK Simulation Society» [5].

В этом исследовании синтез системы диагностики сварки предложено рассматривать, как решение задачи регрессионного анализа.

В качестве математического метода для решения данной задачи, в этом исследовании применяется двухслойная нейронная сеть прямого распространения.

Оценка качества в этом исследовании проводится путем предсказания значения предельной силы разрушения сварного соединения при испытаниях сварки методом “на срез” с помощью разрывной машины.

В процессе сварки, исследователями предложено измерять осциллограммы:

- сварочного тока  $i_{св}$  (ток в рабочем контуре сварочной машины),
- напряжения между электродами  $u_{св}$ .

При этом оценка качества формируется путем анализа вектора значений  $R = (R_1, R_2, \dots, R_m)$  - значения сопротивления на участке между электродами для временного периода с номером  $m$ .

Схема работы данного подхода представлена на рисунке 1.3

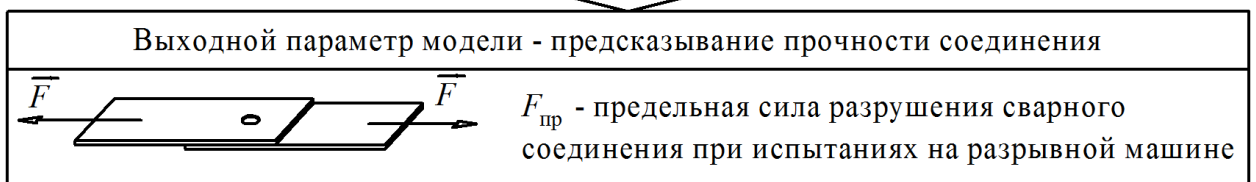
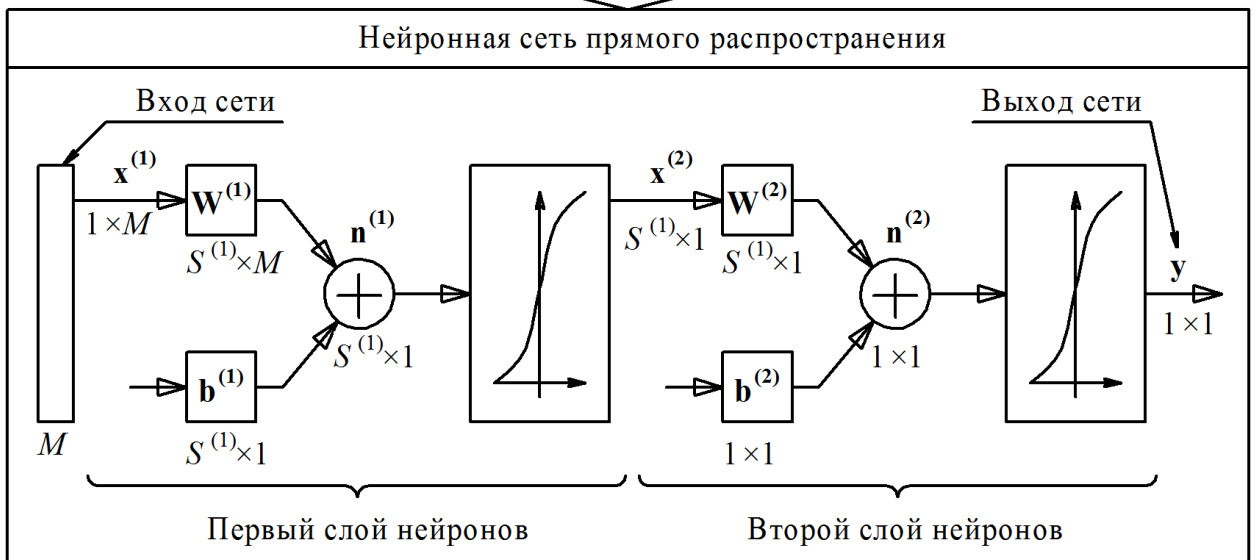
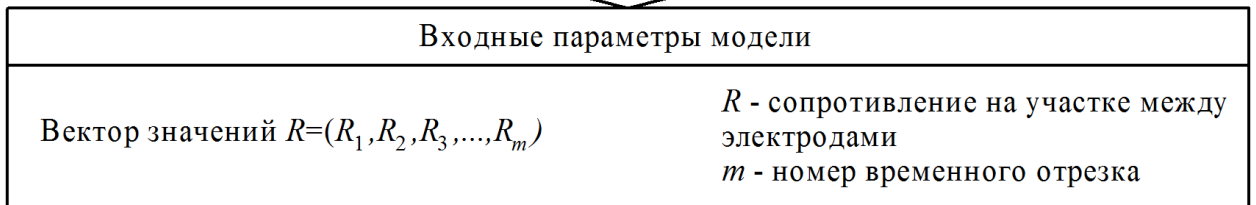
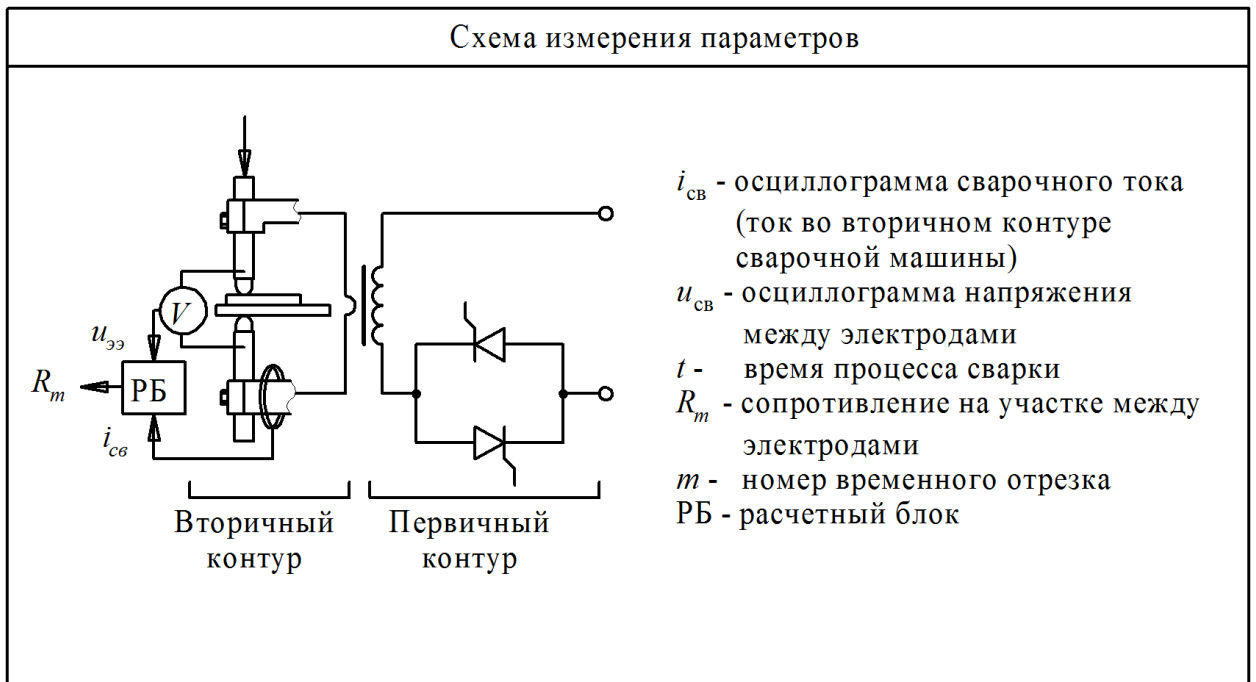


Рисунок 1.3 – Модель диагностики А. Aravinthan, К. Sivayoganathan

Другой подход к использованию машинного обучения в системах диагностики сварки предложили Б.Е. Патон, Н.В. Подола, В.С. Гавриш в статье «Оценка качества контактной точечной сварки с помощью нейронных сетей» журнала «Автоматическая сварка» [6].

В этом исследовании синтез системы диагностики сварки предложено рассматривать, как решение задачи регрессионного анализа.

В качестве математического метода для решения данной задачи, в этом исследовании применяется двухслойная нейронная сеть прямого распространения.

Оценка качества в этом исследовании проводится путем предсказания значения диаметра литого ядра сварного соединения при контроле сварного соединения с применением методов металлографии.

В процессе сварки, исследователями предложено измерять осциллограммы:

- сварочного тока  $i_{св}$  (ток в рабочем контуре сварочной машины),
- напряжения между электродами  $u_{св}$ .

При этом оценка качества формируется путем анализа значений двух векторов:

- $U = (U_1, U_2, \dots, U_m)$  значение напряжения на участке между электродами для временного периода с номером  $m$ .

- $R = (R_1, R_2, \dots, R_m)$  значение сварочного тока для временного периода с номером  $m$ .

Схема работы данного подхода представлена на рисунке 1.4



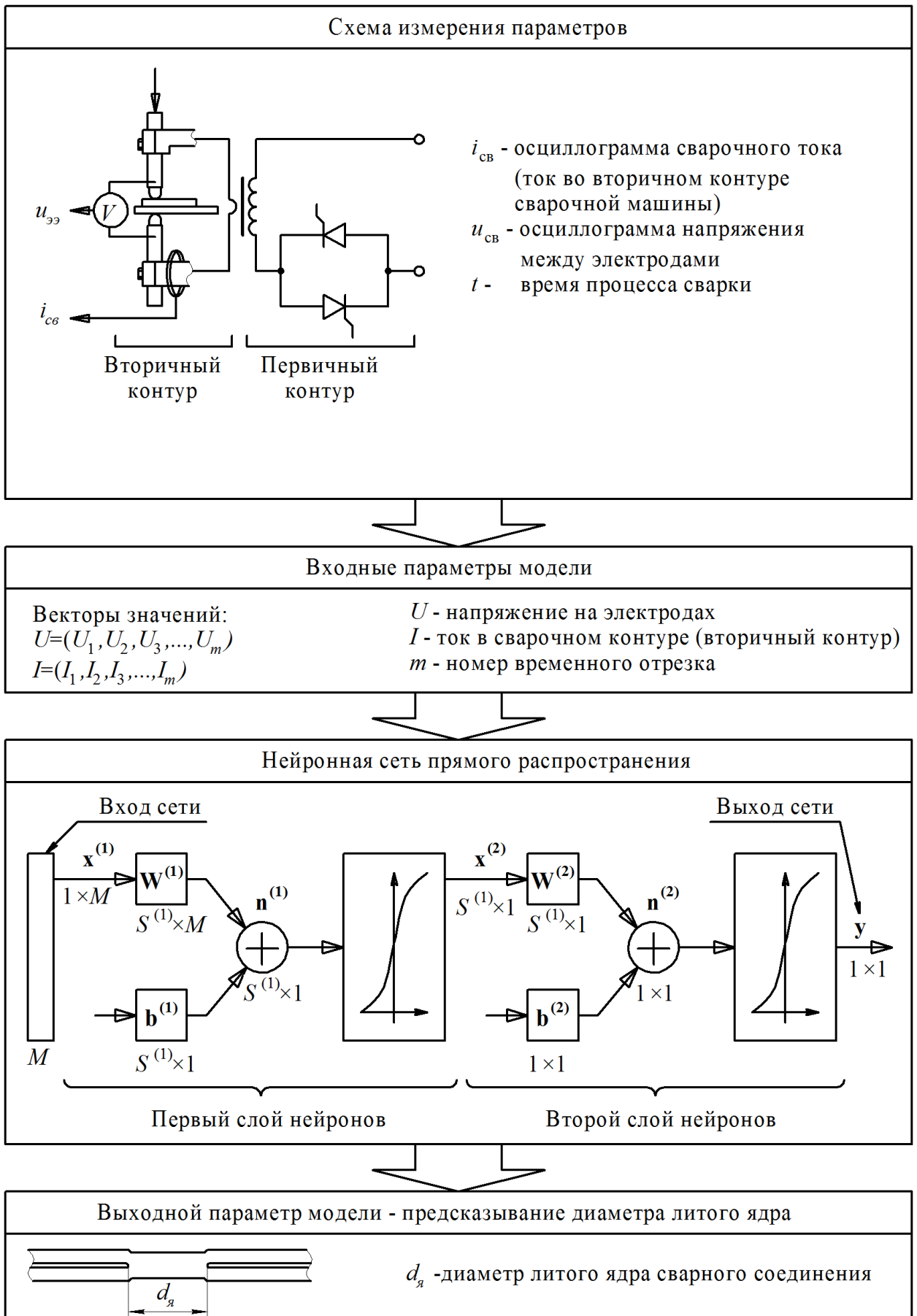


Рисунок 1.4 – Модель диагностики Б.Е. Патон, Н.В. Подола, В.С. Гавриш

Другой подход к использованию машинного обучения в системах диагностики сварки предложили М. El-Banna, D. Filev, R.B. Chinnam в статье «Online qualitative nugget classification by using a linear vector quantization neural network for resistance spot welding» журнала «The International Journal of Advanced Manufacturing Technology» [7].

В этом исследовании синтез системы диагностики сварки предложено рассматривать, как решение задачи классификации.

В качестве математического метода для решения данной задачи, в этом исследовании применяется нейронная сеть LVQ (Learning Vector Quantization).

Оценка качества в этом исследовании проводится путем предсказания принадлежности сварного соединения к одному из трех типов (классов): «Выплеск» (неудовлетворительное качество), «Норма» (удовлетворительное качество), «Непровар» (неудовлетворительное качество).

В процессе сварки, исследователями предложено измерять осциллограммы:

- сварочного тока  $i_{св}$  (ток в рабочем контуре сварочной машины),
- напряжения между электродами  $u_{св}$ .

При этом оценка качества формируется путем анализа вектора значений  $R = (R_1, R_2, \dots, R_m)$  - значения сопротивления на участке между электродами для временного периода с номером  $m$ .

Схема работы данного подхода представлена на рисунке 1.5.

