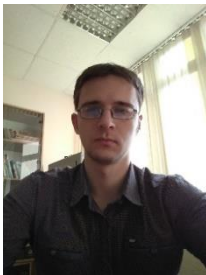


УДК 519.688

## РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА МУРАВЬИНОЙ КОЛОНИИ



**В.М. Басинский**  
Аспирант ГрГУ,  
преподаватель ГрГУ

*Кафедра математического и информационного обеспечения экономических систем факультета экономики и управления Гродненского государственного университета им. Я. Купалы, Республика Беларусь.*

*Гродненский государственный университет им. Я. Купалы, Республика Беларусь.*

*E-mail: vadim.basinskiy@gmail.com.*

### **В. М. Басинский**

*Окончил Гродненский государственный университет им. Я. Купалы. Аспирант ГрГУ. Работает ГрГУ в должности преподавателя. Проводит научные исследования в области Big Data и Data Mining и исследования возможностей применения метаэвристических алгоритмов для решения задач в этих областях.*

**Аннотация.** Типичными примерами современных приложений задачи классификации являются задачи интеллектуального управления бизнесом (англ. business intelligence), анализ социальных сетей, прогнозирование спроса и др. Проблема классификации представляет собой задачу упорядочения объектов по некоторым классификационным признакам и отнесения каждого объекта к какому-то заранее известному классу. Исследование алгоритмов решения задачи классификации является одним из приоритетных направлений. В данной работе описана разработка приложения, решающего задачу классификации при помощи генетического алгоритма и алгоритма муравьиной колонии.

**Ключевые слова:** задача классификации, генетический алгоритм, алгоритм муравьиной колонии.

### **Введение.**

Для использования алгоритма муравьиной колонии для решения задачи классификации выполним для не дискретных атрибутов процедуру дискретизации для уменьшения возможного количества их значений. После этого исходные данные можно представить в виде путей на полносвязном графе, вершинами которого являются пары «attribute = value». Каждый такой путь проходит только через одну вершину, из соответствующих одному атрибуту, а для записей используемых для обучения, валидации и тестирования каждый путь заканчивается в вершине атрибута, соответствующего классу, который требуется предсказать.

### **Материалы и методы.**

Построение и поиск классификационных правил на графе будет осуществлять алгоритм муравьиной колонии [1]. Алгоритм муравьиной колонии имеет набор гиперпараметров, значения которых необходимо подбирать под каждый набор индивидуально. К гиперпараметрам относятся следующие переменные:

- количество муравьев в поколении;
- количество записей для сходимости;
- максимальное количество пропущенных записей;
- минимальное количество записей, подходящих под правило;

- количество групп для разделения при дискретизации атрибутов (выбирается для каждого атрибута отдельно);
- метод расчета эвристической функции;
- метод разбиения данных для обучения алгоритма.

Задачу поиска оптимального набора гиперпараметров решает генетический алгоритм. В качестве фитнес функции для оценки качества модели используется точность классификации на тестовом наборе данных. Общий процесс работы алгоритма представлен на рисунке 1.

