**Краткий обзор**

Для анализа систем хранения данных на принадлежность к нормальным формам используется нейронная сеть, которая обучается генетическим алгоритмом. При таком подходе обучения в качестве генов популяции выступает последовательность весов нейронной сети.

На первом этапе генерируется начальная популяция, гены которой – случайные вещественные числа на промежутке [-1; 1] – значения весов нейронной сети. Далее над популяцией производятся стандартные операции генетического алгоритма.

На втором этапе производится расчет функций приспособленности каждой хромосомы в популяции проверка на наличие решения задачи. Если проверка не пройдена, то алгоритм продолжает работу. Если же решение найдено или превышен лимит итераций – алгоритм завершается.

Третий этап – селекция. В качестве метода отбора особей используется турнирный подход. Из популяции случайным образом выбирается три особи и затем из этой подгруппы выбирается для скрещивания та, значение функции приспособленности которой наилучшая в данной подгруппе. И так повторяется до тех пор, пока размер множества победителей турнира не станет равным размеру популяции.

Далее множество победителей турнира разбивается на 2 группы, и попарно над представителями разных групп производится операция одноточечного кроссинговера. При этом подходе выбирается случайное число X от 1 до N – 1, где N - количество генов в хромосоме. Затем создаются 2 новые последовательности, которые будут генами двух потомков этих родителей. У первого ребенка гены в позициях от 0 до X – гены первого родителя, а от Х до N – второго. У второго потомка - гены в позициях от 0 до X – гены второго родителя, а от Х до N – первого. Это четвертый этап алгоритма – скрещивание.

После размножения выполняется пятый этап генетического алгоритма - операция мутации. В ходе ее выполнения хромосома из популяции с определенной малой вероятностью подвергается мутации – изменения значения одного из генов на вещественное число из промежутка [-0.3; 0.3].

В качестве функции приспособленности выбрана следующая функция:

Это средняя квадратичная ошибка для всех выходных значений нейронной сети при использовании в качестве ее весов одной из хромосомы из текущей популяции генетического алгоритма.

**Обзор приложения**

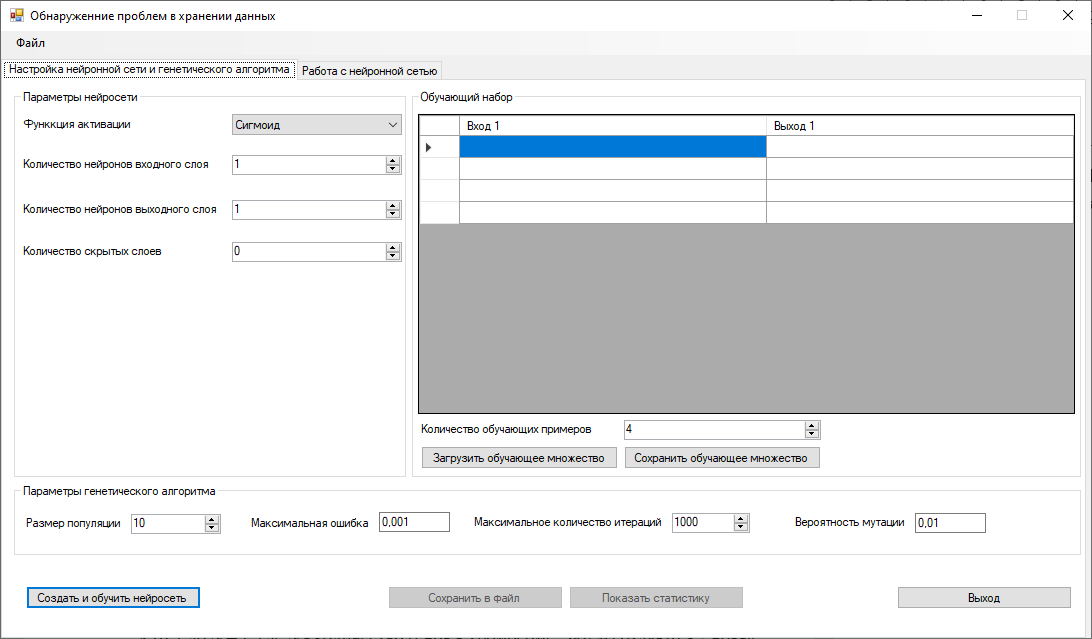


Рисунок 1 – «Первая вкладка главной формы приложения»

В первой вкладке главной формы приложения можно создать и обучить нейронную сеть для обработки данных. Слева находится блок параметров нейронной сети. Здесь можно выбрать функцию активации нейронов, а так же количество и размерность слоев нейронной сети.

