ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ МОСКОВСКОЙ ОБЛАСТИ

«Университет «Дубна»

ИНСТИТУТ СИСТЕМНОГО АНАЛИЗА И УПРАВЛЕНИЯ

Кафедра системного анализа и управления

КУРСОВАЯ РАБОТА

по дисциплине

«Теория принятия решений»

Принятие решений в задаче распознавания образов с использованием алгоритма Энтропия

ТЕМА: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(наименование темы)

2255

Выполнил: студент группы \_\_\_\_\_\_\_\_\_\_

Яшина Дарья Валентиновна

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

(Ф.И.О.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись студента)

Руководитель:

Ст. преподаватель Бархатова Ирина Александровна

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ученая степень, ученое звание, занимаемая должность, ФИО)

Дата: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись руководителя)

Дубна, 2022

Содержание

[Введение 3](#_Toc104188552)

[Теоретическая часть 4](#_Toc104188553)

[Метод решения задачи 5](#_Toc104188554)

[Постановка задачи 7](#_Toc104188555)

[Решение поставленной задачи 8](#_Toc104188556)

[Практическая часть 11](#_Toc104188557)

[Реализация алгоритма «Энтропия» 15](#_Toc104188558)

[Вывод 19](#_Toc104188559)

[Список литературы 20](#_Toc104188560)

Введение

При решении задач распознавания на количестве образов большем, чем один, возникала потребность в использовании максимально эффективных алгоритмов, поскольку с увеличением количества образов, объектов, относящихся к этим образам, и свойств ресурсоёмкость задачи распознавания растёт нелинейно быстро.

Одним из наиболее эффективных методов, оперирующих количеством образов превышающим единицу, в теории распознавания образов является алгоритм «Энтропия».

Данный алгоритм предлагает решение задач распознавания на *K* образов, где *K* > 1, причем для каждого объекта формируется свое решающее правило.

Метод позволяет эффективно реализовать подобные задачи, оперируя понятием энтропии. Собственно, энтропия (в свою очередь понятие заимствовано из термодинамики) мера неопределенности или непредсказуемости информации, неопределенность появления какого-либо символа первичного алфавита. При отсутствии информационных потерь численно равна количеству информации на символ передаваемого сообщения. В алгоритме же данное понятие применяется как показатель меры неопределенности отношения материла экзамена к материалу обучения.

Алгоритм очень удобен с той точки зрения, что позволяет воспользоваться математическим аппаратом аналитической геометрии - многомерные объекты (*n*-мерные сферы) и вектора. Основным же преимуществом этого алгоритма является его ориентированность на реализацию индивидуальных правил для каждого объекта, при этом возможно оперировать любыми объектами.

B данной курсовой работе будет исследовано отношение миграции людей, убывших из Российской Федерации и прибывших, на основе решения задачи распознавания образов относительно их заработной платы и их возраста.

Теоретическая часть

У алгоритма «Энтропия» есть некоторые условия применимости:

1. В исходных данных, представленных в виде таблице, присутствуют представители всех образов.
2. Для каждого объекта указана его принадлежность к образу.
3. В процессе распознавания определяется принадлежность объектов экзамена к одному из образов.
4. Таблица должна быть без пропусков; свойства - арифметические, логические 1-го и 2-го рода.

Стоит заметить, что метод «Энтропия» рекомендуется использовать в случаях, когда объекты образов перемешаны или объекты одного образа находятся в окружении объектов другого образа (рис. 1).

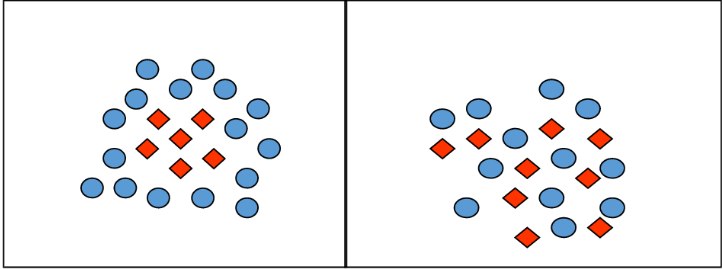


Рис.1. Расположение объектов относительно их образов

Метод решения задачи

Этот метод основан на том, что для каждого объекта из материала экзамена требуется свое решающее правило. Для чего вокруг каждого объекта экзамена описывается система концентрических сфер. Для каждой из этих сфер определяется функция энтропии, характеризующая преобладания точек одно из образов этой сфере. Результат определяется по той сфере, где значение функции оптимально. Объект экзамена относится к тому объекту образа, который в сфере преобладает. (рис. 2).

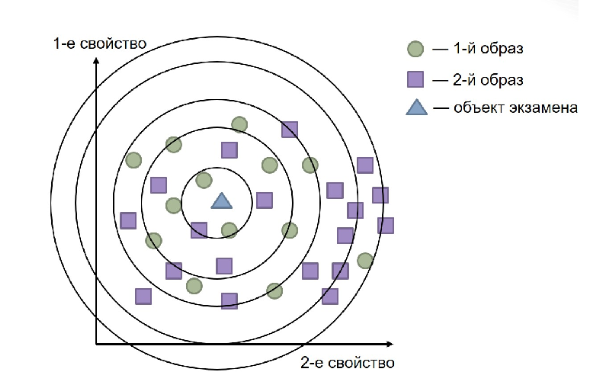


Рис.2. Пример построения сфер

Рассмотрим данный алгоритм более подробно. На первом этапе необходимо вычислить матрицы мер сходства по каждому свойству, а затем — общую матрицу мер сходства . Из общей матрицы мер сходства  выбирается максимальная  и минимальная  меры сходства. (рис.3).