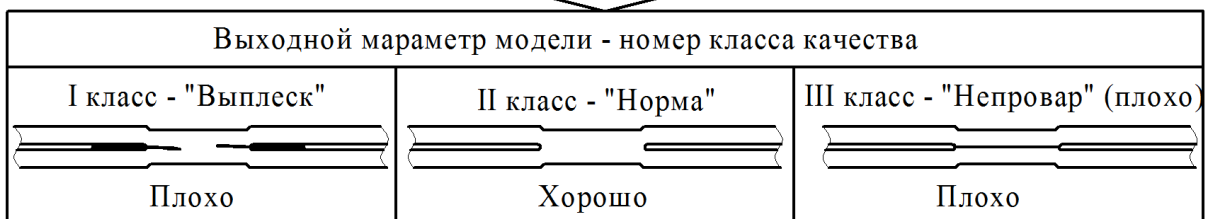
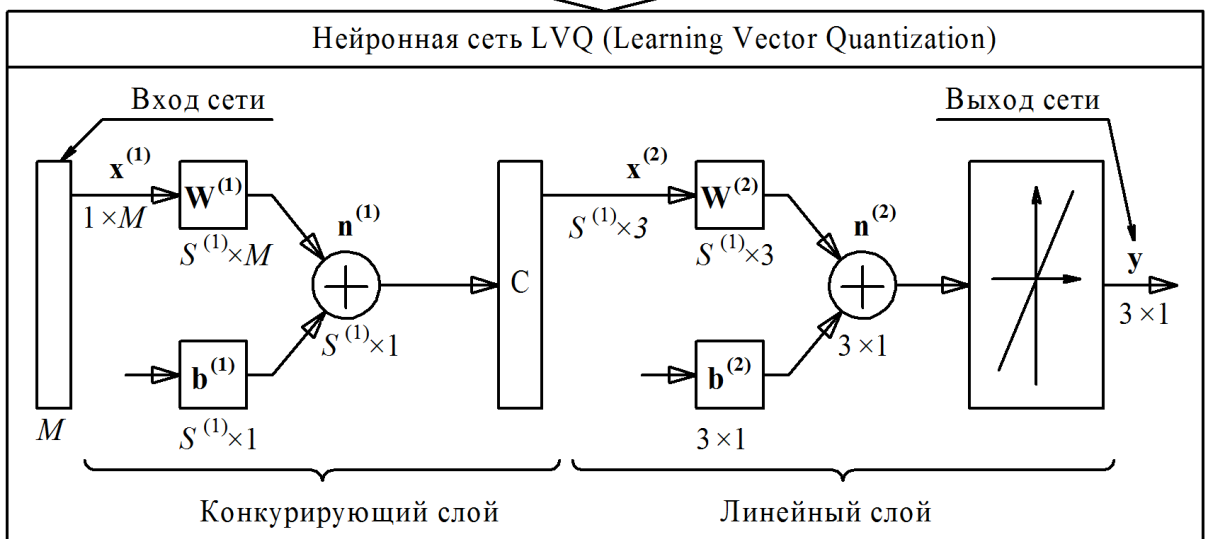
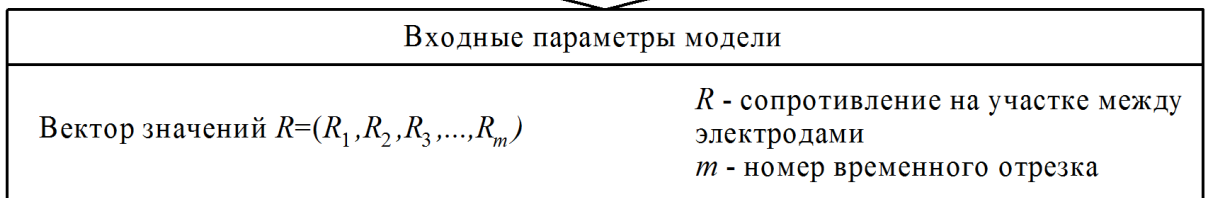
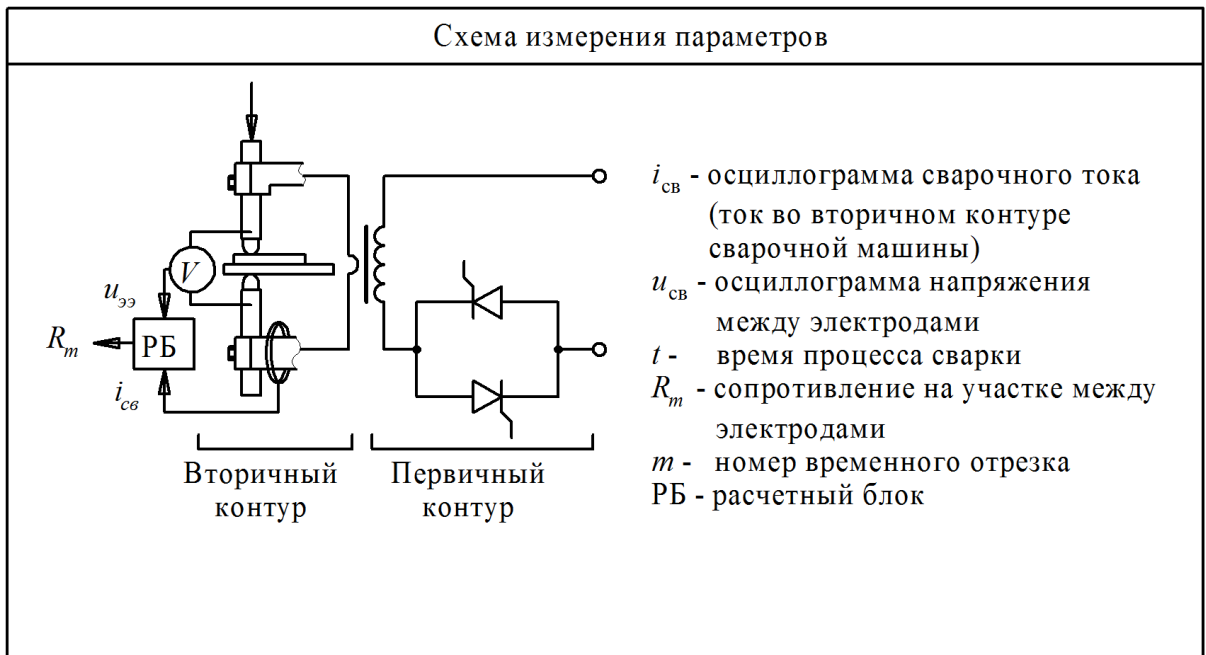


Рисунок 1.5 – Модель диагностики М. El-Banna, D. Filev, R.B. Chinnam

Другой подход к использованию машинного обучения в системах диагностики сварки предложили Y.J. Park, H. Cho в статье «Quality evaluation by classification of electrode force patterns in the resistance spot welding process using neural networks» журнала «Journal of Engineering Manufacture» [8].

В этом исследовании синтез системы диагностики сварки предложено рассматривать, как решение задачи классификации.

В качестве математического метода для решения данной задачи, в этом исследовании применяется нейронная сеть LVQ (Learning Vector Quantization).

Оценка качества в этом исследовании проводится путем предсказания принадлежности сварного соединения к одному из трех типов (классов): «Insufficient welding», «Poor welding», «Good welding», «Excessive welding».

В процессе сварки, исследователями предложено измерять осциллограмму изменения силы сжатия электродов  $f_{сж}$ .

При этом оценка качества формируется путем анализа вектора значений  $F = (F_1, F_2, \dots, F_m)$  - значения силы сжатия электродов на временном периоде с номером  $m$ .

Схема работы данного подхода представлена на рисунке 1.6

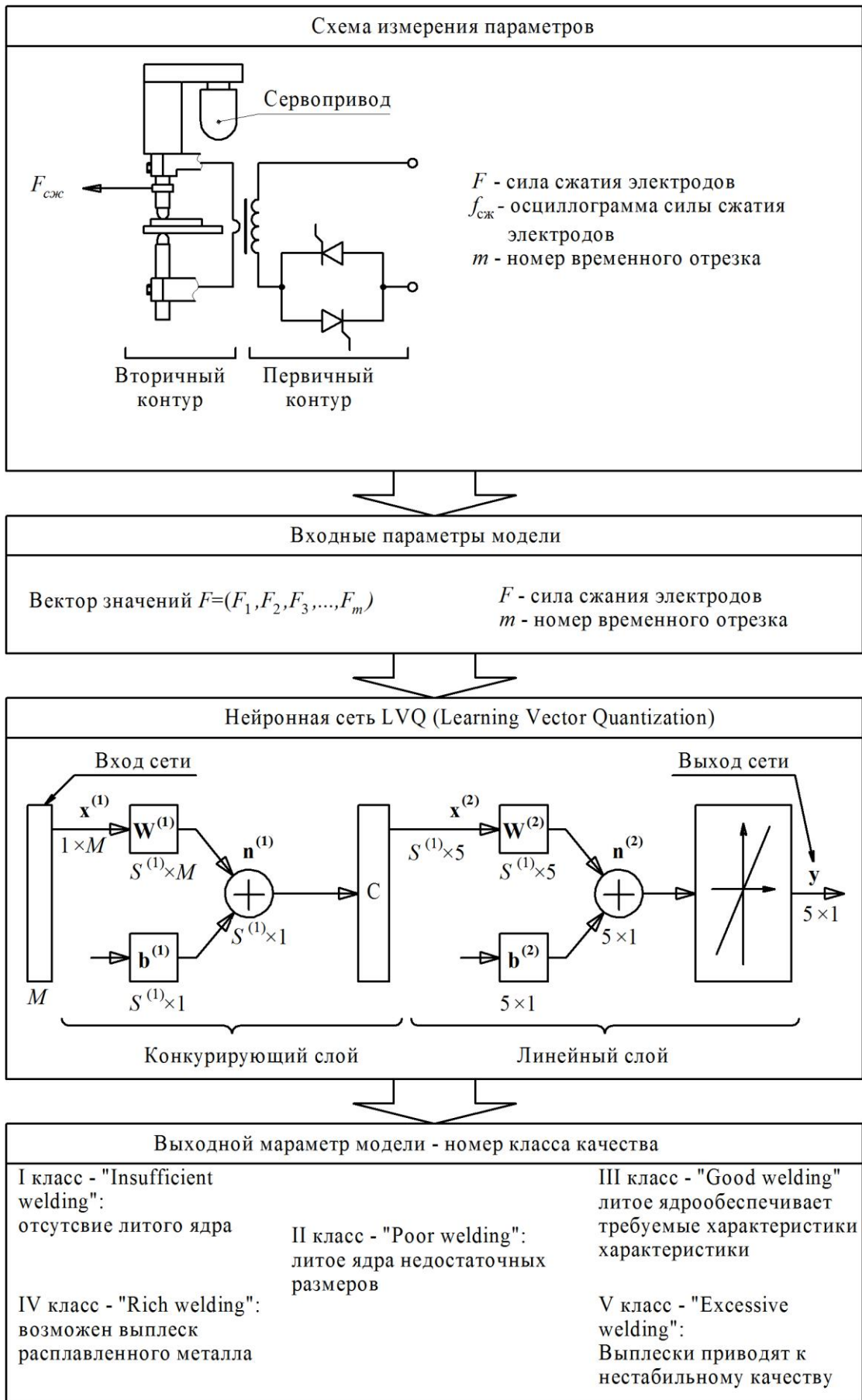


Рисунок 1.6 – Модель диагностики Y.J. Park, H. Cho

Таким образом, в результате анализа литературных источников можно сделать вывод, разработчики систем диагностики сварки в своих исследованиях пытаются связать изменение электрических и механических параметров сварки с получаемыми характеристиками соединений. Необходимо решить задачу регрессионного анализа.

Для решения данной задачи (поиска зависимости между изменением электрических параметров процесса сварки и получаемыми прочностными характеристиками) они предпочитают использовать нейронные сети. Для этого они путем проведения экспериментальных сварок накапливают базу примеров, которые впоследствии анализируют с помощью LVQ, рекуррентных и нейронных сетей прямого распространения с целью поиска скрытых закономерностей.

Обзор, проанализированный исследований, представлен в таблице 1.1.

Но у нейронных сетей есть существенные недостатки, которые усложняют процесс синтеза систем диагностики сваркой, к ним относятся:

- Стохастичный процесс обучения. Это означает, что по одним и тем же данным нейронная сеть может обучаться с разной точностью. А с учетом таких проблем обучения сети, как паралич сети и попадание в локальный минимум сеть может и вовсе не обучиться.
- Нейронная сеть воспринимает значимость всех входных сигналов как одинаковую. Это означает, что в процессе обучения существует вероятность, что нейронная сеть найдет закономерность там, где ее в действительности нет (например, связь между курсом доллара и расположением звезд).

С развитием машинного обучения появились и другие алгоритмы способные решать задачи регрессионного анализа, но лишенные описанных выше недостатков. Например, алгоритм построения деревьев принятия решений CART (полное название алгоритма – «Classification and Regression Tree»)[14].

Однако не исследована возможность применения алгоритма CART для решения задач диагностики сварки. Поэтому в данной исследовательской работе проверяется гипотеза о возможности использования алгоритма CART для прогнозирования прочностных характеристик получаемых сварных соединений.

Таким образом, целью работы является обеспечение качества сварной продукции за счет разработки алгоритма диагностики сварки основанного на методе CART.

Поставленная цель достигается путем последовательного решения ряда задач:

1. Анализ состояния вопроса об использовании алгоритмов искусственного интеллекта при решении задачи диагностики сварки.
2. Исследование возможности применения алгоритма CART для решения задачи восстановления зависимостей по эмпирическим данным при синтезе систем диагностики качества сварки.
3. Разработка алгоритма диагностики сварки с использованием алгоритма CART.
4. Апробация реализации предложенного подхода на практике.

## 2 Разработка алгоритма диагностики сварки, основанного на методе CART

### 2.1 Алгоритм CART как метод построения регрессионной модели

Формально задача построения регрессионной модели нескольких свободных переменных можно описать следующим образом.

Задана выборка данных, которую можно записать, как множество:

$$X = \{x_1, x_2, \dots, x_N \mid x \in \mathbb{R}^M\} \quad (2.1)$$

При этом существует, соответствующее ему, множество зависимых переменных которое можно записать как:

$$Y = \{y_1, y_2, \dots, y_N \mid y \in \mathbb{R}\}, \quad (2.2)$$

где  $N$  – размер обучающей выборки,  $M$  – количество независимых переменных.

Эти множества обычно обозначаются как  $D$ , множество исходных данных для настройки модели, которое можно записать как:

$$\{(x, y)_i\} \quad (2.3)$$

При этом вектор  $x_i$  описывается набором значений числовых атрибутов  $P$ , а  $y_i$  – описывается единственным числовым значением:

$$x_i = (P_1, P_2, \dots, P_M), \quad (2.4)$$

Для задания регрессионной модели используется параметрическое семейство функций вида  $f(w, x)$ , что можно записать как:

$$f : W \times X \rightarrow Y, \quad (2.5)$$

где  $w \in W$  – пространство настраиваемых параметров.