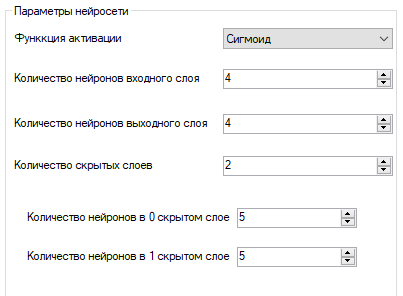


Рисунок 2 – «Блок настройки нейронной сети»

Ниже расположен блок настройки параметров генетического алгоритма для обучения нейронной сети. Пользователь может настроить такие параметры, как размер популяции, максимальная ошибка, то есть максимальное значение функции приспособленности одной из хромосом, при котором алгоритм завершает свою работу и выдает решение. А так же ограничение количества итераций и вероятность мутации.

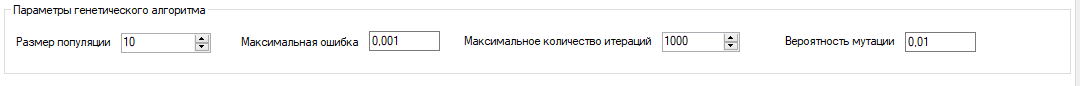


Рисунок 3 – «Блок настройки генетического алгоритма»

В правой части вкладки расположен блок обучающих примеров. Здесь можно вводить примеры для обучения нейронной сети. Так же есть возможность сохранить или загрузить из файла обучающее множество. В данном приложении входы и выходы обучающего множества – вещественные числа.

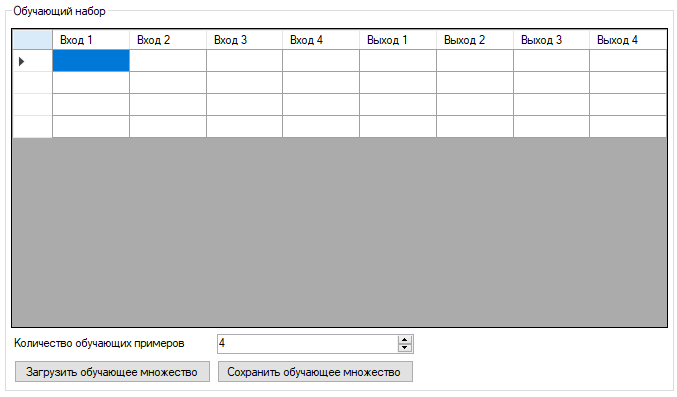


Рисунок 4 – «Блок обучающего набора»

После всех настроек нейронная сеть обучается и после ее обучения можно вызвать окно, где показана статистика обучения нейронной сети. А так же можно сохранить структуру и веса нейронной сети в файл.