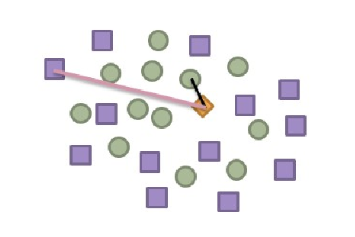


Рис.3. Максимальная и минимальная мера сходства

На следующем этапе введем некоторый шаг

 ,

где  — заданное нами постоянное число сфер.

Вычисляется мера сходства для каждого объекта экзамена  с каждым объектом МО . Затем выбираем все объекты (рис.4), для которых мера сходства с объектами МО удовлетворяют условию

 , 0,1...

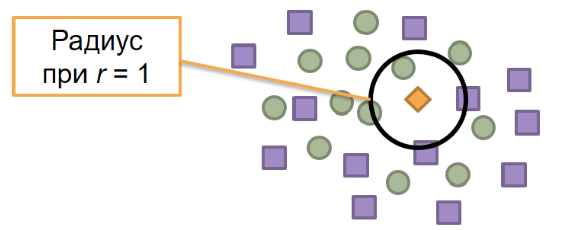


Рис.4. Пример сферы (2-х мерного образа) при радиусе равном единице

В результате получаем набор вложенных сфер (концентрических). Для каждой из этих сфер мы вычисляем энтропию

,

где — число объектов -го образа, попавших в - ую сферу;

 — число объектов, попавших в - ую сферу.

Из всех полученных энтропий  выбираем минимальную . Затем находим номер сферы с минимальной энтропией . Эта сфера, в которой объектов одного образа намного больше объектов другого образа, т. е. .

Последним этапом решения является распознавание. Распознавание производится при помощи правила Байеса. Объект  будет отнесен к -му образу, если

,

где — число объектов -го образа в -ой сфере, — общее число объектов -го образа,  — число объектов -го образа в -ой сфере, — общее число объектов -го образа. Иначе, если

,

то объект  будет отнесен к -му образу.

Постановка задачи

Таблица 1. ТОС 1образца Таблица 2. ТОС 2 образца

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Образец 1 | |  | Образец 2 | |
| свойство 1 | свойство 2 |  | свойство 1 | свойство 2 |
| **Возраст** | **Заработная плата** |  | **Возраст** | **Заработная плата** |
| 22 | 125 727,35 р. |  | 18 | 143 625,35 р. |
| 51 | 96 301,51 р. |  | 56 | 68 810,69 р. |
| 49 | 59 355,96 р. |  | 33 | 101 338,56 р. |
| 48 | 108 824,67 р. |  | 54 | 162 441,25 р. |
| 41 | 116 989,25 р. |  | 31 | 131 238,04 р. |
| 34 | 134 967,57 р. |  | 40 | 115 138,89 р. |
| 50 | 165 408,31 р. |  | 46 | 52 799,71 р. |
| 48 | 95 042,21 р. |  | 52 | 39 344,33 р. |
| 25 | 105 839,47 р. |  | 33 | 131 126,17 р. |
| 44 | 122 426,78 р. |  | 48 | 134 677,97 р. |
| 26 | 102 632,24 р. |  | 26 | 88 655,88 р. |
| 37 | 47 829,75 р. |  | 39 | 52 319,77 р. |
| 47 | 65 565,58 р. |  | 22 | 104 352,92 р. |
| 20 | 109 018,97 р. |  | 27 | 74 716,26 р. |
| 40 | 119 239,50 р. |  | 32 | 111 899,15 р. |
| 36 | 105 027,57 р. |  | 54 | 104 934,63 р. |
| 39 | 50 454,82 р. |  | 40 | 99 625,89 р. |
| 68 | 107 419,98 р. |  | 58 | 74 404,50 р. |
| 45 | 124 412,70 р. |  | 40 | 92 964,16 р. |
| 40 | 115 893,15 р. |  | 39 | 82 832,95 р. |
| 47 | 152 474,75 р. |  | 46 | 85 831,63 р. |
| 29 | 86 396,60 р. |  | 57 | 113 603,43 р. |
| 41 | 27 975,11 р. |  | 50 | 113 438,02 р. |
| 55 | 102 411,16 р. |  | 52 | 46 146,78 р. |
| 42 | 32 712,85 р. |  | 28 | 120 433,17 р. |
| 45 | 107 784,91 р. |  | 34 | 99 440,04 р. |

Имеются данные опросов на границе Российской Федерации от людей, которые решили покинуть страну или же в нее приехать на постоянное место жительства в таблице. Основные критерии – Возраст мигранта и его заработная плата, до момента миграции. (табл.1, табл.2, табл.3)

Таблица 3. ТОС материала экзамена

|  |  |  |
| --- | --- | --- |
| **Образец** | **Возраст** | **Заработная плата** |
| **A** | 25 | 60 000,00 р. |
| **B** | 80 | 82 000,00 р. |
| **C** | 18 | 35 000,00 р. |
| **D** | 50 | 90 000,00 р. |
| **E** | 36 | 120 000,00 р. |

Решение поставленной задачи

Сначала нанесем на график наши исходные данные двух образов (материалов обучения) и материалов экзамена. (рис.5)

Рис.5. Объекты материалов обучения

Как видно на рисунке необходимо определить к какому из образов относится человек (МЭ).

Во-первых, находим минимум и максимум из свойств объектов образов.

|  |  |  |
| --- | --- | --- |
| *min* | 15 | 25524 |
| *max* | 68 | 165408 |

Затем находим матрицу мер сходства по каждому свойству и вычисляем общую матрицу мер сходства:

Таблица 4. Фрагмент таблицы мер сходства материалов обучения

|  |  |  |
| --- | --- | --- |
| **Мер. Сход 1** | **Мер. Сход 2** | **µ** |
| 1 | 1 | 1 |
| 0,998983379 | 0,822082258 | 0,910533 |
| 0,99904166 | 0,598698081 | 0,79887 |
| 0,999079957 | 0,897801154 | 0,948441 |
| 0,999344711 | 0,947166715 | 0,973256 |
| 0,999564997 | 0,944130808 | 0,971848 |
| 0,998996927 | 0,760076706 | 0,879537 |
| 0,999083028 | 0,814468156 | 0,906776 |
| 0,999913197 | 0,879751713 | 0,939832 |
| 0,999227534 | 0,980043683 | 0,989636 |

Выбираем максимальную и минимальную меру сходства для каждого объекта экзамена:

* Минимальный – 0,584348486.
* Максимальный – 0,999626729.