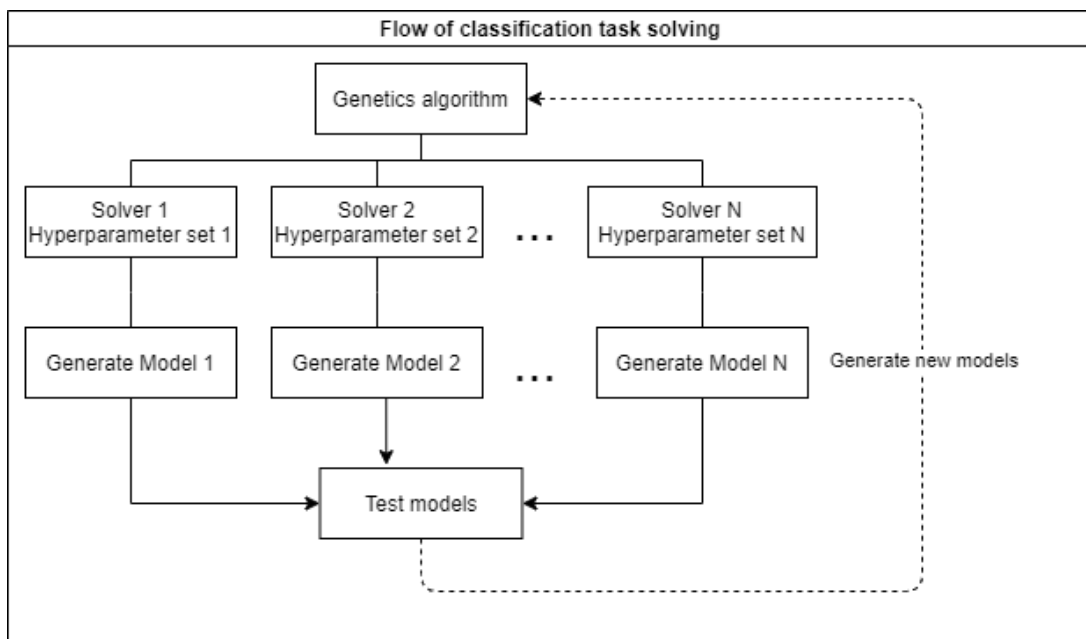


Рисунок 1. Диаграмма деятельности для алгоритма решения задачи классификации

Каждая модель представляет собой набор классификационных правил, построенных алгоритмом муравьиной колонии. Классификационное правило – набор выражений вида «attribute = value» для описания построенного пути с обязательным значением class – одно из множества значений переменной, которую модель предсказывает. Для каждого правила класс, предсказываемый им, определяет набор записей, которые оно покрывает [2].

Теоретическое описание процессов построения правил и некоторых используемых модификаций с уточнением содержательного смысла параметров и гиперпараметров рассматриваемых алгоритмов приведено в [3]. Рассмотрим основные структурные элементы разработанного приложения.

В рамках тестирования приложения использовались наборы, имеющиеся в открытом доступе. Наборы данных доступны в формате .arff, который позволяет получать данные в строго описанном формате и избежать возникновения ошибок при чтении данных.

Основные алгоритмы построения классификационных правил реализует класс Ant Класс Solver обеспечивает взаимодействие между муравьями одной колонии при построении классификационной модели. Параметры каждой колонии определяет класс GeneticOptimizer, который является начальным этапом в решении задачи классификации. Структуру и взаимодействие этих классов можно видеть на рисунке 2.

