УДК 519.68

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ АЛГОРИТМА ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ С НАСТРОЙКОЙ КОЭФФИЦИЕНТОВ В ЗАДАЧЕ СИМВОЛЬНОЙ РЕГРЕССИИ

Становов В.В.

Научный руководитель проф. СеменкинЕ.С.

Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева

Рассматривается эффективность работы самонастраивающегося алгоритма генетического программирования для задачи символьной регрессии с настройкой коэффициентов и без. Разработана программная среда, реализующая алгоритм. Получены результаты работы на тестовых и практических задачах.

Задача восстановления регрессионной зависимости может быть решена различными способами, в том числе с использованием метода генетического программирования для символьной регрессии. В процессе генерации выражений в них могут появляться некие константы – действительные числа из заданного пользователем интервала, от которых зависит результат вычисления. Настройка этих констант производилась при помощи генетического алгоритма.

Для исследования эффективности настройки коэффициентов была разработана программная среда, реализующая алгоритм. В программе имеется возможность выбора трех типов селекции: пропорциональная, ранговая и турнирная с настраиваемым размером турнира; два типа скрещивания: стандартное и одноточечное; два типа мутации: точечная и выращиванием поддерева с настраиваемой вероятностью мутации. Былииспользованыдваметодасамонастройки: Population-LevelDynamicProbabilities (PDP) иIndividual-LevelDynamicProbabilities (IDP), описанные в [1].

Генетический алгоритм, использованный для настройки коэффициентов, применялся для всех без исключения индивидов в популяции. В нем присутствует один тип селекции – турнирная, 3 типа скрещивания (одноточечное, двухточечное и равномерное)и 3 типа мутации (слабая, средняя и сильная).

В качестве тестовых функций были взяты функции двух переменных с различными параметрическими структурами, приведенные в [2]. Исследование проводилось по следующей схеме: тестовые выборкиобъемом 400 были сгенерированы для каждой функции, в каждом независимом прогоне алгоритма устанавливалось число индивидов равным 100 и число поколений равным 1000 (для алгоритма без настройки) либо 100 (для алгоритма с настройкой), среднеквадратичная ошибка и длина выражения (без учета скобок) усреднялись по 20 прогонам. Для генетического алгоритма, настраивающего параметры устанавливалось число индивидов – 10, число поколений – 10, одноточечное скрещивание, слабая мутация. Результаты приведены в таблице ниже:

	ср. ошибка	ср. длина	дисп. ошибки	дисп. длины
PDP+настр.	0,0472259	24,4729	3,88E-07	3,75099
IDP+настр.	0,00498988	25,3833	4,00E-07	4,09779
PDP без настр.	0,004412822	35,7962	3,79E-07	12,19943
IDP без настр.	0,004176	37,73083	3,99E-07	11,32794
Обычный ГП	0,005307	34,11875	3,48E-07	11,40988

По результатам, приведенным в таблице можно заключить, что, хотя использование настройки коэффициентов приводит к увеличению средней ошибки, длина выражения в этом случае оказывается в среднем в 1.5 раза меньше. Также стоит отметить значительно меньшую величину дисперсии длины выражения.

Кроме того, при помощи описанного алгоритма были решены задачи банковского скоринга по базам данных Australia-1 и Germany-1, приведенным в [3] и [4].Данные задачи сводятся к задачам классификации, которые решались путем построения разделяющей поверхности в виде символьного выражения. Результаты работы алгоритма представлены ниже (представленный алгоритм обозначен GP_RPN):

SCGP	0,9022	0,795
MGP	0,8985	0,7875
2SGP	**0,9027**	**0,8015**
GP	0,8889	0,7834
Fuzzyclassifier	0,891	0,794
C4.5	0,8986	0,7773
LR	0,8696	0,7837
Bayesianapproach	0,847	0,679
Boosting	0,76	0,7
Bagging	0,847	0,684
RSM	0,852	0,677
CCEL	0,866	0,746
k-NN	0,715	0,7151
CART	0,8744	0,7565
MLP	0,8986	0,7618
GP_RPN	0,896	0,755

Анализ результатов показывает работоспособность предложенных методов не только на тестовых функциях, но и при решении реальных задач классификации и построения регрессии.

Библиографическиессылки

1. Niehaus, J.,Banzhaf, W.Adaption of Operator Probabilities in Genetic Programming. In: Miller J. et al. (Eds.): EuroGP 2001, LNCS 2038, pp. 325-336, 2001.
2. URL:http://coco.gforge.inria.fr/lib/exe/fetch.php?media=download3.6:bbobdocexperiment.pdf
3. URL: http://archive.ics.uci.edu/ml/datasets/Statlog+%28Australian+Credit+Approval%29
4. URL: http://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29