Постоянную *l* задаем равную 3 и вычисляем шаг: *h* = 0,138426081.

Для каждого объекта экзамена вычисляем энтропию (табл.5), отыскиваем минимальную энтропию и производим распознавание.

Для каждой сферы вычисляем показатели энтропии (меры неопределенности) по формуле:

.

Где — число объектов -го образа, попавших в -ую сферу;  — число объектов, попавших в -ую сферу.

Для данного случая получаем *Hmin*= 0,28806973. Следовательно, распознавание

необходимо проводить по данным первой сферы.

Таблица 5. Фрагмент таблицы вычислений энтропии

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **МЭ\_1** |  |  |  |
|  |  | I1 | I2 | I3 |
| 1 | 0,74 | 0 | 1 | 1 |
| 2 | 0,66 | 0 | 1 | 1 |
| 3 | 0,80 | 1 | 1 | 1 |
| 4 | 0,64 | 0 | 1 | 1 |
| 5 | 0,67 | 0 | 1 | 1 |
| 6 | 0,65 | 0 | 1 | 1 |
| 7 | 0,41 | 0 | 0 | 1 |
| 8 | 0,69 | 0 | 1 | 1 |
| 9 | 0,83 | 1 | 1 | 1 |
| 10 | 0,62 | 0 | 1 | 1 |
| 16 | 0,60 | 0 | 1 | 1 |
| 17 | 0,74 | 0 | 1 | 1 |
| 18 | 0,68 | 0 | 1 | 1 |
| 19 | 0,76 | 1 | 1 | 1 |
| 20 | 0,80 | 1 | 1 | 1 |
| 21 | 0,74 | 0 | 1 | 1 |
| 22 | 0,55 | 0 | 1 | 1 |
| 23 | 0,60 | 0 | 1 | 1 |
| 24 | 0,73 | 0 | 1 | 1 |
| 25 | 0,76 | 1 | 1 | 1 |
| 26 | 0,79 | 1 | 1 | 1 |
|  | Hr | 0,28807 | 0,300653 | 0,30103 |

Из , следует, что объект А относиться к первому образу. Аналогичным способом переходим к распознаванию всего материала экзамена целиком (табл.6), получим:

Таблица 6. Распознавание материалов экзамена

|  |  |
| --- | --- |
| **Образец** | **Возраст** |
| **A** | 1 обр |
| **B** | 2 обр |
| **C** | 1 обр |
| **D** | 2 обр |
| **E** | 1 обр |

Практическая часть

В ходе выполнения данной работы была написана программа реализующая алгоритм распознавания образов методом «Энтропия».

Данная программа определяет принадлежность предоставленного объекта, имеющего 2 свойства, к двум предоставленным ранее образам. Для того чтобы реализовать данное распознавание необходимо ввести материал обучения первого и второго образа, а также объекты распознавания. Для того чтобы ввести данные материала обучения и материала экзамена можно воспользоваться двумя реализованными способами

ввода:

1. Сгенерировать данные;

2. Чтение данных из *Excel* (см. рис. 6).

