# ВВЕДЕНИЕ

Классификация заключается в прогнозировании значения категориального атрибута (класса) на основе значений других атрибутов. Цель классификации – взять входной вектор *х* и назначить его к одному из *K* дискретных классов *Сk*, где *k* = 1, ..., *K*. Каждый экземпляр данных определяется в соответствии с набором атрибутов или переменных.

Для решения задачи классификации выполняется определенный вид предварительной обработки, известной как преобразование представления, которое, учитывая исходный вектор признаков *F0* и обучающее множество *L*, состоит в создании представления *F*, полученного из *F0*, которое максимизирует некоторый критерий, и, по крайней мере, так же хорошо, как *F0* по отношению к этому критерию.

Разработка хорошего пространства признаков является предпосылкой для достижения высокой производительности в любой задаче машинного обучения. Однако часто неясно, каким должно быть оптимальное представление признака. В результате общий подход заключается в том, чтобы использовать все доступные атрибуты в качестве признаков и оставить проблему идентификации полезных наборов признаков модели обучения. Такой упрощенный подход не всегда работает хорошо. С развитием технологий аппаратного и программного обеспечения и увеличением объема данных количество признаков, используемых системами машинного обучения, измеряется десятками тысяч и миллионами признаков.

Для решения проблем нерелевантности признаков и взаимодействия признаков используются методы построения признаков. Конструирование признака включает в себя преобразование заданного набора входных признаков для создания нового набора более мощных признаков, которые затем используются для прогнозирования. Поскольку вновь созданные признаки учитывают взаимодействия в предыдущем пространстве признаков, они более значимы и приводят к более кратким и точным классификаторам.

В данной работе применяется генетическое программирование для автоматического конструирования признаков с целью повышения различающей эффективности классификатора

Актуальность выбранной темы обосновывается тем, что проблема классификации не может быть правильно решена, если важные взаимодействия и отношения между оригинальными признаками, не принимаются во внимание. Таким образом, многие исследователи согласились, что выделение признаков является наиболее важным ключом к любому распознаванию образов и проблеме классификации: «Точный выбор признаков, пожалуй, самая сложная задача распознавания образов» (Micheli-Tzanakou, 2000), «идеальная функция извлечения даст представление, что сделает работу классификатора тривиальной» (Duda, Hart, & Stork, 2001). В большинстве случаев выделение признаков осуществляется человеком на основе знаний исследователя, опыта и / или интуиции.

Цель – нахождение и отработка методики построения признаков для решения задачи классификации.

Для достижения поставленной цели необходимо выполнить следующие задачи:

* изучить соответствующую литературу;
* разработать алгоритм построения признаков;
* реализовать программу по данному алгоритму;
* оценить эффективность работы программы;
* обобщить полученные результаты и сделать соответствующие выводы.

Объектом исследования выступает классификация изображений. Предметом исследования является алгоритм построения признаков для классификации.

Методы исследования: анализ и классификация литературных источников, моделирование системы, системный анализ, объектно-ориентированное проектирование и программирование.

Основные результаты:

* разработан метод построения признаков для классификации;
* оценен результат применения метода генетического программирования для построения признаков.

Практическая значимость. Результаты данной работы могут быть применены в медицинских учреждениях для распознавания опухолей.

Структура работы. Диссертационная работа включает введение, три главы, заключение, список использованных источников, заключение и 3 приложения.