Регрессионная модель является настроенной, когда зафиксированы её параметры, то есть модель задаёт отображение для фиксированных значений  $W$ :

$$f : X \rightarrow Y \quad (2.6)$$

Одним из возможных представлений регрессионной модели является дерево принятия решений. Дерево решений – это связный ациклический



граф, представляющий параметрическое семейство функций  $f$  в иерархической последовательной структуре.

Дерево состоит из узлов и листьев. В каждом узле расположена проверка по одному из атрибутов вектора  $x_i$ . В каждом листе расположено одно из значений множества  $Y$ .

Чтобы определить значение выходного параметра для любого объекта, описываемого набором значений  $(P_1, P_2, \dots, P_M)$  необходимо произвести спуск по дереву принятия решения от самого верхнего узла (т.н. корня) до одного из листьев. При этом необходимо следовать условиям перехода, которые содержатся в ребрах между узлов. В каждом из листьев будет содержаться одно из значений множества  $Y$ . В зависимости от значений  $(P_1, P_2, \dots, P_M)$  спускаясь по дереву можно прийти к различным листьям, содержащим значения множества  $Y$  (рисунок 2.1).

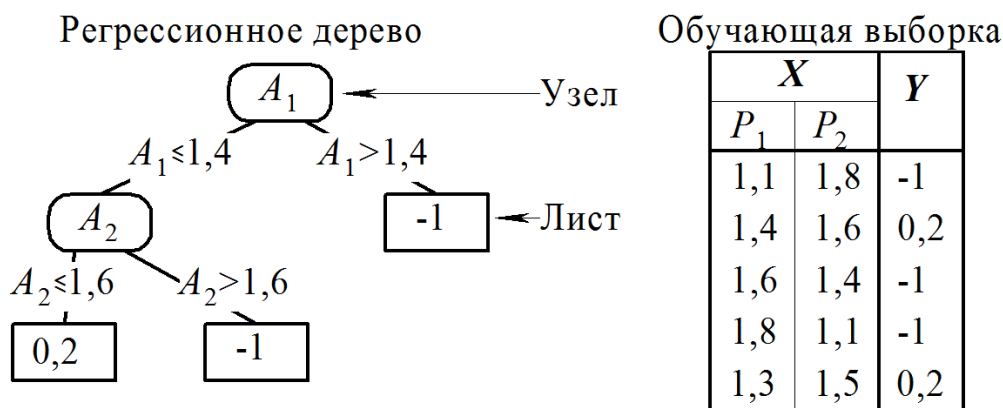


Рисунок 2.1 – Пример обучающей выборки и соответствующее ей регрессионное дерево

Общий принцип построению деревьев решений, заключается в рекурсивном разбиении множества объектов из обучающей выборки  $D$  на подмножества  $T_j$ , содержащие объекты, с одинаковыми значениями выходного параметра  $y_i$ .

Относительно любого подмножества  $T$  и в зависимости от значений выходного параметра возможны три ситуации:

1) Множество  $T$  содержит один или более объектов, с одинаковыми значениями выходного параметра  $y_i$ . Тогда дерево решений для  $T$  - это лист, определяющий возвращающий значение  $y_i$ ;

2) Множество  $T$  не содержит ни одного объекта (пустое множество). Тогда данное множество  $T$  – это лист, с выходным значением, рассчитанное как среднеарифметическое выходных значений объектов, находящихся в узле-родителе;

3) Множество  $T$  содержит объекты, с разными значениями выходного параметра  $y_i$ . В этом случае следует разбить множество  $T$  на некоторые подмножества. Для этого выбирается один из атрибутов  $P_1, P_2, \dots, P_M$  и  $T$  разбивается на подмножества 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”).

Основная идея алгоритма CART – разбиение всего пространства признаков на области разного размера, в которых выходная переменная  $y_i$  считается постоянной.

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.

В алгоритме CART выбор наилучшего разбиения для текущего узла, содержащего в себе множество  $T$  объектов  $t$ , происходит следующим образом.

Сначала определяются все возможные варианты разбиения. Для каждого атрибута  $P_i$  существует несколько вариантов разбиений, количество которых равно мощности множеству  $Z$  порогов. Множество порогов находится путем записи всех уникальных значений (в обучающей выборке  $D$ ) рассматриваемого атрибута  $P_i$  в порядке возрастания без повторов с отбрасыванием наибольшего значения. Затем каждое значение порога  $z$  выступает в роли точки деления исходного множества  $T$ , состоящего из объектов  $t$ , на два подмножества  $T_1$  и  $T_2$  (2.7):

$$\begin{cases} T_1 = \{ t \in T \mid a(t) \leq z \\ T_2 = \{ t \in T \mid a(t) > z \} \end{cases}, \quad (2.7)$$

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

Затем каждый вариант разбиения оценивается следующим образом. Для подмножеств  $T_1$  и  $T_2$  рассчитываются регрессионные оценки, соответственно,  $f_1$  и  $f_2$  по следующим формулам:

$$\begin{cases} f_1 = \frac{1}{|T_1|} \sum_{t \in T_1} y(t) \\ f_2 = \frac{1}{|T_2|} \sum_{t \in T_2} y(t) \end{cases} \quad (2.8)$$

Данные регрессионные оценки используются для проверки эффективности рассматриваемого разбиения. Для этого сначала вычисляется ошибка  $E$  по формуле:

$$E = \sum_{t \in T_1} (y(t) - f_1)^2 + \sum_{t \in T_2} (y(t) - f_2)^2 \quad (2.9)$$

В соответствии с алгоритмом CART выбирается наилучшим разбиением для текущего узла, содержащего в себе множество  $T$  объектов  $t$ , считается то, которое обеспечивает наименьшую сумму  $E$  квадратов ошибок:

$$E \rightarrow \min \quad (2.10)$$

Таким образом, алгоритм CART при построении регрессионной модели самостоятельно определяет, какой атрибут из набора  $P_1, P_2, \dots, P_M$  и какое значение порога  $z$  необходимо использовать на каждой итерации построения дерева.

## 2.2 Основные параметры контактной сварки

Контактная сварочная машина подключается к двум фазам трехфазной сети питания на предприятии с напряжением промышленной частоты (50 Гц).

В первичном контуре сварочной машины максимальное напряжение  $u_c$  составляет  $380\sqrt{2}$  В. За счет трансформатора во вторичном контуре напряжение  $u_{ce}$  составляет всего несколько вольт, при этом действующее значение сварочного тока  $i_{ce}$  измеряется кА.

Для управления мощностью вкладываемой в формирование сварной точки во вторичном контуре установлен контактор, который может прерывать подачу напряжения  $u_c$  на сварочный трансформатор. Напряжение на трансформаторе со стороны первичного контура, с учетом работы контактора, имеет вид  $u_2$  (как это показано на рисунке 2.2).

Таким образом, импульсы сварочного тока (если учитывать только временные характеристики) обладают следующими параметрами:

- задержка  $\alpha$  подачи импульса тока  $i_{ce}$  относительно перехода сетевого напряжения  $u_{ce}$  через ноль;
- Длительность  $\lambda$  подачи импульса тока  $i_{ce}$ .

Причем значение  $\alpha$  задается системой управления сваркой, а вот значение длительности  $\lambda$  (подачи импульса тока) является саморегулируемым параметром. Т.е. подачу импульса тока нельзя “вручную” прервать – контактор будет пропускать ток, пока его значение не достигнет нуля.

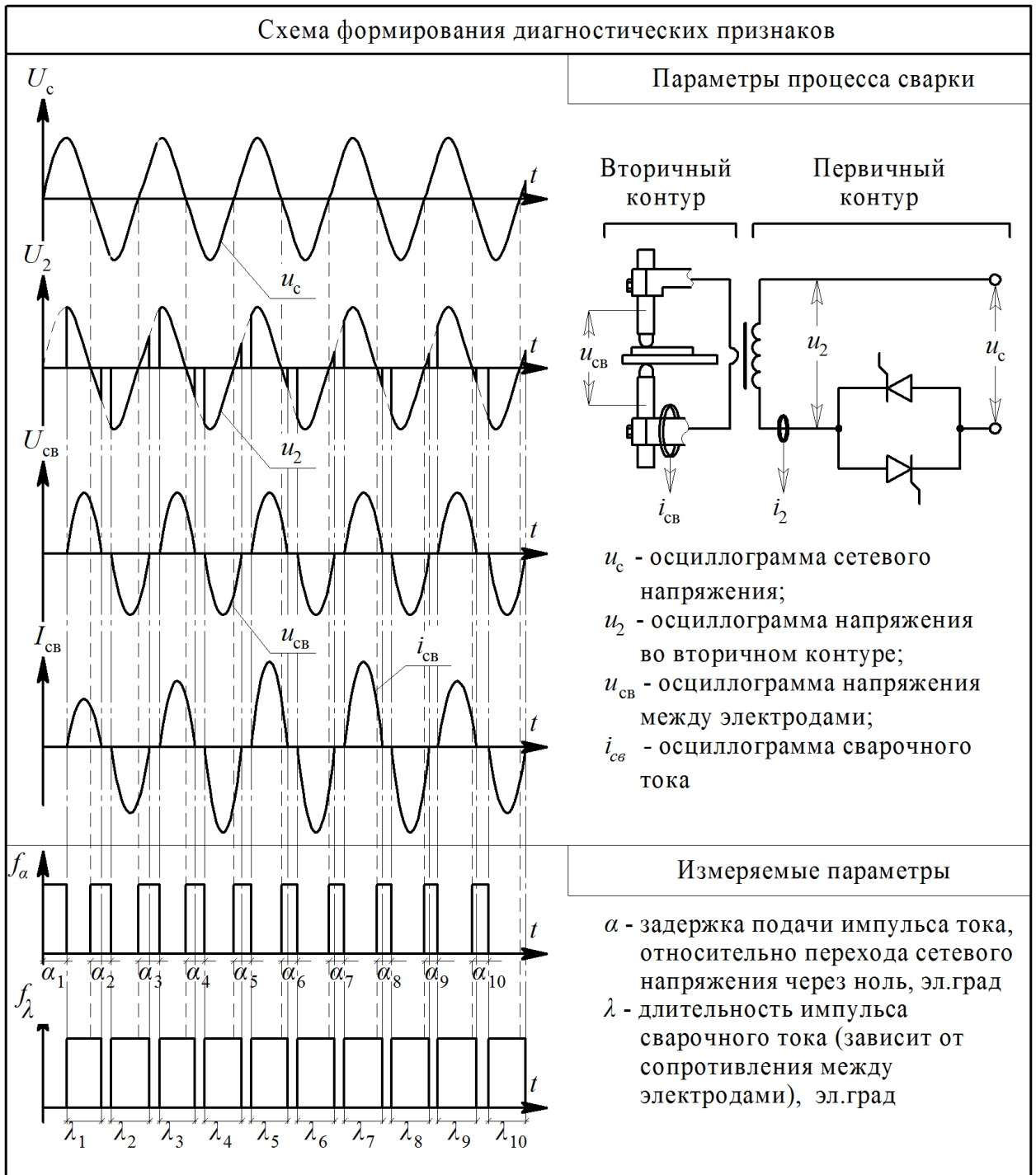


Рисунок 2.2 – Зависимость основных параметров процесса сварки и измеряемых параметров, по которым производится диагностика качества

Значение длительности  $\lambda$  импульса сварочного тока зависит как от значения  $\alpha$ , так и от изменения сопротивления на участке между электродами. А это сопротивление, в свою очередь меняется вследствие протекания металлургических процессов в зоне сварки (нагрев, увеличение зоны контакта, зарождения и рост литого ядра в зоне сварки). Именно металлургические процессы обеспечивают прочностные характеристики сварного соединения.

Таким образом, сочетание параметров задержки  $\alpha$  и длительность  $\lambda$  обладает диагностическим потенциалом. Так как в течении сварки импульсов тока для формирования соединения подается несколько, то при измерении  $\alpha$  и  $\lambda$  по окончанию сварки будет сформировано два вектора:  $\mathbf{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_m)$ ,  $\mathbf{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_m)$ , где  $m$  – количество импульсов тока или количество периодов сварки (периоды сетевого напряжения в которых подаются импульсы тока называют периодами сварки).

Пара значений  $\alpha$  и  $\lambda$  связана с коэффициентом мощности сварки  $\cos(\varphi)$  следующим известным равенством:

$$\sin(\lambda + \alpha - \varphi) / \sin(\alpha - \varphi) = e^{-\lambda \cdot \text{ctg} \varphi} \quad (2.10)$$

Решение равенства относительно  $\varphi$ , позволяет рассчитывать коэффициент мощности для каждого периода сварки.

Таким образом, изменение коэффициента мощности в процессе сварки предложено использовать в качестве диагностического признака, характеризующего протекание металлургических процессов при сварке.

Предлагается анализировать изменение коэффициента мощности в процессе сварки с целью предсказания силы разрушения  $F_{\text{пр}}$  сварного соединения при испытаниях на срез (рисунок 2.3).

### Алгоритм построения регрессионной модели

Для каждого сварного соединения из обучающей выборки (в результате измерений), имеем два вектора:

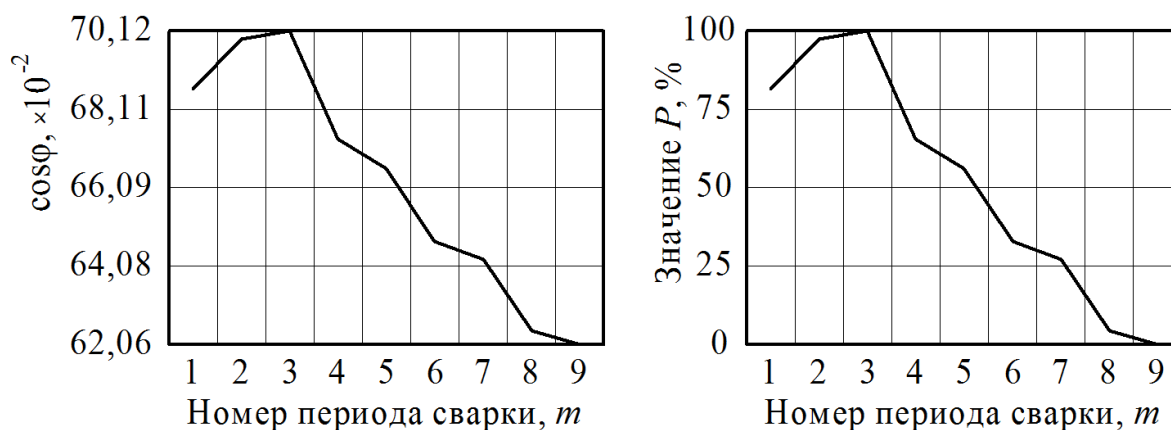
$$\alpha = (\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_m) \text{ и } \lambda = (\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_m)$$

Используя представленное равенство, для каждой пары  $\alpha_i$  и  $\lambda_i$  рассчитываем значение коэффициента мощности сварки  $\cos(\varphi_i)$

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

Полученный набор значений  $\cos \varphi$  будет характеризовать протекание различных этапов формирования сварного соединения

Так как для диагностики интересна форма кривой  $\cos \varphi$ , то значения вектора  $\cos \varphi$  нормируется к диапазону от 0 до 1 (или от% 0 до 100%):  $\cos(\varphi_i) \rightarrow P_i$



Таким образом, каждая сварка из обучающей выборки **D** описывается:

- 1) Вектором **P** ( $P_1, P_2, P_3, \dots, P_m$ ) - нормированные значения  $\cos$ .
- 2) Выходным значением  $F_{\text{пр}}$  (измеренная сила прочности сварного соединения)

Обучающей выборки **D** подается алгоритму CART для построения регрессионного дерева принятия решений.