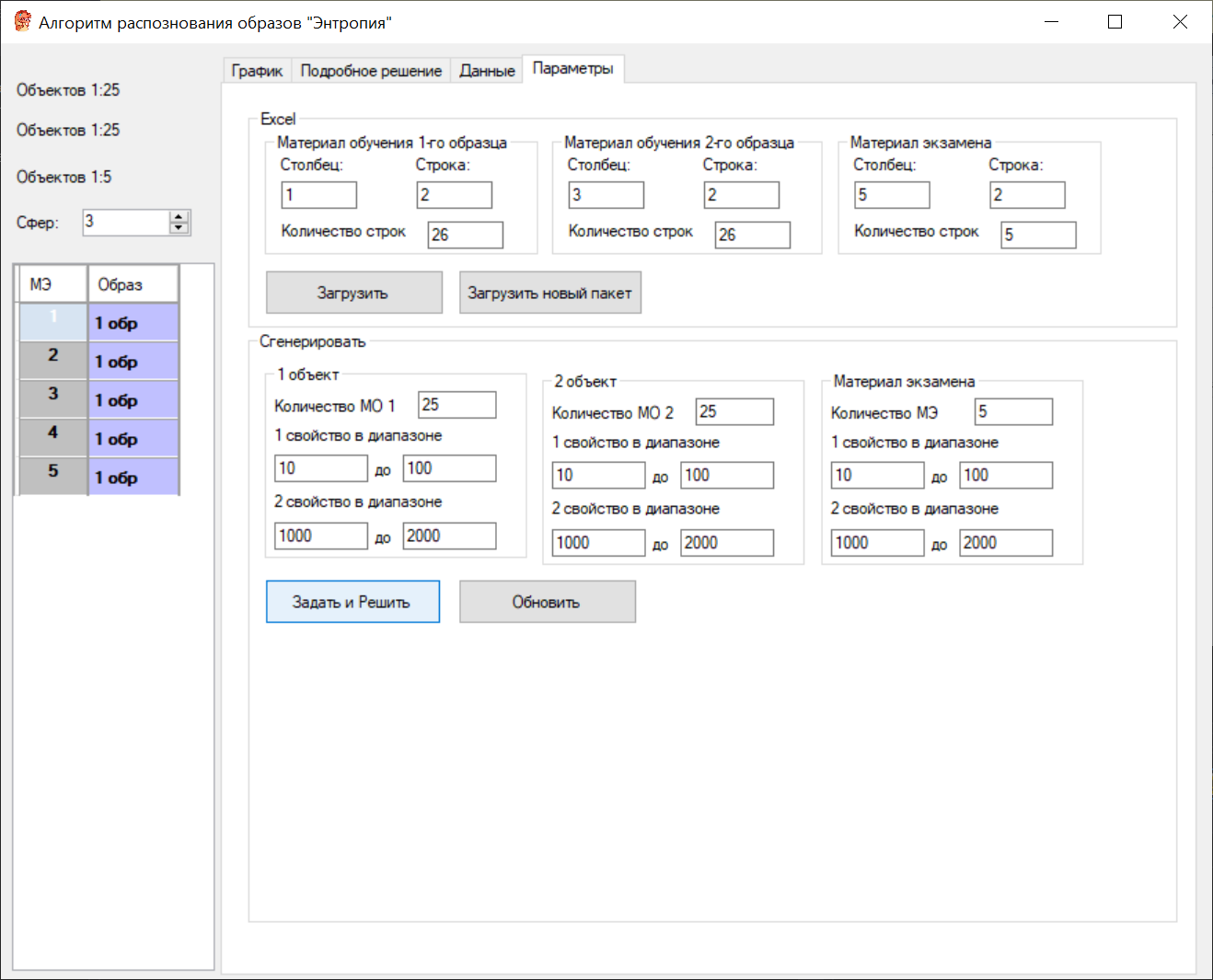


Рис.6. Вкладка программы для ввода данных «Параметры»

Для того чтобы сгенерировать данные используется вкладка «Параметры» (см. рис. 6). Пользователю предлагается сгенерировать значения исходя из диапазона, который он может задать. Также пользователю предлагается выбрать необходимое количество объектов образов и материала экзамена. При нажатии кнопки на экран выводятся результаты вычислений. В отдельной вкладке «Данные» (см. рис. 7) Пользователь может проверить введённые или сгенерированные данные.

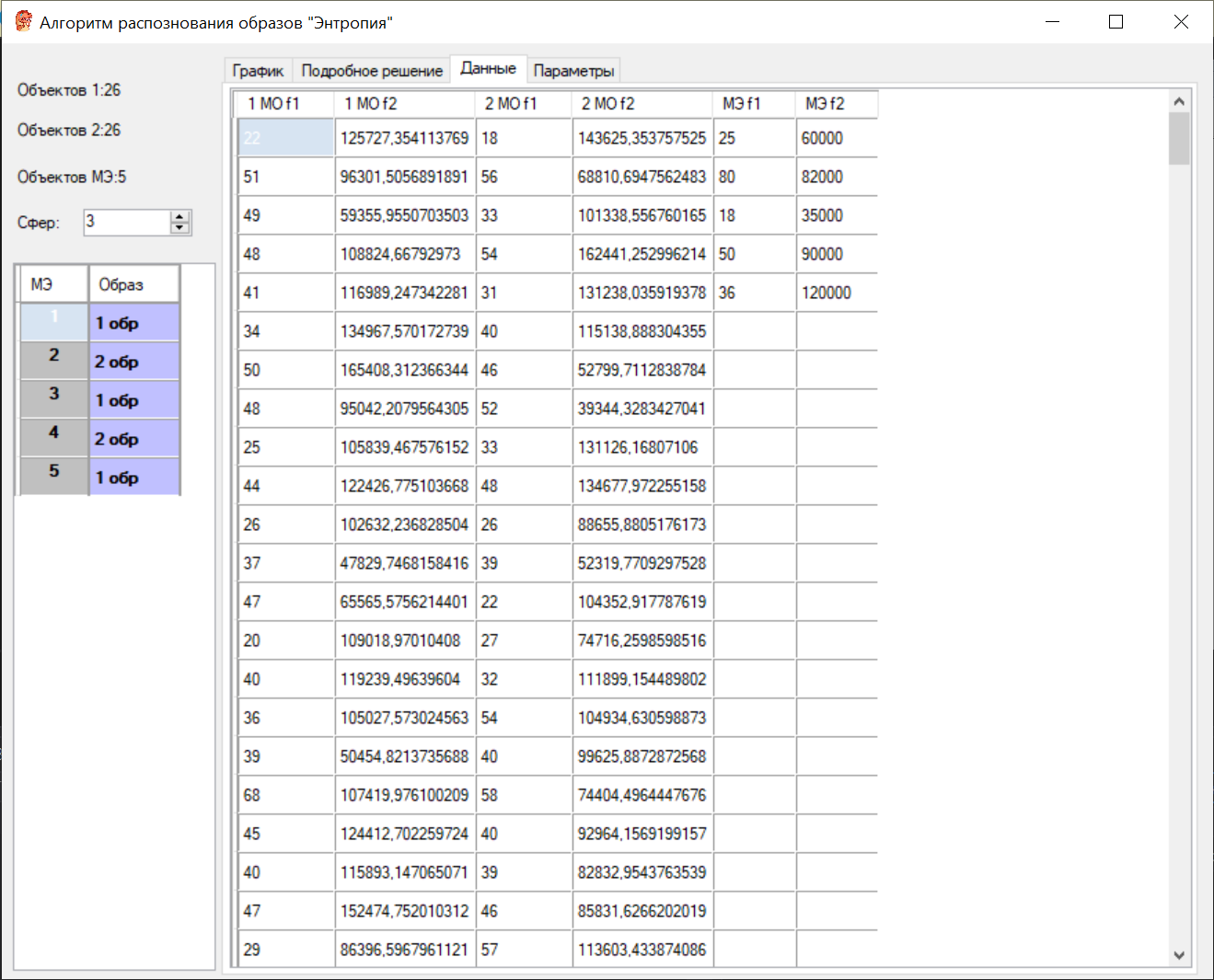


Рис.7. Вкладка программы «Данные»

Для удобства и наглядности работы алгоритма добавлена вкладка «График» (см. рис. 8). После загрузки или генерации данных график строится автоматически. При изменении данных, график так же автоматически перестраивается при поиске нового решения.