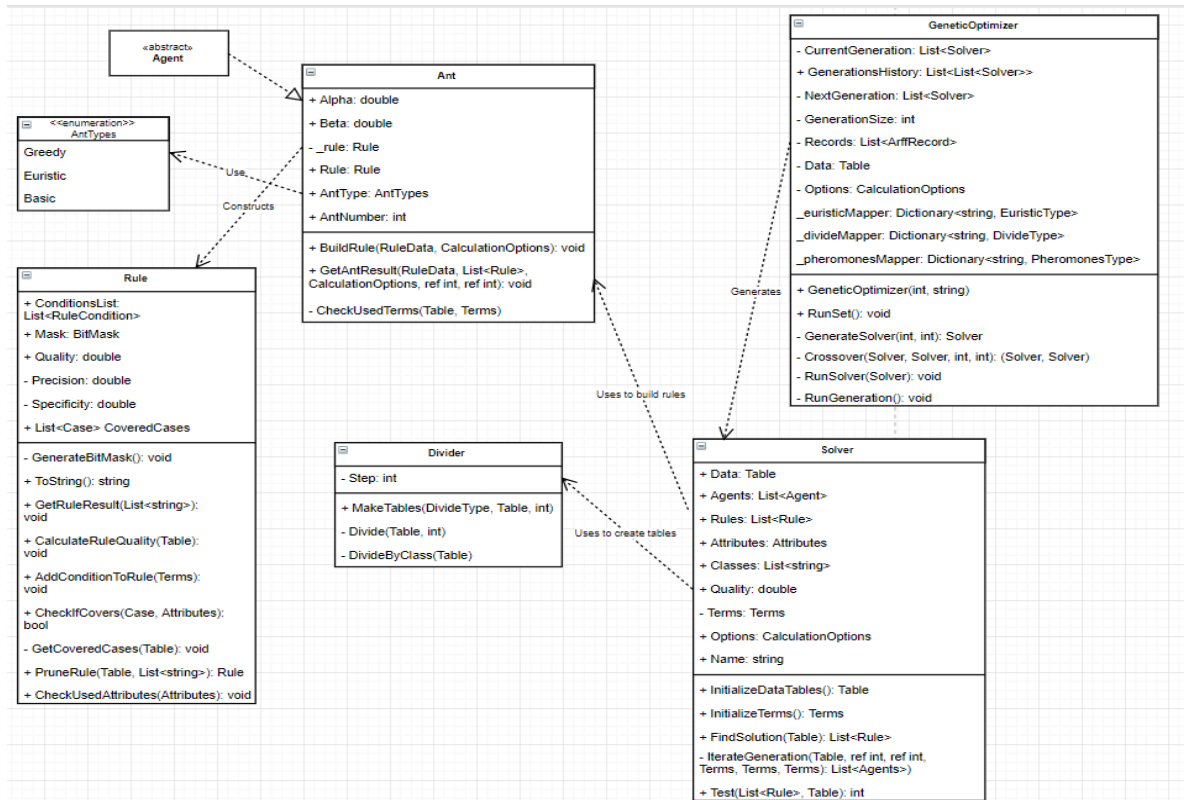


Рисунок 2. Классы, реализующие построение классификационных моделей

Для тестирования алгоритмов были использованы наборы доступные в открытом доступе. Для работы с файлами данных в формате arff использовались классы, показанные на рисунке 3.