Рисунок 2.3 – Алгоритм подготовки данных для построения регрессионного дерева с помощью алгоритма CART

### 2.3 Алгоритм диагностики сварки с использованием алгоритма CART

Сначала 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  $\alpha$  и длительность  $\lambda$  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  $(P_1, P_2, \dots, P_m)$ , где количество компонентов  $n$  равно количеству периодов сварки и значение, указывающее на качественную характеристику соединения (например, сила разрушения сварного соединения при испытаниях на срез).

Затем полученный массив данных (также называемый обучающей выборкой) анализируется с помощью алгоритма CART, для автоматического определения такого сочетания проверок по значениям компонентов  $P_1, P_2, \dots, P_m$ , которое позволяет прогнозировать качественную характеристику соединения. Результат работы алгоритма можно представить в виде регрессионного дерева, описывающего зависимость между значений компонентов  $P_1, P_2, \dots, P_m$  и получаемой качественной характеристикой соединения.



### 3 Практическая реализация предложенных решений

#### 3.1 Моделирование построения CART в среде Matlab

Моделирование построения деревьев проводилось в среде Matlab. Для этого использовался комплект *Statistic and Machine Learning Toolbox*.

Для того, что бы построить дерево принятия решений по алгоритму CART необходимо сначала создать две переменные в области глобальных переменных. Одну переменную (предположим  $X$ ) в которой будет храниться матрица независимых переменных, обучающей выборки (в нашем случае – это значения  $P_1, P_2, \dots, P_n$ ). И вторая переменная (например,  $y$ ), которая будет содержать в себе зависимые значения переменной (в нашем случае – это значения силы разрушения соединения  $F_{пр}$ ).

После того, как обучающая выборка необходимо ввести команду: `rtree = fitrtree(X,y)`. Функция `fitrtree` отвечает за построение полного бинарного дерева регрессии по алгоритму карт. Построенное дерево сохраниться в переменной `rtree`. Для просмотра сгенерированных правил регрессии можно воспользоваться командой `view(rtree)`. Для просмотра дерева в виде графа необходимо ввести `view(rtree,'mode','graph')` (рисунок 3.1).

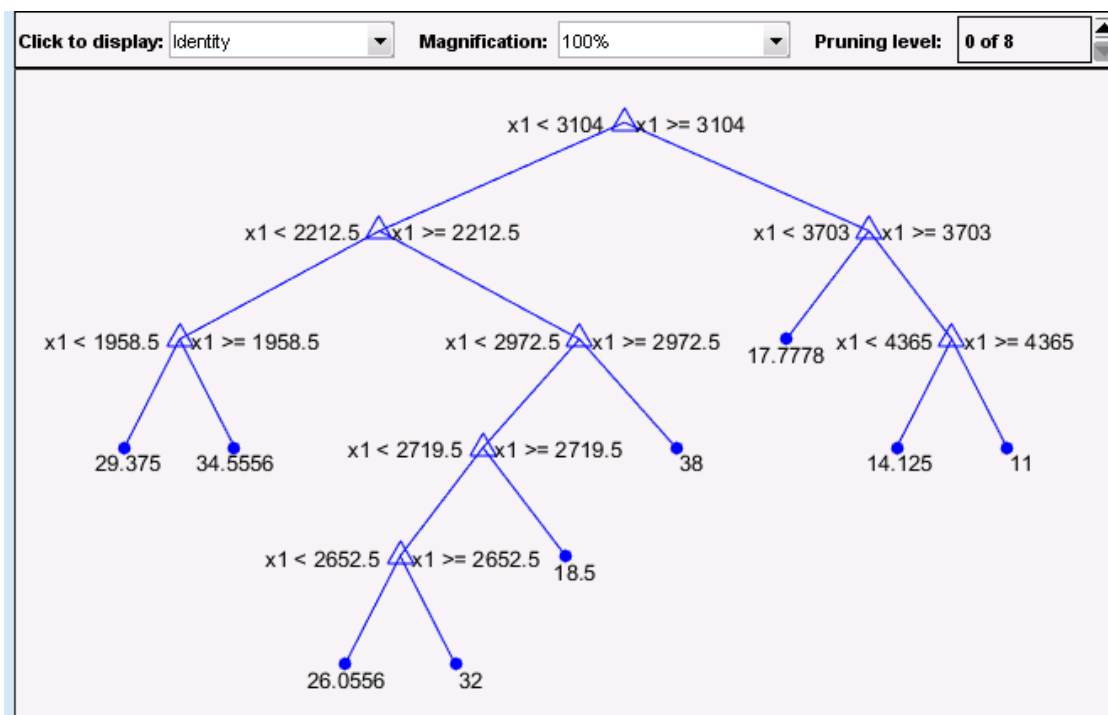


Рисунок 3.1 – Пример отображения дерева в Matlab

### **3.2 Практическая апробация алгоритма и исследование точности полученной регрессионной модели**

С использованием технологического оборудования лаборатории “Контактная сварка” института машиностроения ТГУ было проведено около 100 экспериментальных сварок пакета заготовок из стали 08кп.

Испытания сварных соединения для определения силы разрушения соединений при статических нагрузках на срез проводились на оборудовании НИИ “Прогрессивные технологии”. Результаты испытаний оформлены в виде протокола испытательного центра при НИИ с действующей государственной аккредитацией.

Данные от 59 сварок использовались при построении регрессионной модели в виде дерева принятия решений (по алгоритму CART).

Данные от 10 сварок использовались при тестировании точности полученной модели.

Полученное полное дерево регрессии представлено в приложении А.

Полное дерево регрессии обладает максимальным количеством узлов (117 шт.) и листов (59 шт.). Однако не всегда оправдано использование полной версии регрессионного дерева. Возможно что, существует усеченная версия дерева, более компактная и функционирующая с требуемой точностью. У усеченного дерева точность всегда меньше чем у полного дерева.

Существует два механизма построения усеченного дерева:

1. Ограничение минимального количество примеров, содержащихся в каждом узле дерева. Такой механизм используется при построении дерева.

2. Проверка для каждого узла состав содержащихся в нем объектов обучающей выборки. Если значения прогнозируемого параметра всех объектов в данном узле близки (критерий близости регулируется параметром  $M$ ) к среднему (по узлу) значению, то такой узел объявляется листом. Такой лист возвращает среднее значение выходного параметра.

Исследование влияния механизмов усечения на точность регрессионной модели исследовалась с использованием следующих показателей точности:

- MSE (mean squared error) – среднеквадратичная ошибка, которая рассчитывается на тестовой выборке данных как:

$$MSE = \frac{1}{n} \sum_{i=1}^n (p_i - a_i)^2 \quad (3.1)$$

- RMSE (root mean square error) – квадратный корень из среднеквадратичной ошибки, которая рассчитывается на тестовой выборке данных как:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - a_i)^2} \quad (3.2)$$

- MAE (mean absolute error) – средняя абсолютная ошибка, которая рассчитывается на тестовой выборке данных как:

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - a_i| \quad (3.3)$$

- RSE (relative squared error) – относительная квадратичная ошибка, которая рассчитывается на тестовой выборке данных как:

$$RSE = \frac{\sum_{i=1}^n (p_i - a_i)^2}{\sum_{i=1}^n (\bar{a} - a_i)^2} \quad (3.4)$$

- RRSE (root relative squared error) – квадратный корень из относительной квадратичной ошибки, который рассчитывается на тестовой выборке данных как:

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (p_i - a_i)^2}{\sum_{i=1}^n (\bar{a} - a_i)^2}} \quad (3.5)$$

- RAE (relative absolute error) – относительная абсолютная ошибка, которая рассчитывается на тестовой выборке данных как:

$$RAE = \frac{\sum_{i=1}^n |p_i - a_i|}{\sum_{i=1}^n |\bar{a} - a_i|} \quad (3.6)$$

- R2 (coefficient of determination) - Коэффициент детерминации  $R^2$ , квадратичной ошибка, который рассчитывается на тестовой выборке данных как:

$$R^2 = 1 - \frac{\sum_{i=1}^n (p_i - a_i)^2}{\sum_{i=1}^n (\bar{a} - a_i)^2} = \frac{\sum_{i=1}^n (\bar{p} - p_i)^2}{\sum_{i=1}^n (\bar{a} - a_i)^2} \quad (3.7)$$

- При этом  $a_i$  – фактические значения выходного параметра для  $i$ -ого объекта тестовой выборки,  $p_i$  – выходное значение, выдаваемое регрессионной моделью для входных параметров  $i$ -ого объекта тестовой выборки,  $n$  – количество объектов тестовой выборки.

Результаты исследования влияния механизмов усечения на точность регрессионной модели показаны на рисунках 3.2 – 3.8.

Исследована точность прогнозирования силы разрушения соединения построенной регрессионной моделью (полное дерево): MSE=18520.88 Н<sup>2</sup>, RMSE = 136.0915 Н, MAE=133.5556 Н, RSE = 0.0301, RRSE = 0.1736, RAE = 0.2044, R2 = 0.9699

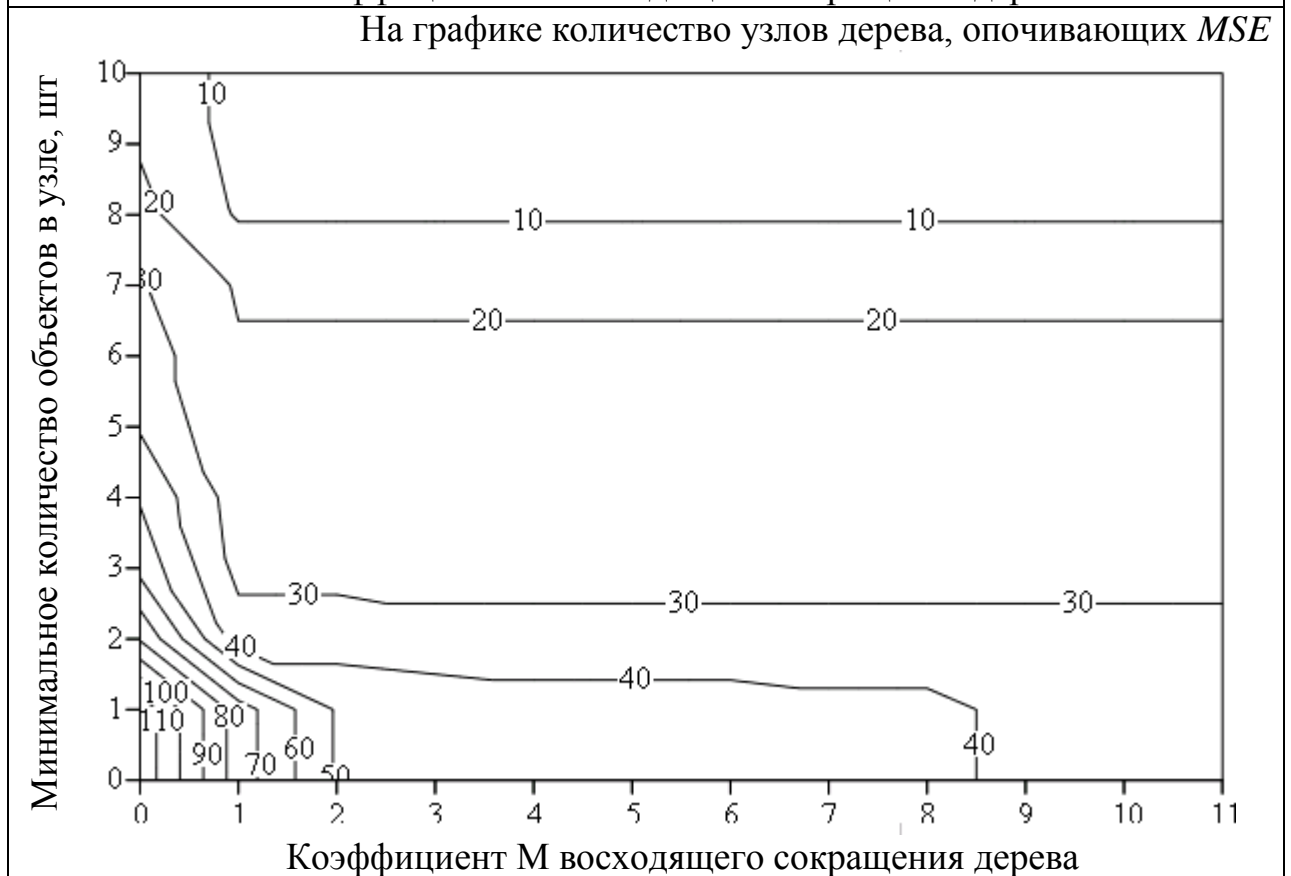
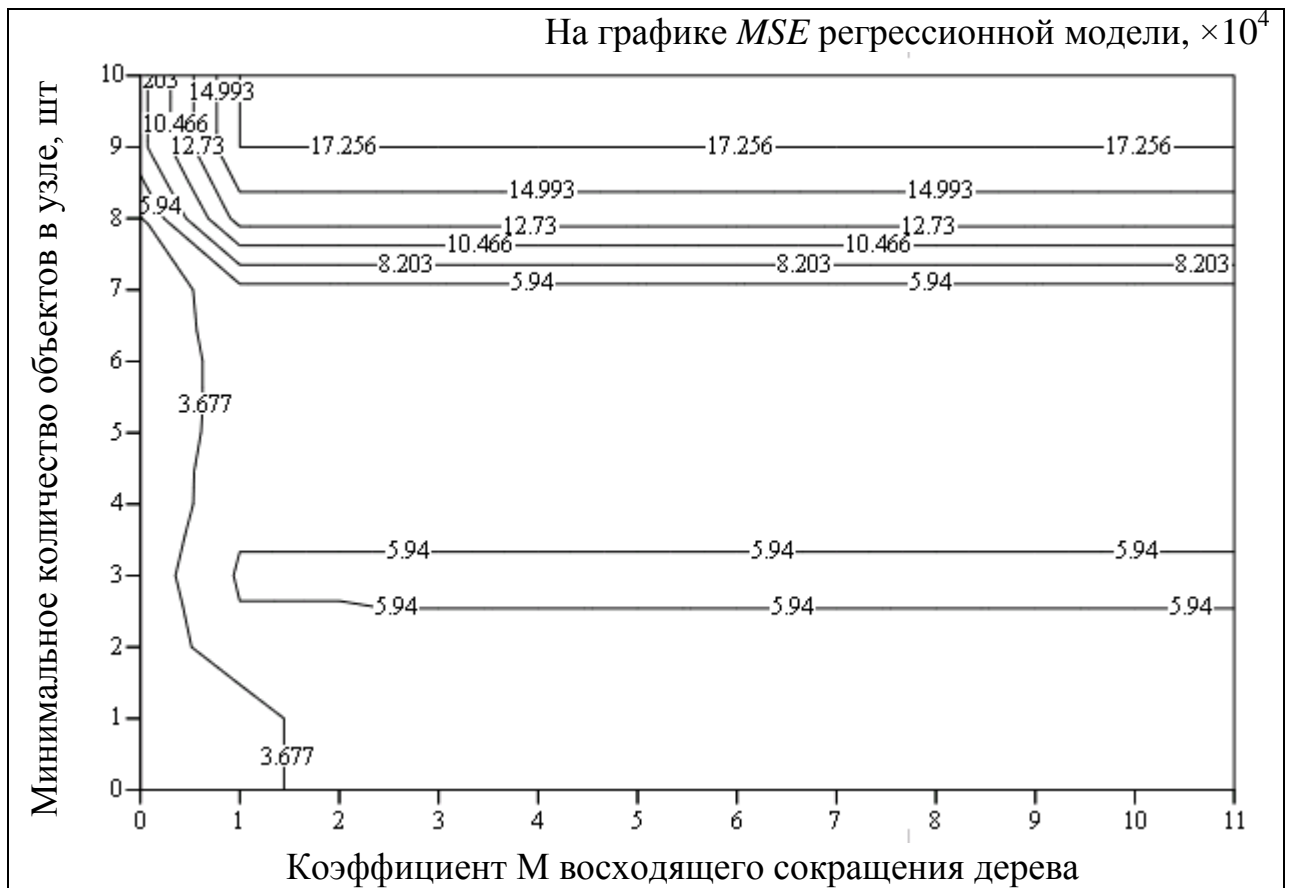