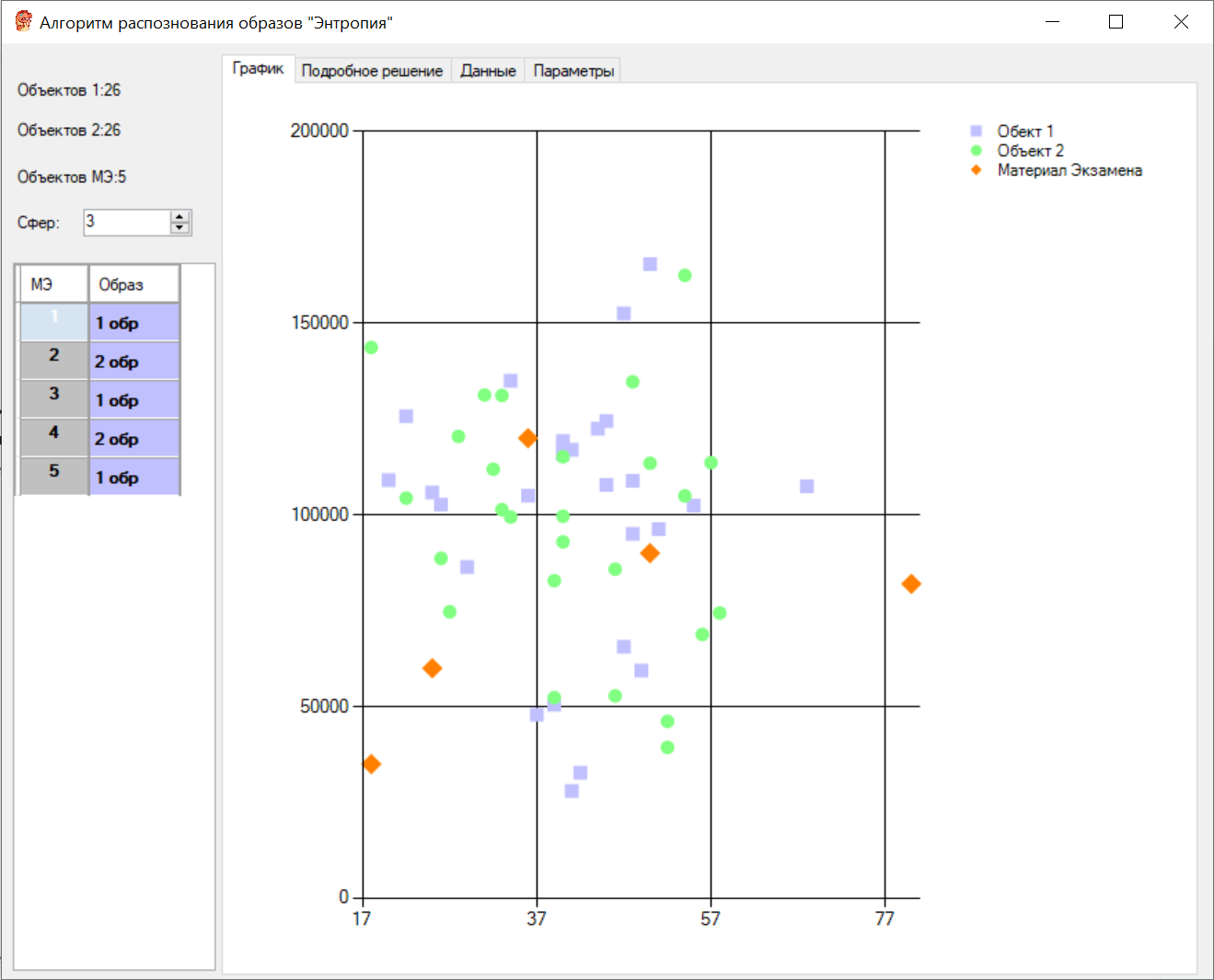


Рис.8. Вкладка программы «График»

Для того чтобы распознать образы материала экзамена пользователю нужно указать количество колец и нажать на кнопку на вкладке «Параметры».

После нажатия на форме появится таблица принадлежности объектов к образу.

Если есть потребность посмотреть промежуточные результаты расчетов, то по нажатию на вкладку «Подробное решение» на экран будет выведена форма с промежуточными результатами вычислений (см. рисунок 9).