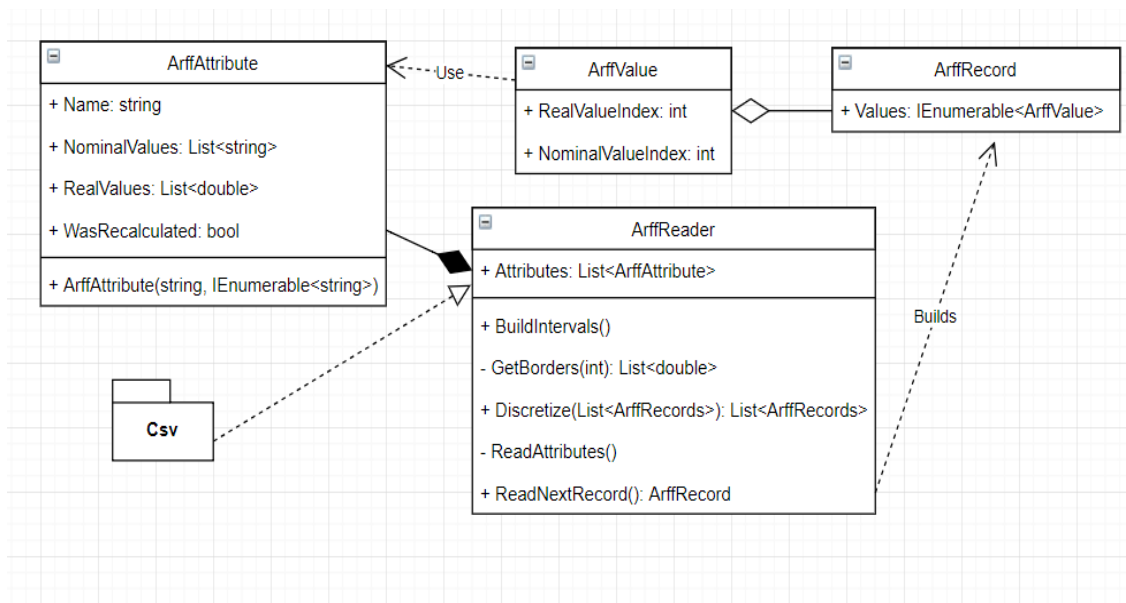


Рисунок 3. Классы, реализующие чтение данных

Для хранения и представления данных внутри приложения были разработаны классы данных, структура и взаимодействие которых отображены на рисунке 4.

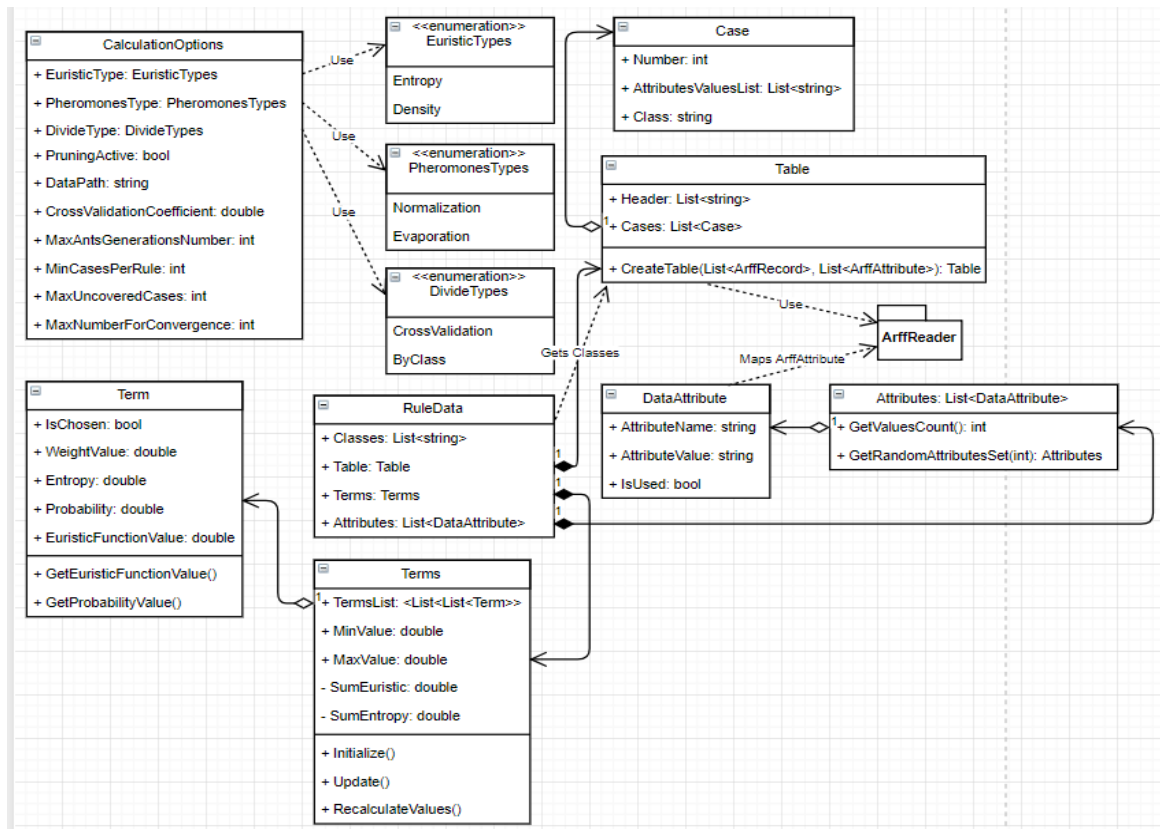


Рисунок 4. Классы, реализующие хранение, преобразование и представление данных

### Результаты.

Рассмотрим результаты тестирования реализованного приложения на наборе данных vote [4]. Для проверки влияния генетического алгоритма на качество классификационных моделей запускалось обучение на различное количество поколений: 1, 5, 100 и 100. Начальные значения всех переменных отличались для различных запусков, но совпадало для различных моделей в рамках одного запуска. Результаты тестирования приведены на рисунке 5.