Рисунок 3.2 – Изменение значения  $MSE$  в зависимости от параметров построения дерева и количества узлов

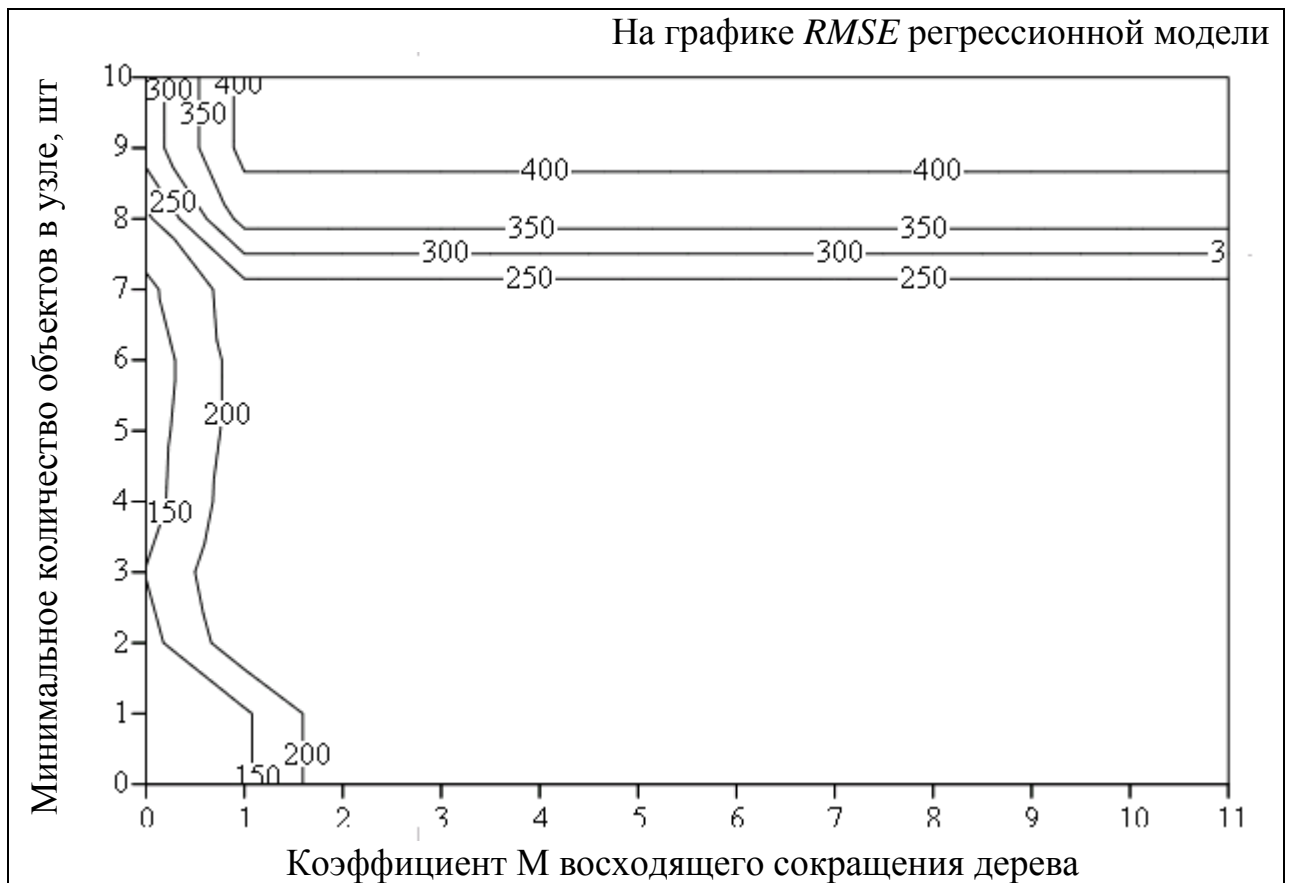


Рисунок 3.3 – Изменение значения *RMSE* в зависимости от параметров построения дерева и количества узлов

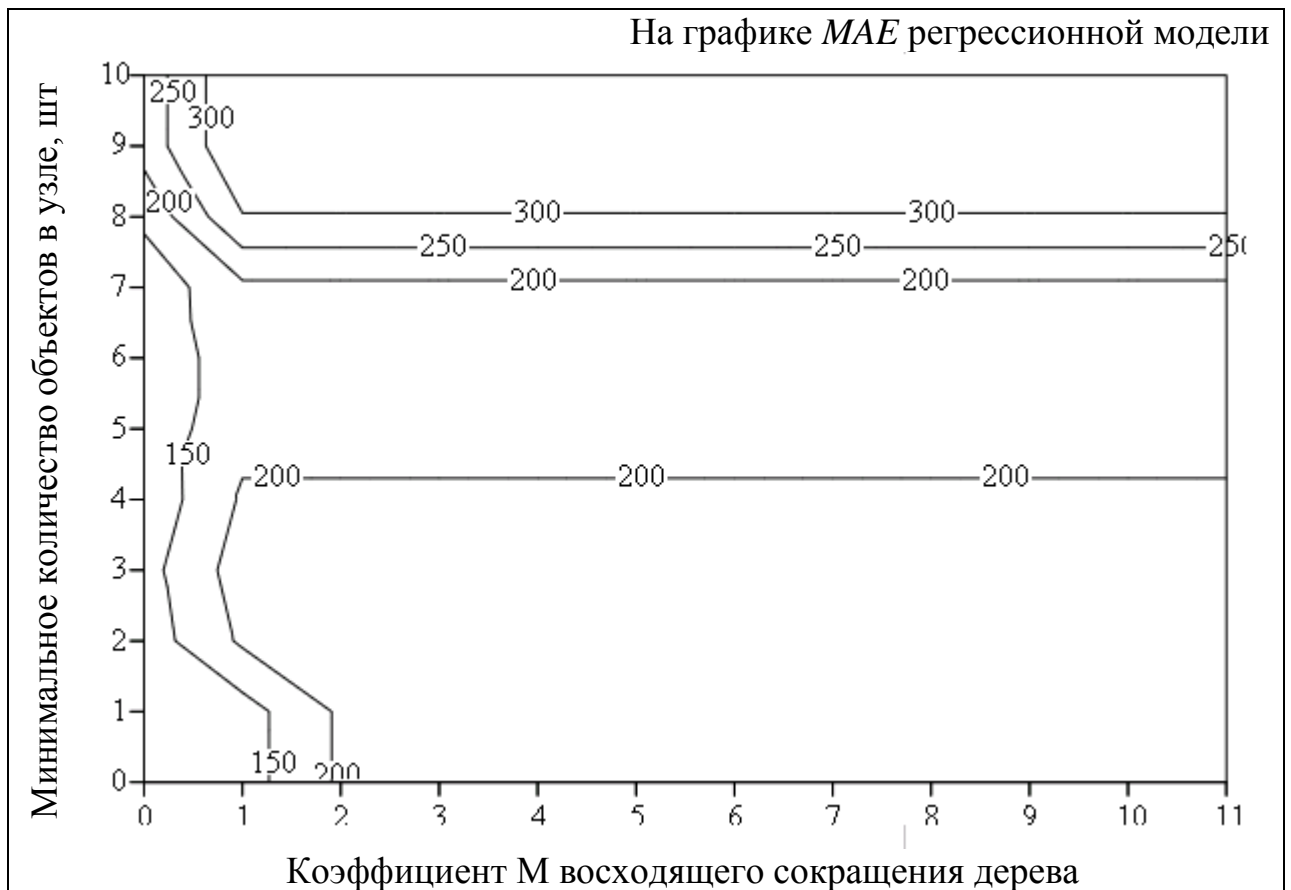


Рисунок 3.4 – Изменение значения *MAE* в зависимости от параметров построения дерева и количества узлов





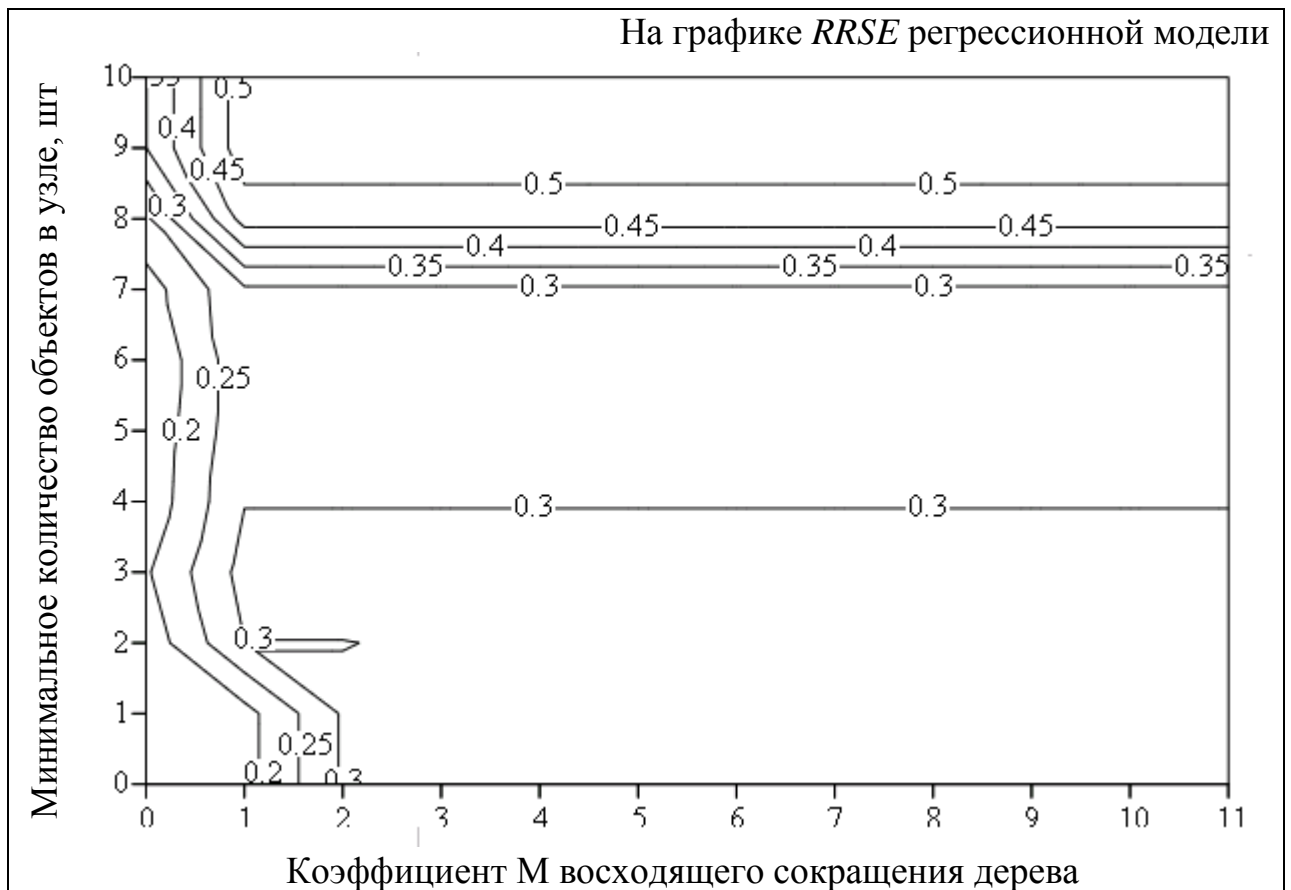


Рисунок 3.6 – Изменение значения *RRSE* в зависимости от параметров построения дерева и количества узлов