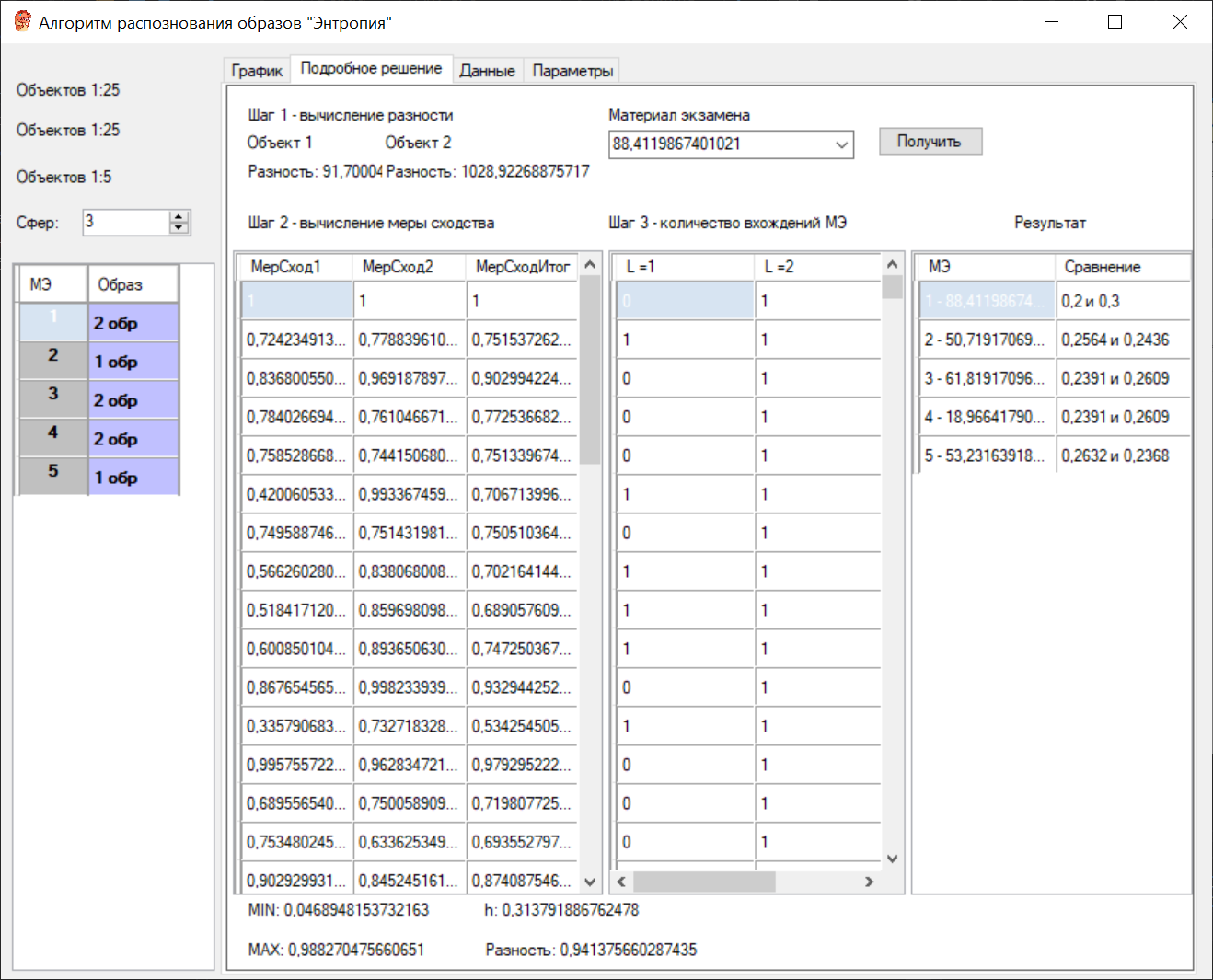


Рис.9. Вкладка программы «Подробное решение»

Реализация алгоритма «Энтропия»

Первым шагом данного алгоритма является построение матриц свойств каждого из образов (см. рис. 10).

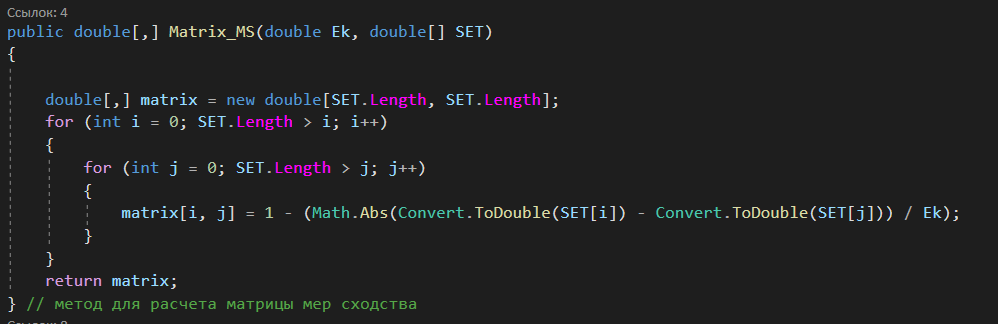


Рис.10. Метод для расчета матрицы мер сходства

По полученным матрицам находим общую меру сходства (см. рис. 11).

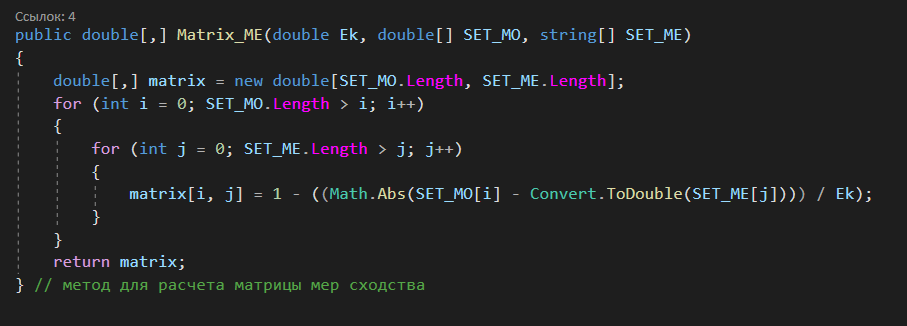


Рис.11. Метод для расчета общей матрицы мер сходства

После нахождения матрицы мер сходства мы можем найти максимум и минимум и рассчитать шаг. Далее находим количество вхождений объектов материала обучения попавших в определенный круг и по формуле вычисляем энтропию. В итоге получаем матрицу в соотношении количества сфер на количество материалов экзамена (см. рис. 12).

