

## ПРИМЕНЕНИЕ КОЛЛЕКТИВОВ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ КЛАССИФИКАЦИИ

Хритonenко Д.И.

Научный руководитель д-р техн. наук Семенкин Е.С.  
Сибирский государственный аэрокосмический университет  
имени академика М. Ф. Решетнева

Искусственные нейронные сети (ИНС) на сегодняшний день получили довольно широкое распространение при решении различных задач анализа данных. Для их корректной работы выбирается структура, а также производится процедура обучения. После чего, полученная ИНС, как правило, используется для решения поставленной задачи. Эффективность такого решения можно повысить посредством одновременного использования нескольких нейронных сетей. В таком случае встает вопрос о выборе используемых ИНС из общего, заранее сгенерированного, множества, а также о методе учета «мнений» каждого из членов полученного коллектива нейронных сетей.

В качестве методов, решающих перечисленные проблемы предлагается использование самоконфигурируемых эволюционных алгоритмов, а именно алгоритма генетического программирования (ГП) и генетического алгоритма (ГА). Эти алгоритмы способны в автоматическом режиме проектировать и обучать ИНС и их коллективы, а самоконфигурирование позволит избежать дополнительных настроек алгоритмов.

Функциональное множество алгоритма ГП, используемого для формирования коллектива будет включать в себя различные математические операции и функции (умножение, синус и т. д.). Терминальное – все предварительно отобранные ИНС, а также ряд параметров, настраиваемых в ходе эволюционного поиска алгоритмом. Таким образом, полученный алгоритм будет являться некоторым обобщением существующих стандартных методов учета мнения коллектива (среднее, взвешенное среднее и т.д.). За счет ввода в терминальное множество настраиваемых параметров алгоритм не может получить решение хуже, чем лучший член коллектива ИНС. Работу описанного алгоритма можно представить в виде блок-схемы:

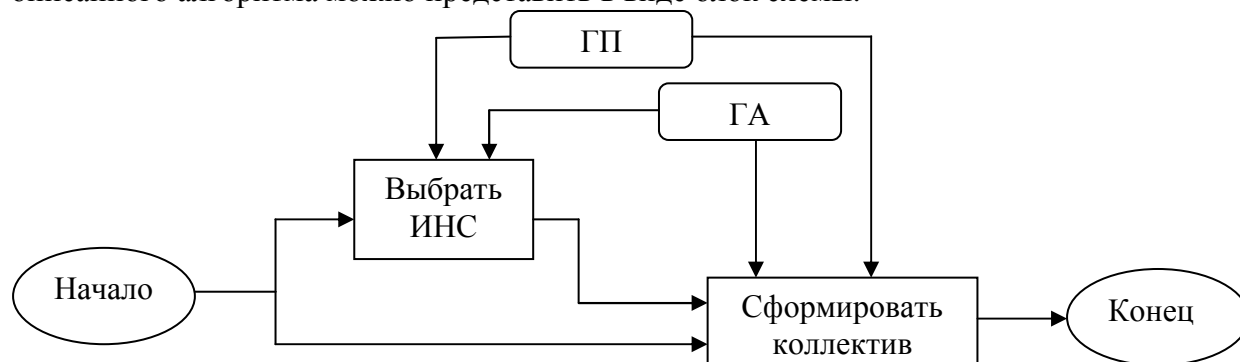


Рисунок 1 – Блок-схема алгоритма

Представленный алгоритм был реализован в виде программной системы и протестирован на следующих задачах с репозитория машинного обучения UCI:

Таблица 2 – Тестовые задачи

Задача	Объем выборки	Число атрибутов	Число классов
Ирисы	150	4	3
Вина	178	13	3
Кредиты (Австралия)	690	14	2
Кредиты (Германия)	1000	20	2

Для тестирования все выборки разбивались на тестовую и обучающую случайным образом. Вероятность попадания в тестовую выборку была равна 0,3. Усреднение проводилось по 50-ти прогонам. Ниже представлены результаты тестирования на данных задачах.

Таблица 3 – результаты тестирования

Алгоритм	Ирисы	Вина	Кредиты (Австралия)	Кредиты (Германия)
SC_GP+ANN	6,3%	3,9%	11,3%	23%
SC_GP+ANN+E	2,5%	2,4%	9,9%	20%

В таблице 2 представлена относительная ошибка классификации на тестовой выборке. В данной таблице SC\_GP+ANN означает, что использовалась ИНС сгенерированная самоконфигурируемым алгоритмом генетического программирования. Алгоритм SC\_GP+ANN+E использовал коллектив ИНС.

Из представленных выше результатов видно, что использование коллективов искусственных нейронных сетей позволяет снизить ошибку классификации. Однако использование такого подхода приводит к существенному увеличению временных затрат.

Ниже также приведено сравнение представленного алгоритма классификации с существующими аналогами известными из научной литературы:

Таблица 4 – Результаты тестирования

Алгоритм	Кредиты (Австралия)	Кредиты (Германия)	Алгоритм	Кредиты (Австралия)	Кредиты (Германия)
SCGP	0.9022	0.7950	Boosting	0,7600	0,7000
MGP	0.8985	0.7875	Bagging	0,8470	0,6840
2SGP	0.9027	0.8015	RSM	0,8520	0,6770
GP	0.8889	0.7834	CCEL	0,8660	0,7460
Fuzzy classifier	0,8910	0,7940	CART	0,8744	0,7565
C4.5	0.8986	0.7773	MLP	0,8986	0,7618
LR	0.8696	0.7837	SC_GP+ANN	<b>0,8874</b>	<b>0,7697</b>
k-NN	0.7150	0.7151	SC_GP+ANN+E	<b>0,9011</b>	<b>0,8004</b>

Значение критерия, указанного в таблице выше вычислялось следующим образом:

$$I = 1 - \frac{ERROR}{n}$$

В данной формуле *ERROR* – число неверно классифицированных объектов, *n* – общее число объектов.

Из приведенных результатов видно, что усредненные показатели эффективности алгоритма формирования коллективов ИНС (SC\_GP+ANN+E) выигрывают у большинства из представленных в таблице 3. Все это показывает эффективность данного метода классификации. Стоит отметить, что описанный подход позволяет использовать внутри алгоритма ГП любые из существующих видов интеллектуальных информационных технологий. Это может способствовать дальнейшему увеличению эффективности генерируемых классификаторов.

Таким образом, был разработан и программно реализован подход последовательного применения самоконфигурируемого алгоритма генетического программирования, который позволяет формировать эффективные коллективы искусственных нейронных сетей для решения задач классификации.