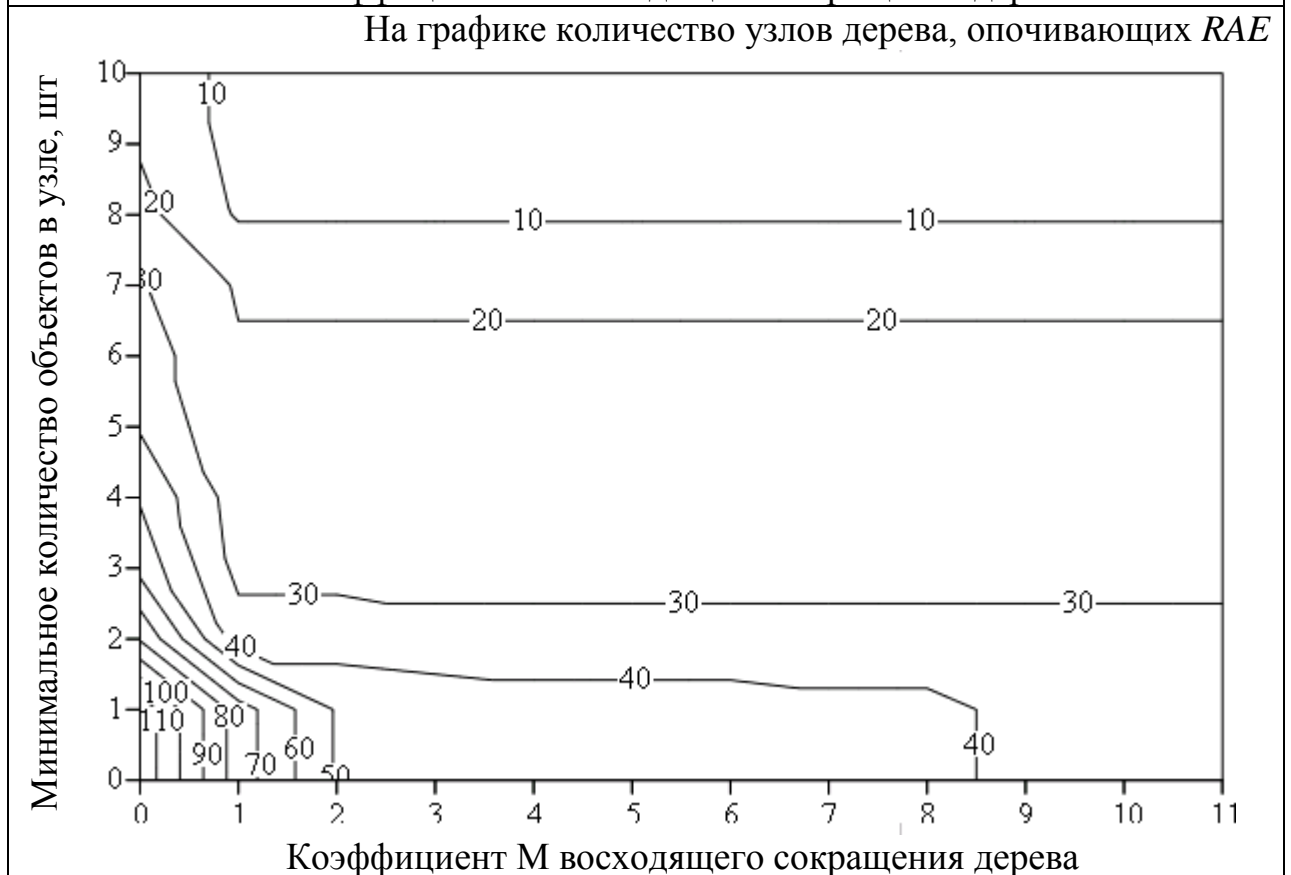
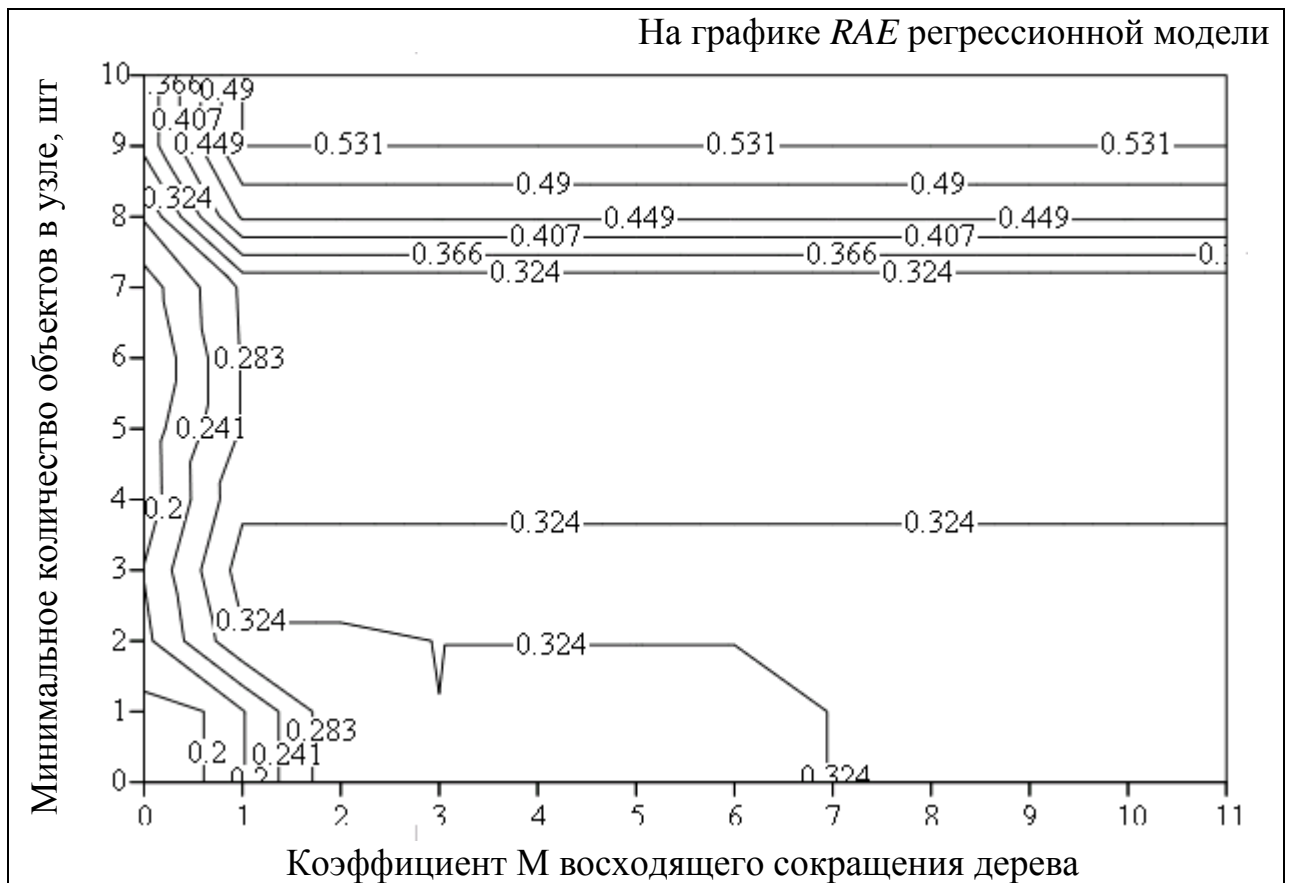


Рисунок 3.7 – Изменение значения *RAE* в зависимости от параметров построения дерева и количества узлов

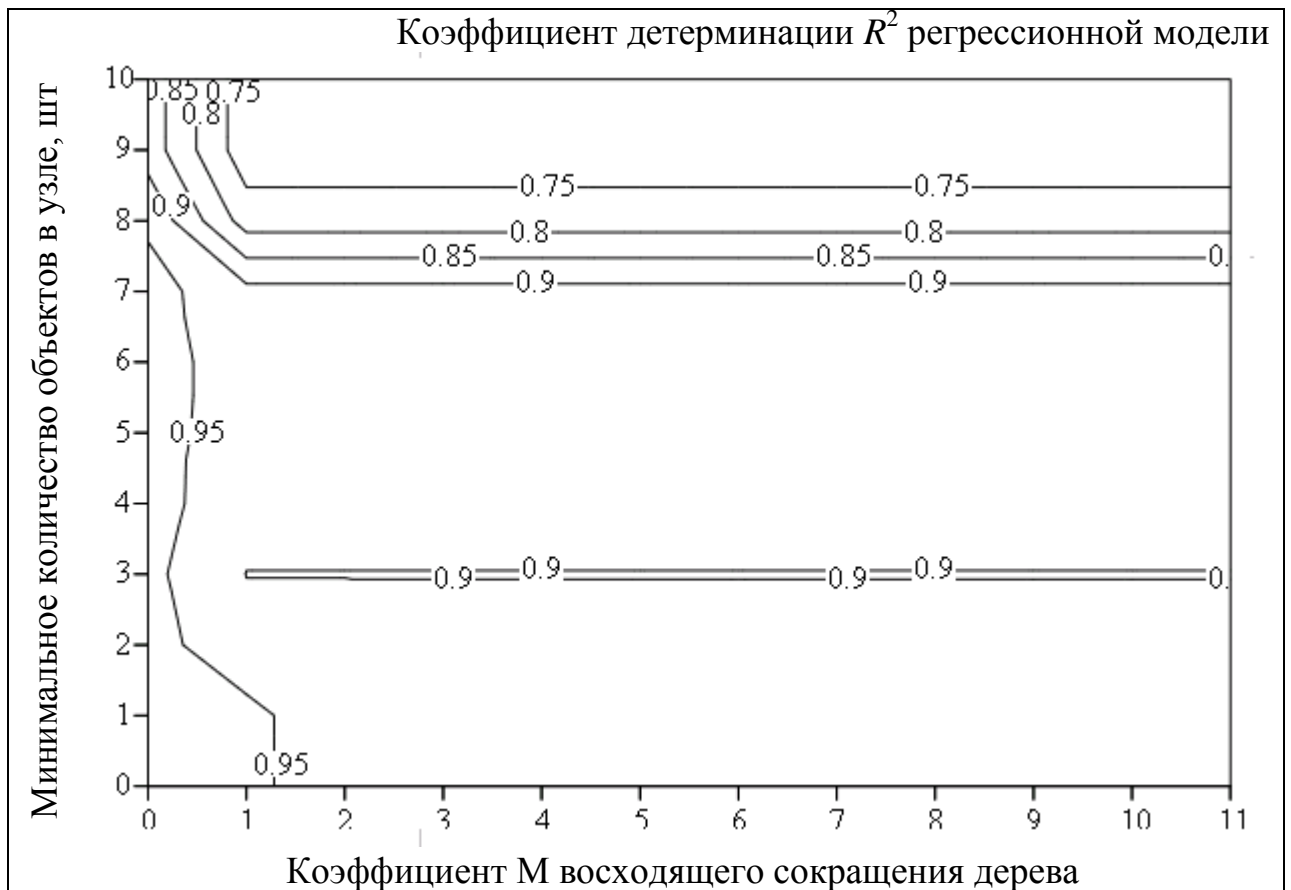


Рисунок 3.8 – Изменение значения коэффициент детерминации  $R^2$  в зависимости от параметров построения дерева и количества узлов

## Заключение

По результатам выполнения выпускной квалификационной работы были сделаны следующие выводы:

1. На основе изучения научных статей по теме исследования установлено, что совершенствование систем диагностики качества контактной сварки возможно за счет применения в них алгоритмов интеллектуального анализа данных. Использование данных алгоритмов позволяет строить модели для контроля качества получаемых соединений по измеряемым в процессе сварки параметрам.

2. Показано, что диагностики качества сварки приводится к решению задачи построения регрессионной модели. При этом входными параметрами регрессионной модели являются измеряемые характеристики процесса сварки, а выходным – прогнозируемые прочностные характеристики.

3. Доказано, что регрессионная модель для оценки качества сварки может быть построена автоматически с использованием алгоритмов индуктивного алгоритма CART. 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  $y$  модели предельная сила  $F$  разрушения сварного соединения на срез, измеряемая в Ньютонах.

4. С использованием технологического оборудования лаборатории “Контактная сварка” института машиностроения ТГУ было проведено около 100 экспериментальных сварок пакета заготовок из стали 08кп. Испытания сварных соединения проводились на оборудовании НИИ “Прогрессивные технологии”. Полученные данные использовались при построении и

тестировании регрессионной модели для прогнозирования силы разрушения сварного соединения при статических нагрузках на срез (приложение А).

5. Исследована точность прогнозирования силы разрушения соединения построенной регрессионной моделью (полное дерево):  $MSE=18520.88 \text{ Н}^2$ ,  $RMSE = 136.0915 \text{ Н}$ ,  $MAE=133.5556 \text{ Н}$ ,  $RSE = 0.0301$ ,  $RRSE = 0.1736$ ,  $RAE = 0.2044$ ,  $R2 = 0.9699$ .

6. Исследовано влияния механизмов получения усеченных деревьев на точность модели. Контурные диаграммы представлены в третьей главе (рисунки 3.2-3.8).

7. Апробация предложенных подходов по построению регрессионных моделей для диагностики качества сварки проводилось в среде Matlab. Результаты апробации доказывают состоятельность предложенных подходов.

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

9. По результатам исследований получен акт внедрения на предприятии ПАО «АвтоВАЗ»

### Список используемой литературы

1. 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
2. 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
3. Sammut, C. Encyclopedia of Machine Learning / Claude Sammut, Geoffrey I. Webb. – Springer US, 2010. – 1031 p.
4. 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.
5. 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
6. 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.
7. El-Banna, M. Online qualitative nugget classification by using a linear vector quantization neural network for resistance spot welding / M. El-Banna, D. Filev, R.B. Chinnam // The International Journal of Advanced Manufacturing Technology. - 2008. - №36. - P. 237-248.
8. 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.
9. Huang, D. Intelligent Computing in Signal Processing and Pattern Recognition / De-Shuang Huang, Kang Li, George William Irwin. – Springer Berlin Heidelberg, 2006 – 1182 p.