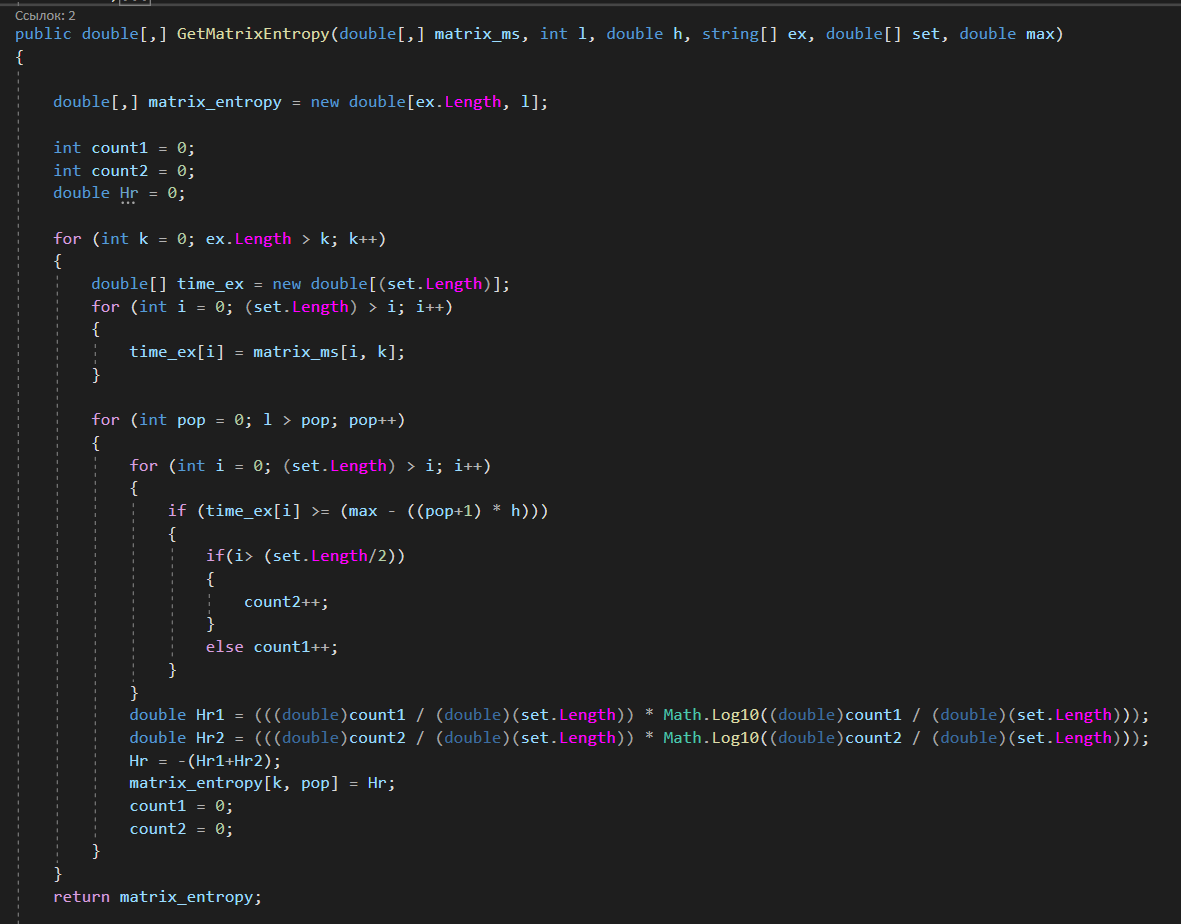


Рис.12. Метод для нахождения матрицы Энтропии

После нахождения матрицы мы с помощью правила Байеса находим к какому образу принадлежит каждый материал экзамена. Подсчитывая количество вхождений элементов и сравнивая их между собой.

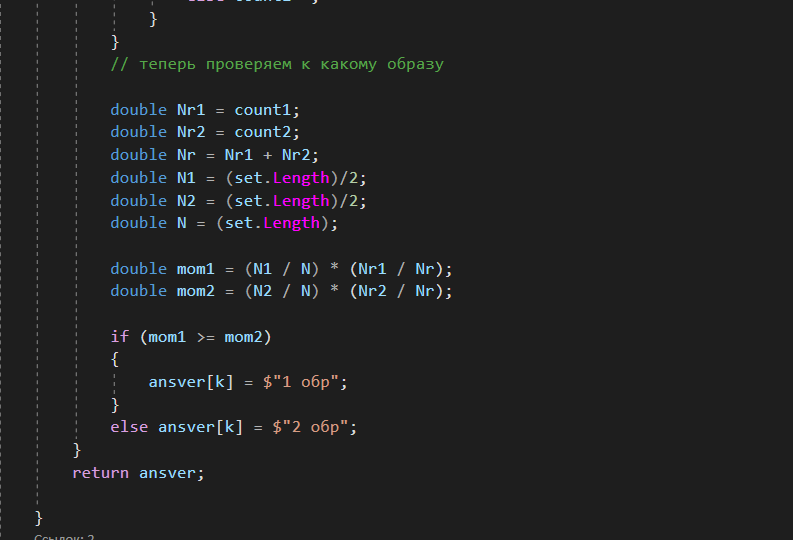


Рис. 12. Метод для нахождения ответа о принадлежности материала экзамена к образам

В целях практической проверки работы программы зададим для нее параметры, которые ставились нами ранее в предметной задаче. Как мы можем заметить программа дала идентичный результат, что и наше решение ранее в *Excel* (см. рис. 13).

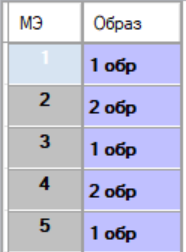


Рис. 13. Результат распознавания

Анализируя полученный результат, можно сказать, что объектов 1-ого образа - 3, объектов 2-ого образа - 2. Из решения задачи распознавания принадлежности материалов экзамена, а именно опрашиваемых людей, к двум категориям (убывших из Российской Федерации и прибывших) стоит отметить, что людей, убывающих из Российской Федерации больше.