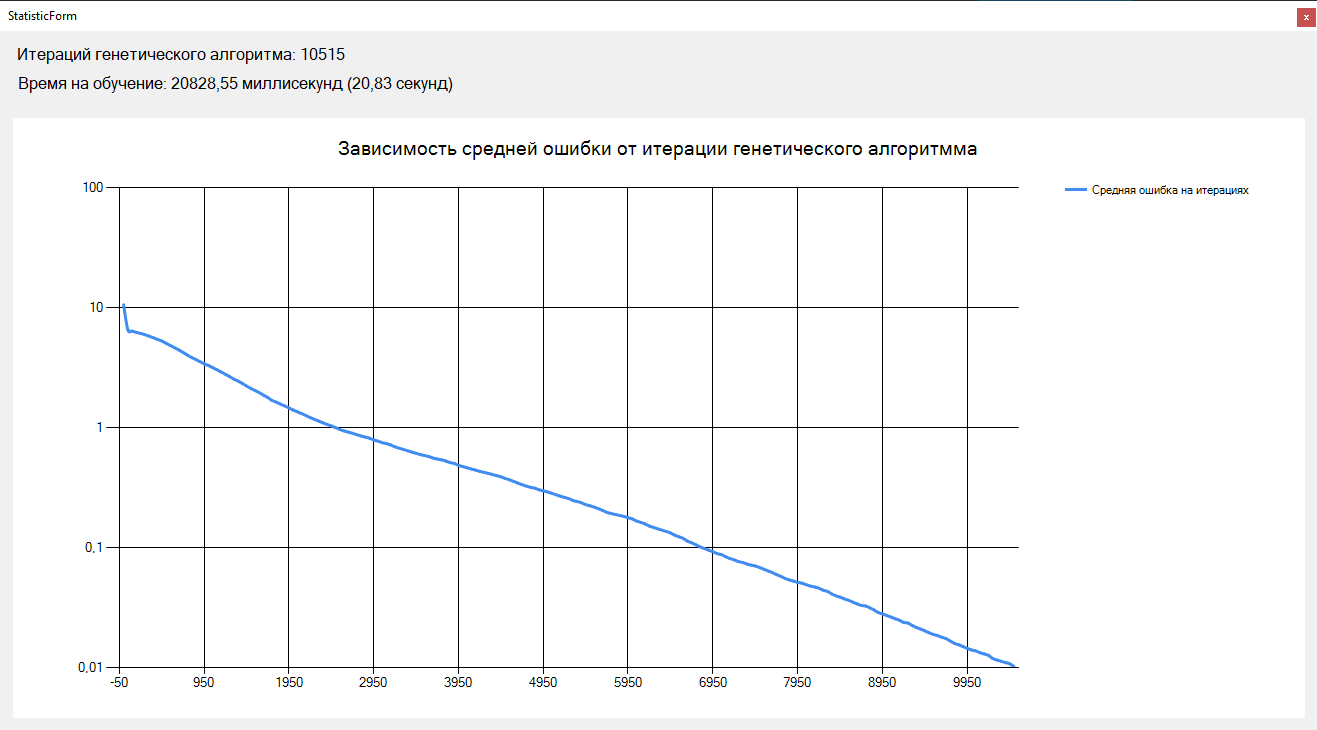


Рисунок 5 – «Статистика обучения»

На графике видно, как изменялось среднее значение функции приспособленности популяции на каждом поколении. А также указано, что нейронная сеть для определения принадлежности отношения к нормальным формам прошла обучение за 20 секунд и за 10000 поколений генетического алгоритма. Для обучения использовались параметры, которые показаны на рисунке 6.