10. 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
11. 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
12. 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
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
21. Azar, A.T. Decision tree classifiers for automated medical diagnosis / A.T. Azar, S.M. El-Metwally // *Neural Computing and Applications*, 2013. – №23(7). – pp. 2387-2403
22. Chen, L.S. Extracting knowledge of customers' preferences in massively multiplayer online role playing games / L.S. Chen, P.C. Chang // *Neural Computing and Applications*, 2012. – №23(6). – pp. 1787–1799



Регрессионная модель, полученная с помощью алгоритма CART

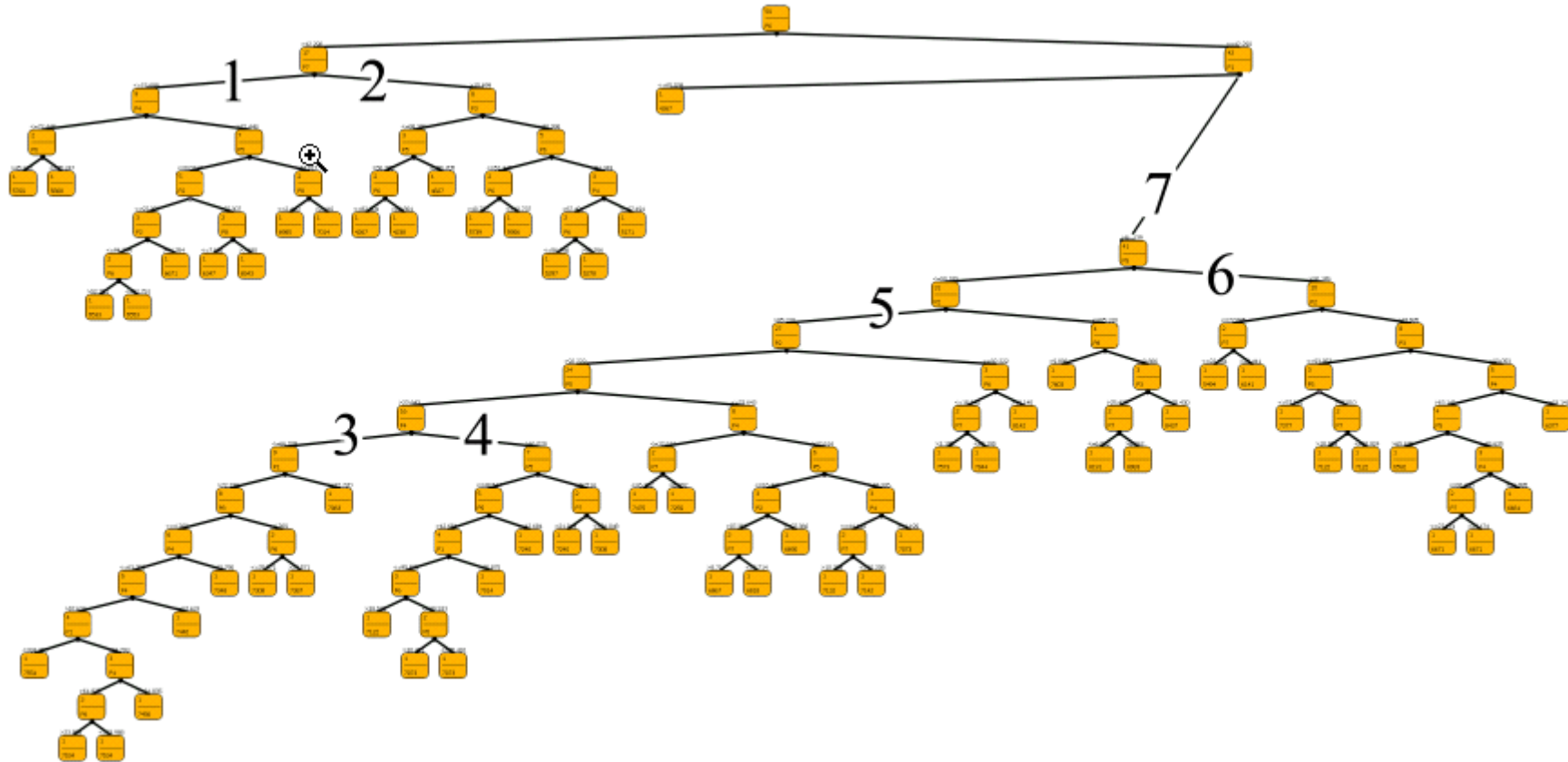


Рисунок А.1 – Обзор регрессионного дерева (цифрами показаны точки перехода)

Так как дерево регрессии получилось достаточно объемным (суммарное количество узлов - 117 штук, а также 59 листов), было принято решение показывать его частями на рисунках А.2 – А.9.

Здесь и далее крупными цифрами показаны точки перехода от одного фрагмента дерева к другому. Для удобства ориентирования по регрессионному дереву на рисунке А.1 представлена его уменьшенное представление.

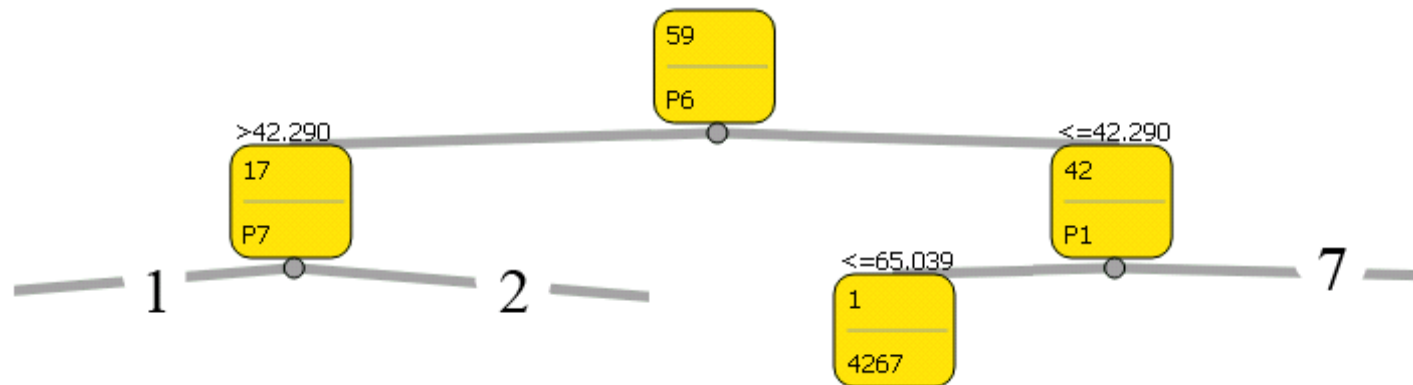


Рисунок А.2 – Корень (цифрами показаны точки перехода)