Проведем с помощью генерации еще несколько опытов для заключения рамок для образов. Возьмем границы заработной платы для материалов обучения 50 000 и 70 000 для 1 образа, и 70 000 и 100 000 для 2 образа. Количество сфер сделаем 3. А материал экзамена возьмем для определенного возраста от 20 до 30 лет и заработной платой от 70 000 до 100 000 (см. рис. 14).

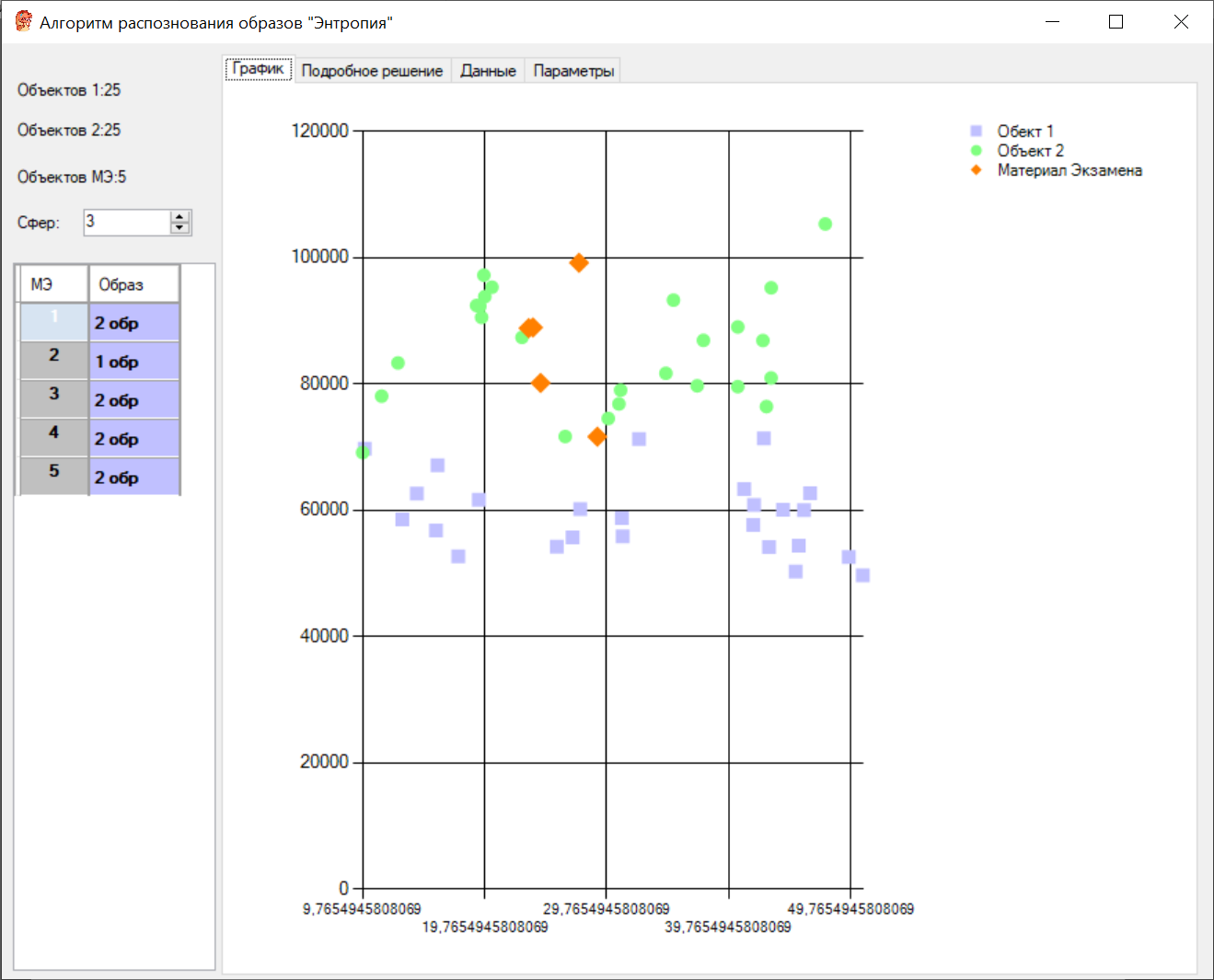


Рис. 14. Дополнительные вычисления для определения рамок

Из чего можно сделать вывод, что материалов второго образа больше, а значит люди с более высоким достатком склонны покидать страну. Чем больше заработная плата, а возраст работоспособный тем больше вероятность, что это 2 образ, а, следовательно, способен мигрировать из Российской Федерации.

Вывод

В данной курсовой работе бы рассмотрен метод распознавания образов как алгоритм «Энтропия». Были изучены: основные понятия, цели и задачи данного метода. Проведена классификация образов при наличии двух исходных образцов материала обучения. Также, была разработана программа реализующая алгоритм распознавания образов «Энтропия».

В итоге были получены результаты по материалам экзамена: объектов 1-ого образа - 3, объектов 2-ого образа - 2. Из решения задачи распознавания принадлежности категории людей, прибывающим и убывающим из Российской Федерации стоит отметить, что людей, убывающих из Российской Федерации больше. Следовательно, можно предположить, что в стране происходит убыль населения.

Список литературы

1. Вапник В. Н., Червоненкис А. Я. Теория распознавания образов. – М.: Наука,

1974. – 416 с.

1. Горелик А.Л., Скрипкин В. А. Методы распознавания. – 4-е изд. – М.: Высшая школа, 1984, 2004. – 262 с.
2. Федеральная служба государственной статистики [Электронный ресурс]: Режим доступа: <https://showdata.gks.ru/report/278934/> Дата обращения: 30.04.2022г.
3. Чабан Л.Н. Теория и алгоритмы распознавания образов. Учебное пособие. М.:

МИИГАиК. 2004. – 70 с.