

УДК004.048:004.622

## **ПОЛУЧЕНИЕ ПРАВИЛ ВЫВОДА НА ОСНОВЕ ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ**

*Т.А. Васяева*

Донецкий национальный технический университет

Современным направлением диагностики и прогнозирования являются методы, основанные на извлечении знаний. Один из подходов формирования знаний заключается в разработке программ, способных обучаться под руководством эксперта-учителя. Наиболее впечатляющим примером обучаемой системы является организм человека или животного, который эволюционировал вместе с окружающим миром. Этот подход к обучению, основанный на адаптации, отражен в генетических алгоритмах и генетическом программировании.

Целью работы является разработка метода прогнозирования с явной формализацией правил вывода продукционной системы.

Для решения поставленной задачи предложено использовать генетическое программирование.

При этом необходимо выполнить предварительные этапы:

- определить терминальное множество;
- определить функциональное множество;
- определить фитнес-функцию;
- определить значения параметров, такие как мощность популяции, максимальный размер особи, вероятности кроссинговера и мутации, способ отбора родителей, критерий окончания эволюции и т.п.

После этого можно разрабатывать непосредственно сам эволюционный алгоритм, реализующий ГП для конкретной задачи.

Решение задачи на основе ГП реализуется следующей последовательностью действий.

1. Установка параметров ГП.

2. Генерация начальной популяции. Популяция представляет собой набор хромосом. Каждая хромосома соответствует определенному дереву, представляющее собой решение. Дерево (хромосома), на начальном этапе генерируется случайным образом и состоит из функциональных и терминальных узлов.
3. Оцениваются значения фитнес-функции особей в популяции.
4. Применение генетических операторов.
5. Проверка критерия останова. При его выполнении переход на шаг 6, иначе шаг 3.
6. Выбор лучшего решения в последней популяции.

Рассмотрим реализацию метода на примере прогнозирования синдрома внезапной смерти грудных детей.

Необходимо выполнить задачу классификации: определить попадает ли пациент в группу риска синдрома внезапной смерти грудных детей (СВСГР) или нет.

В качестве обучающего множества используются реальные данные обследования 240 пациентов, (120 детей, которые умерли в Донецкой области от СВСГД, и контрольная группа из 120 живых детей на первом году жизни). Данные составляют информацию общего характера и образа жизни беременных, перенесенные заболевания и результаты некоторых анализов.

Предлагается модификация способа кодирования особей для генетического программирования. Каждая особь представляет собой дерево, которое соответствует синтаксическому выражению, представляющее множество правил в дизъюнктивной нормальной форме.

На рисунке 1 представлен пример дерева в дизъюнктивной нормальной форме. Дерево представлено двумя правилами. Данное представление особи значительно упрощает интерпретацию результата. В рассмотренном примере расшифровка будет следующей:

ЕСЛИ правило 1 ИЛИ правило 2 ТО результат 1, ИНАЧЕ результат 2.

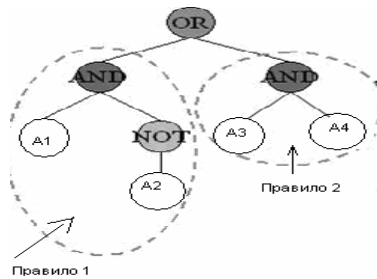


Рисунок 1.- Пример дерева в ДНФ.

Терминальное множество: входное обучающее множество преобразовано в булевы переменные следующим образом:

- место жительства (город – 1, село – 0)
- возраст матери на момент родов (полных лет)  $\leq 17$
- возраст матери на момент родов (полных лет)  $> 31$
- и др.

Наличие каждого фактора принято за единицу, отсутствие за ноль.

Функциональное множество состоит из логических операций: AND, OR, NOT.

В качестве фитнес-функции рассматривается доля пациентов с правильно поставленным диагнозом.

Для реализации поставленной задачи написана программа в среде C++ Builder 6, которая выполняет рассмотренный выше алгоритм.

Разработанный аппарат ГП создан и протестирован на примере прогнозирования СВСГР, но может быть использован и при решении других задач диагностики и прогнозирования.

Литература:

1. W. Banzhaf et all. Genetic Programming – an Introduction. – Morgan Kaufman, Heidelberg:San-Francisco, 1998.
2. Рутковская Д., Пилинский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И.Д. Рудинского. - М.: Горячая линия – Телеком, 2006. – 452 с.