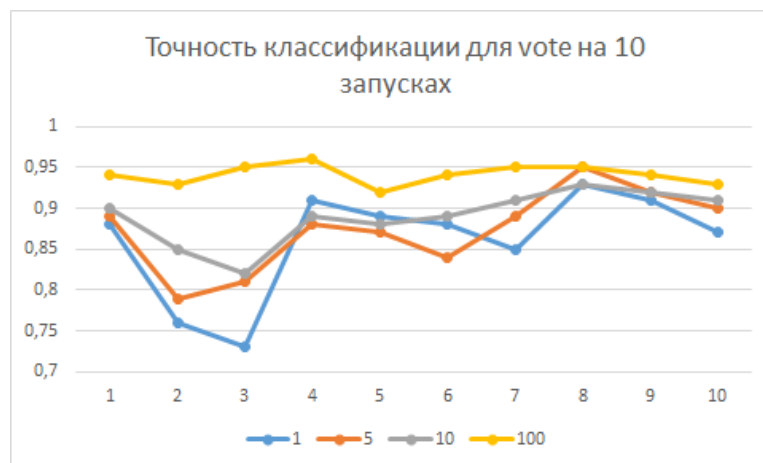


Рисунок 5. Результаты тестирования приложения с различным числом итераций алгоритмов

### Заключение.

Реализованные алгоритмы позволяют получать модели для задачи классификации, которые по своему качеству не уступают результатам большинству широко используемых методов решения задачи классификации, разработанных ранее, а на некоторых наборах данных могут их превосходить.

### **Список литературы**

- [1] Ants Constructing Rule-Based Classifiers / Martens D. [and others] // Swarm Intelligence in Data Mining. – 2006. – Vol. 34. – P. 21-43.
- [2] Freitas, A.A. Ant Colony Algorithms for Data Classification / A.A. Freitas, R.S. Parpinelli, H.S. Lopes // Encyclopedia of Information Science and Technology. – Second Edition – Information Resources Management Association. – USA, 2008. – P. 154–159.
- [3] Басинский, В.М. Программная реализация одной модификации метаэвристического алгоритма для задачи классификации / В.М. Басинский, Ю.Г. Степин // BIG DATA and Advanced Analytics. BIG DATA и анализ высокого уровня: материалы IV Междунар. науч.-практ. конф., Минск, 3–4 мая 2018 г. – Минск, 2018. – С. 321–325.
- [4] OpenML Home Machine Learning, better together [Electronic resource] – 2021. – Mode of access: <https://www.openml.org>.

## **DEVELOPMENT OF AN APPLICATION FOR SOLVING THE CLASSIFICATION PROBLEM USING AN ANT COLONY ALGORITHM**

**V.M. BASINSKII**

*Postgraduate student of the GrSU,  
lecturer of GrSU*

*Department of Mathematical and Information Support of Economic Systems, Faculty of Economics and Management, Y. Kupala Grodno State University, Republic of Belarus  
Y. Kupala Grodno State University, Republic of Belarus  
E-mail: vadim.basinskii@gmail.com*

**Abstract.** Typical examples of modern applications of the classification problem are business intelligence, social network analysis, demand forecasting, etc. The classification problem is the problem of ordering objects according to some classification criteria and assigning each object to some previously known class. The study of algorithms for solving the classification problem is one of the priority areas. This paper describes the development of an application that solves the classification problem using a genetic algorithm and an ant colony algorithm.

**Keywords:** classification problem, genetic algorithm, ant colony algorithm