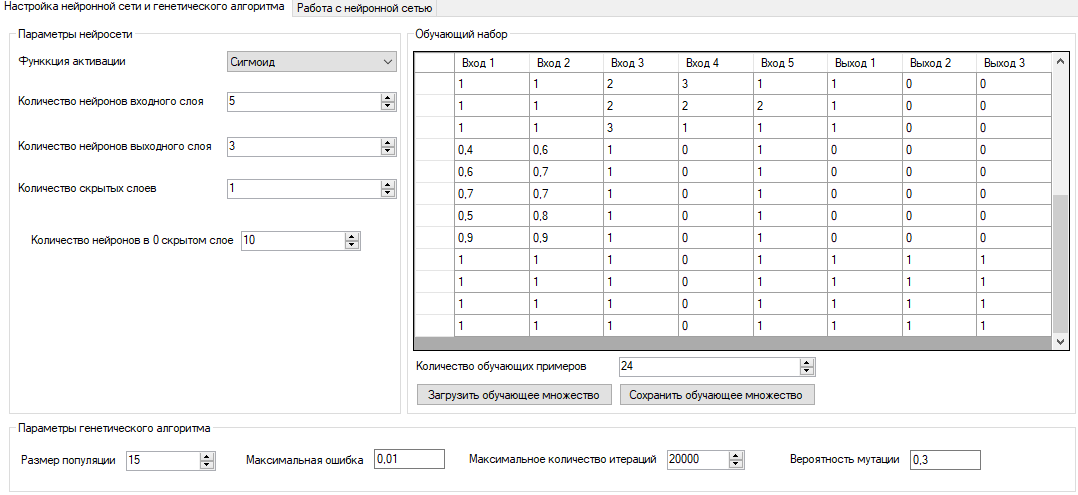


Рисунок 6 – «Параметры обучения»

Нейронная сеть состоит из входного слоя – 5 нейронов, одного скрытого – 10 нейронов и выходного слоя - 3 нейрона. Нейроны на входном слое отвечают за следующие параметры:

1) уникальность значений строк таблицы

2) уникальность значений столбцов таблицы

3) самый длинный путь в связях атрибутов

4) количество отсутствующих связей у атрибутов

5) количество потенциальных ключей отношения

На выходном слое три нейрона выдают результат, где первый нейрон – соответствие 1НФ, второй – 2НФ, 3 – 3НФ соответственно. Если значение на выходе 1 – правила текущей НФ не нарушается, если 0 – правило нарушено.

Для анализа отношения на принадлежность к нормальным формам необходимо загрузить файл, в котором содержатся названия столбцов таблицы, какие в этой таблице ключи, а так же сами данные в этой таблице.

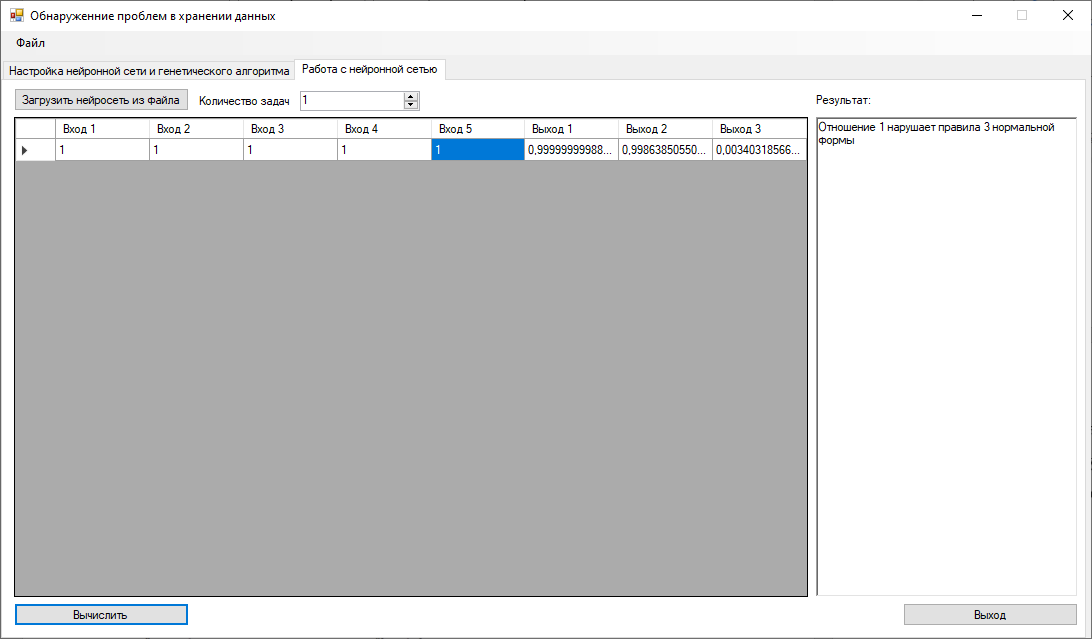


Рисунок 7 – «Вкладка для обработки значений»

В результате анализа в окно результата в правой части вкладки выводится информация о том, правила какой НФ нарушает текущее отношение. Если ни одно из правил не нарушено, то в качестве результата пользователю будет выведена строка «Отношение находится в 3-ей нормальной форме».