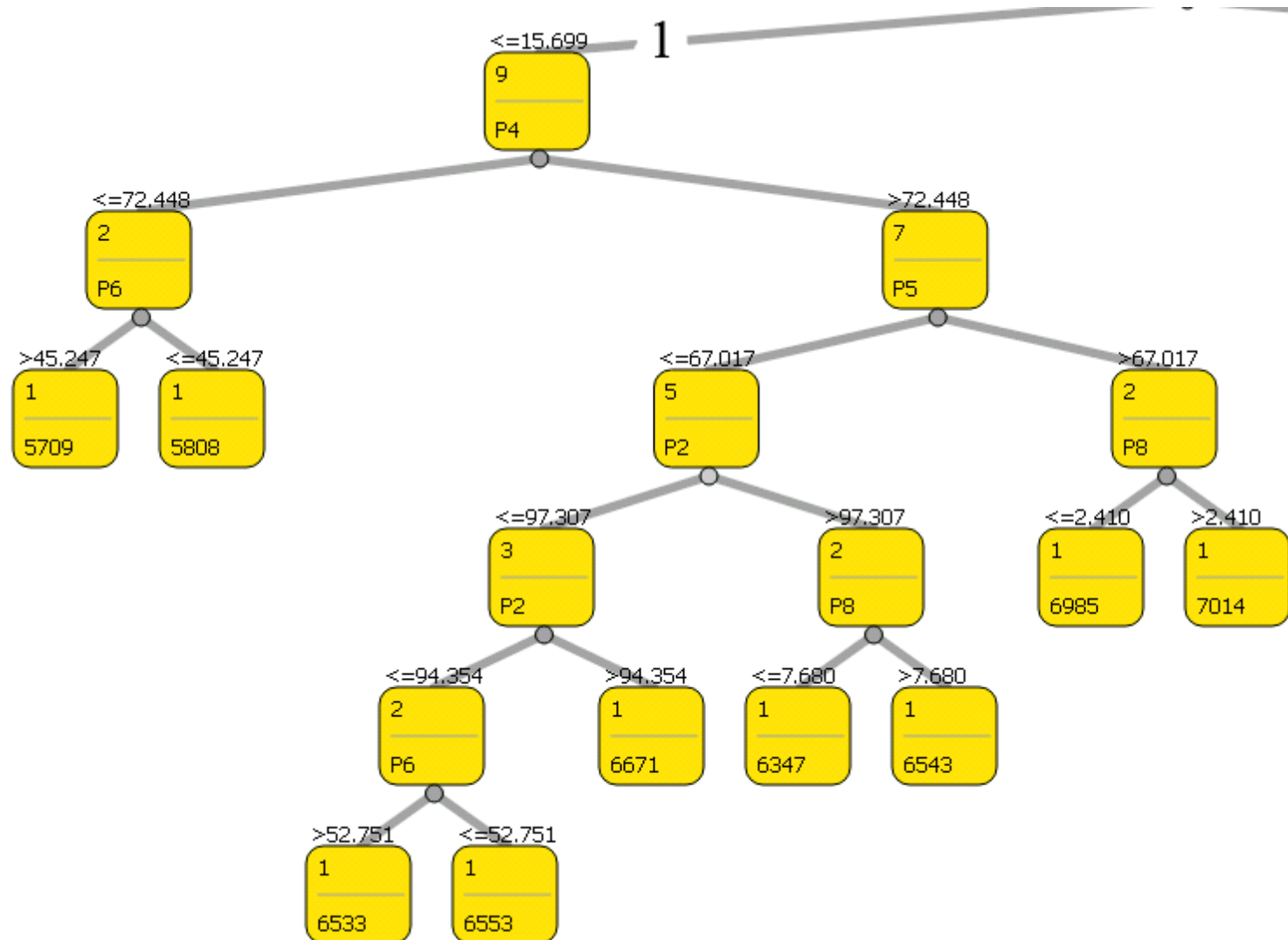


Рисунок А.3 – Продолжение дерева после точки перехода №1

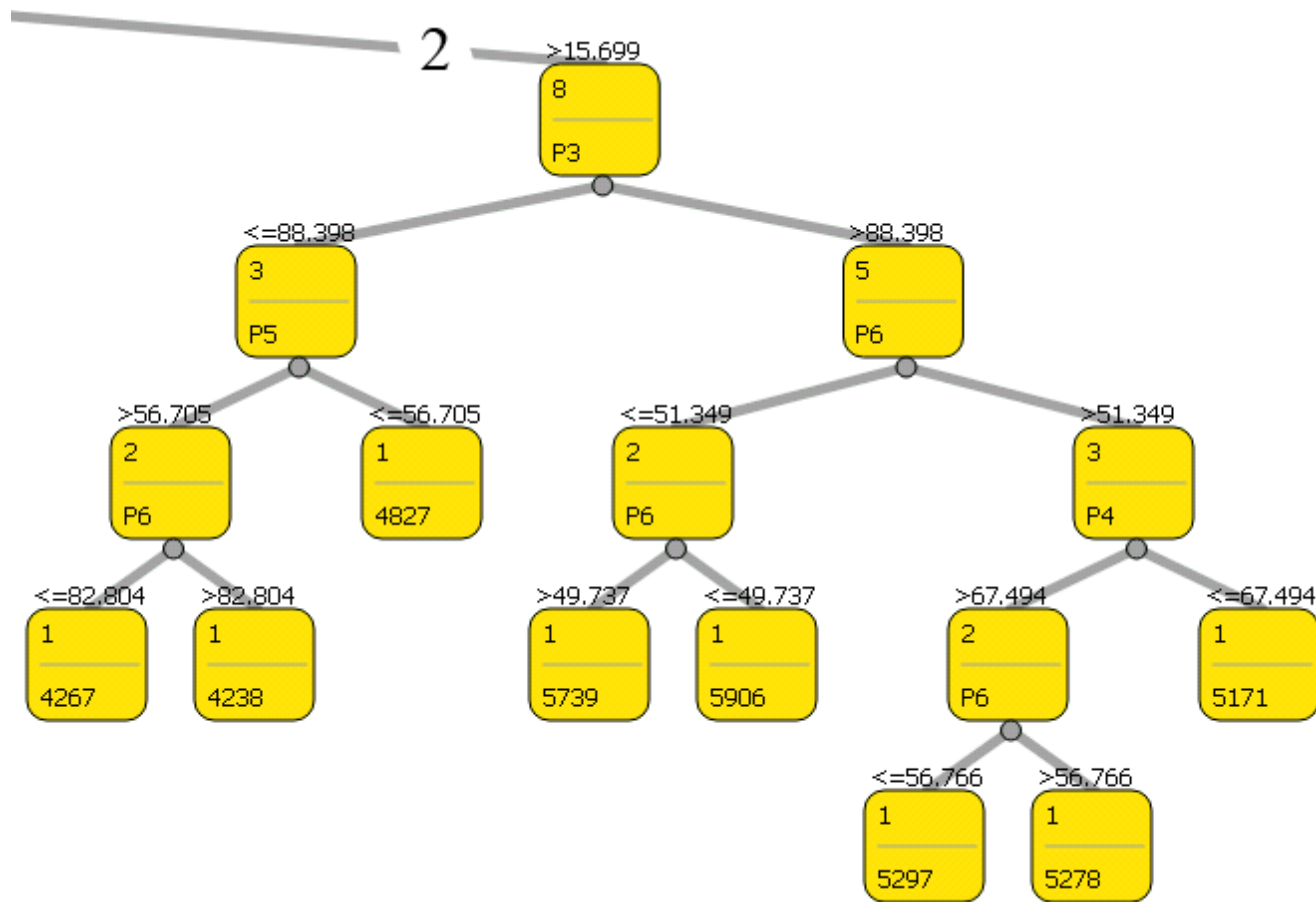


Рисунок А.4 – Продолжение дерева после точки перехода №2

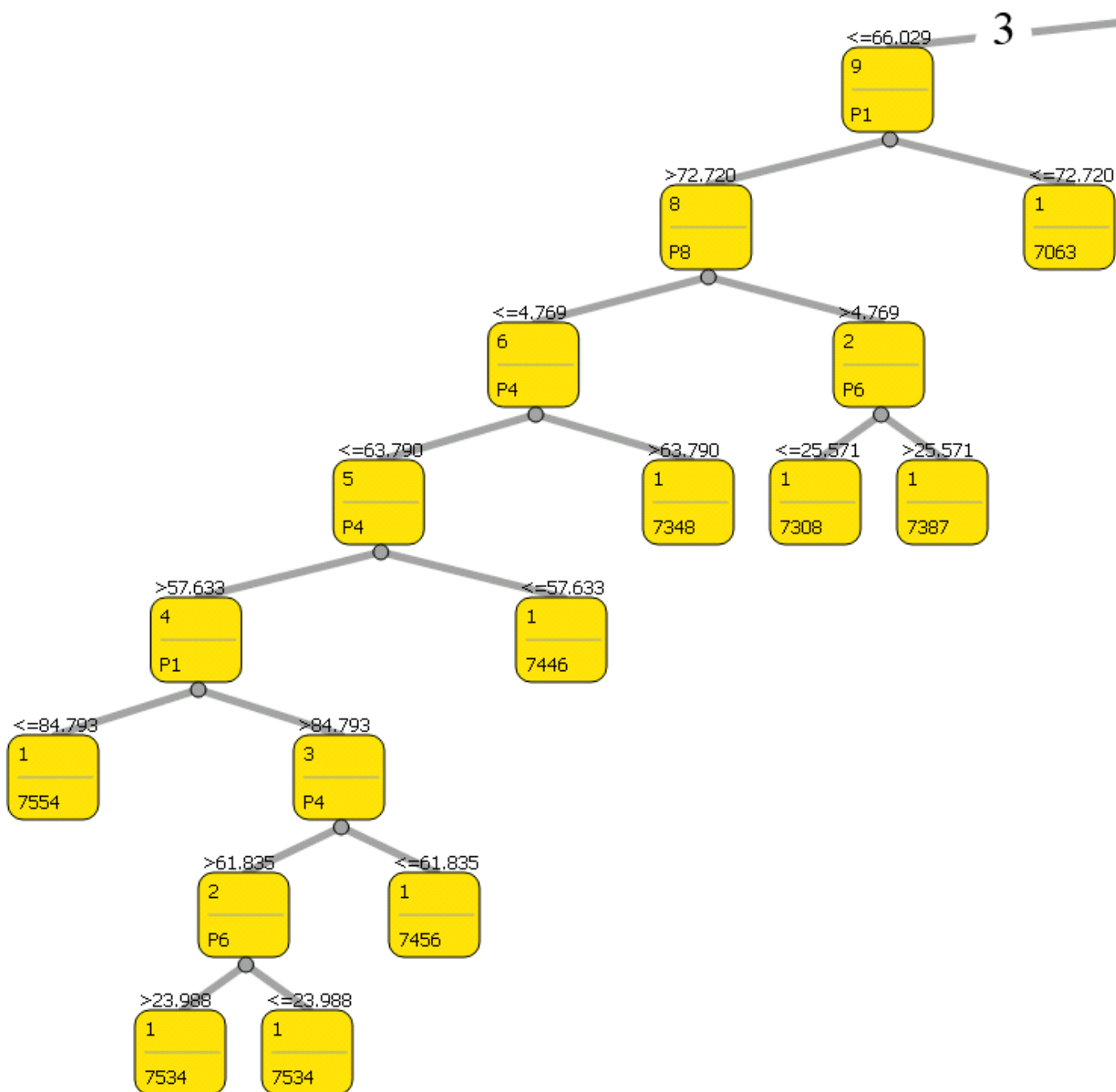


Рисунок А.5 – Продолжение дерева после точки перехода №3

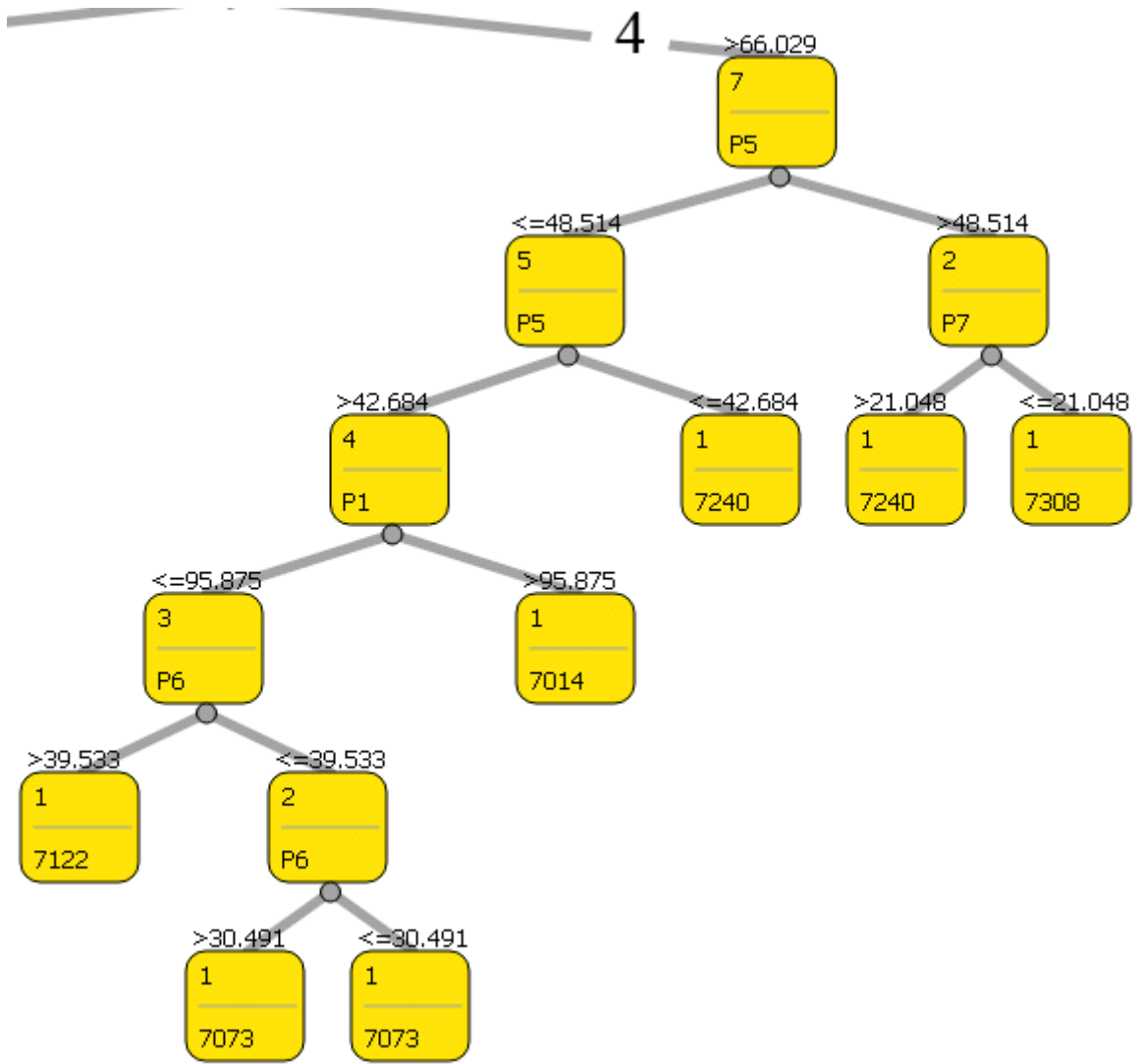


Рисунок А.6 – Продолжение дерева после точки перехода №4



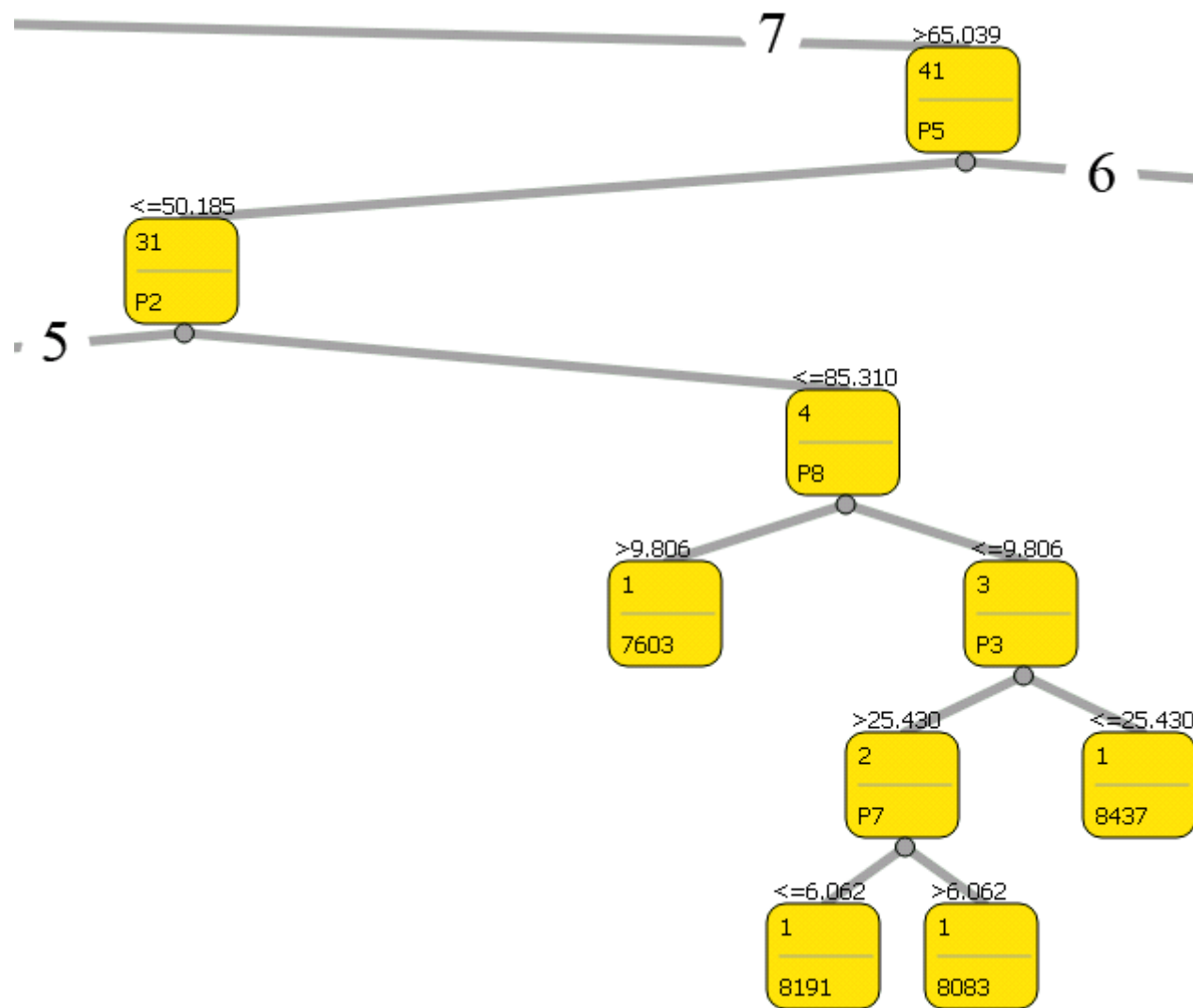


Рисунок А.8 – Продолжение дерева после точки перехода №7



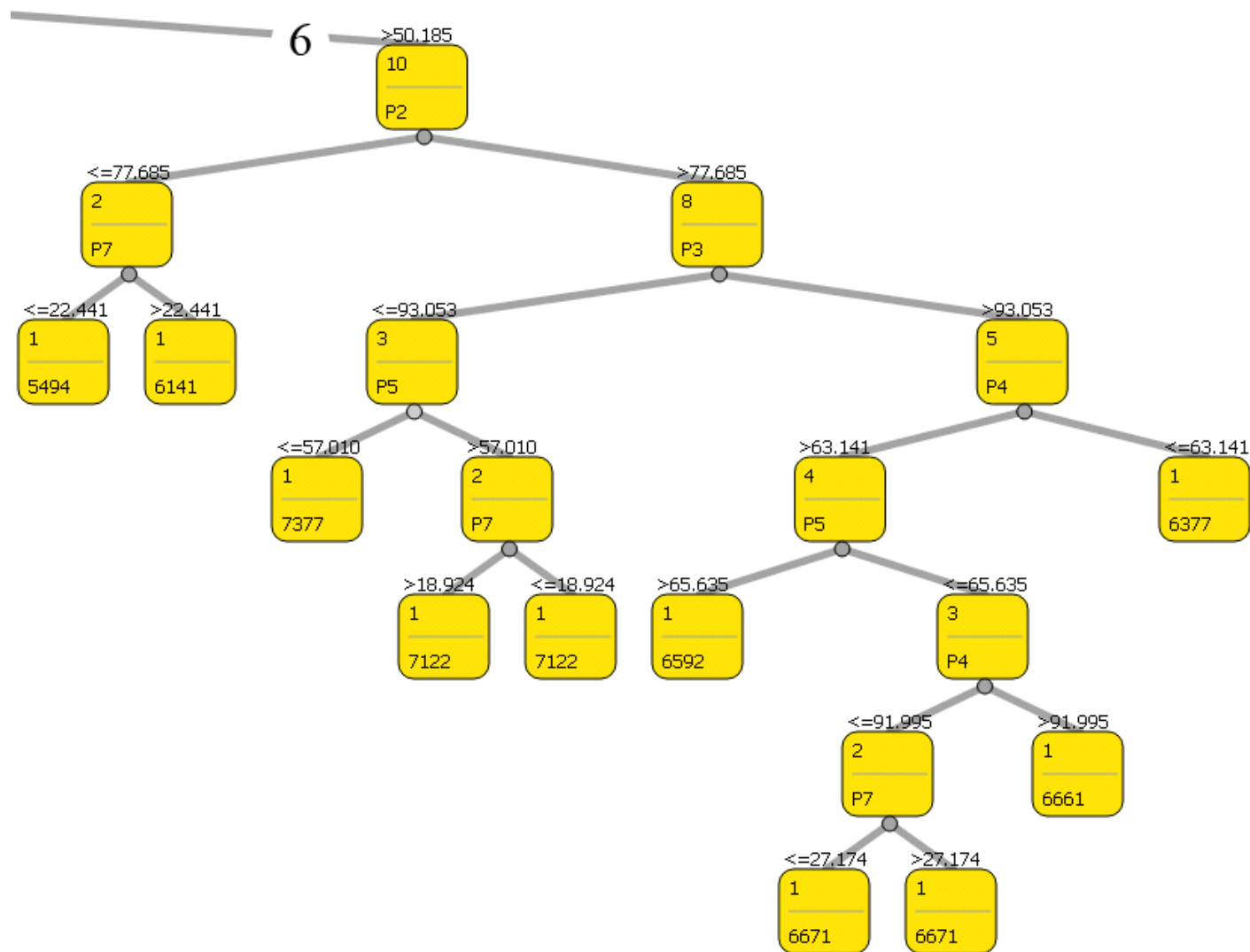


Рисунок А.9 – Продолжение дерева после точки перехода №6