

РАСПАРАЛЛЕЛИВАНИЕ САМОКОНФИГУРИРУЕМОГО АЛГОРИТМА ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ В ЗАДАЧАХ ПРОЕКТИРОВАНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

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) – разновидность интеллектуальных информационных технологий, используемая для решения ряда задач: прогнозирования, кластеризации, классификации и т.д. Использование ИНС позволяет достигать высоких результатов в решении таких задач. Однако существует также ряд проблем, связанных с ИНС – проблема выбора конфигурации нейронной сети, проблема обучения. Эволюционные алгоритмы (ЭА), а в частности алгоритм генетического программирования, позволяют при небольшой своей модификации решать эти проблемы.

На сегодняшний день широкое распространение имеют многоядерные вычислительные системы. Учитывая большую вычислительную сложность ЭА, используемых в задачах проектирования ИНС, полезной модификацией будет распараллеливание вычислений ЭА. Учитывая факт зависимости эффективности ЭА от его конфигурации (выбора тех или иных генетических операторов), также полезной модификацией является самоконфигурирование. Эффективность самоконфигурируемых ЭА была доказана ранее на различных задачах.

Как показали исследования, вычислительные системы в состоянии генерировать лишь псевдослучайные последовательности чисел. Кроме того, в случае генерирования последовательностей псевдослучайных чисел различными потоками, они зачастую оказываются либо одинаковыми, либо сдвинутыми на некоторое число элементов. Вышесказанное важно учитывать, т.к. эффективность стохастических алгоритмов (к которым относятся и ЭА) зависит от работы генератора (ГСЧ). Ниже представлена общая схема алгоритма генетического программирования (ГП), а также предлагаемый метод параллелизации:

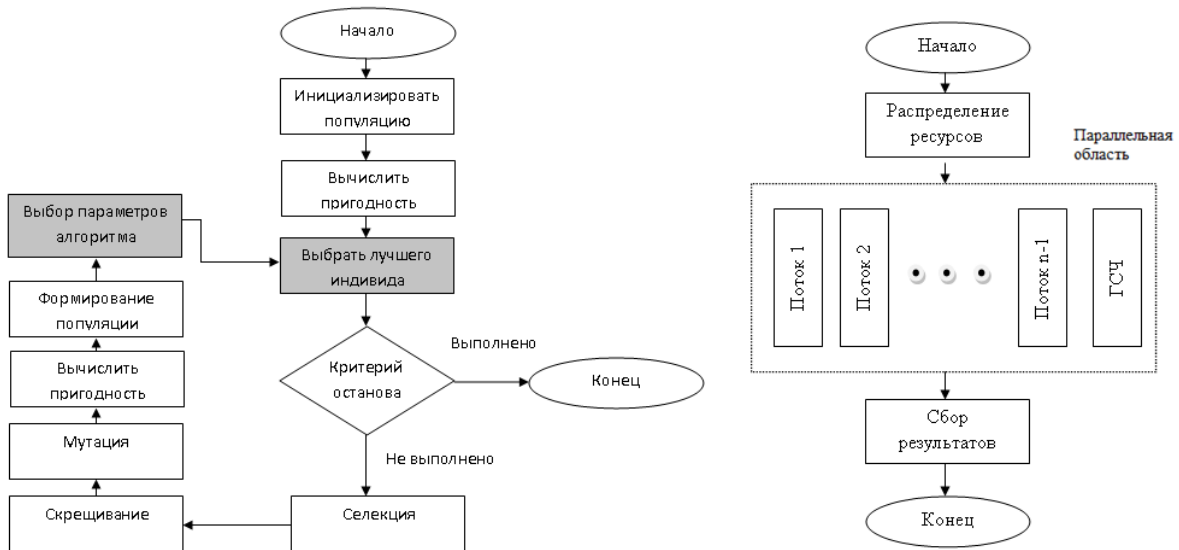


Рисунок 1. Параллельный самоконфигурируемый ГП

Затемненные блоки представляют собой части алгоритма, не подлежащие параллелизации. В алгоритме каждый из представленных потоков будет обрабатывать лишь свою часть популяции. Проблема с ГСЧ была решена посредством введения потока-мастера, отвечающего за генерацию псевдослучайных чисел. Проведенные ранее исследования говорят об эффективности такого подхода.

Самоконфигурируемый алгоритм ГП был модифицирован для проектирования ИНС и реализован в виде программной системы (ANN_GP). Эффективность предложенной системы была проверена на следующих задачах классификации:

- *Australian credit* – задача о выявлении подозрительных операций с банковскими картами. Данная задача имеет 14 признаков и 2 класса;
- *German credit* – задача о выдаче банковского кредита. Данная задача имеет 24 признака и 2 класса. Критерий эффективности алгоритма:

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

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

Для проверки эффективности алгоритма было проведено десять стократных прогонов. По критерию Манна-Уитни-Уилкоксона с доверительной вероятностью 0.95 алгоритм ANN_GP является неразличимым со своей распараллеленной версией. Ниже представлены усредненные результаты работы алгоритма:

Таблица 1 – Результаты тестирования

Название алгоритма	Australian credit	German credit	Название алгоритма	Australian credit	German credit
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
C4.5	0.8986	0.7773	CART	0.8744	0.7565
LR	0.8696	0.7837	MLP	0.8986	0.7618
k-NN	0.7150	0.7151	ANN_GP	0.9257	0.8393

Из приведенных выше результатов видно, что предложенный подход решения задачи выигрывает у всех представленных аналогов, известных из научной литературы. Кроме того, важным является тот результат, что при неизменной эффективности время работы было уменьшено (приблизительно в 2 раза). Нелинейное увеличение быстродействия вычислительной системы объясняется сложностью её устройства. Зависимость времени работы от числа потоков представлена на графике ниже:

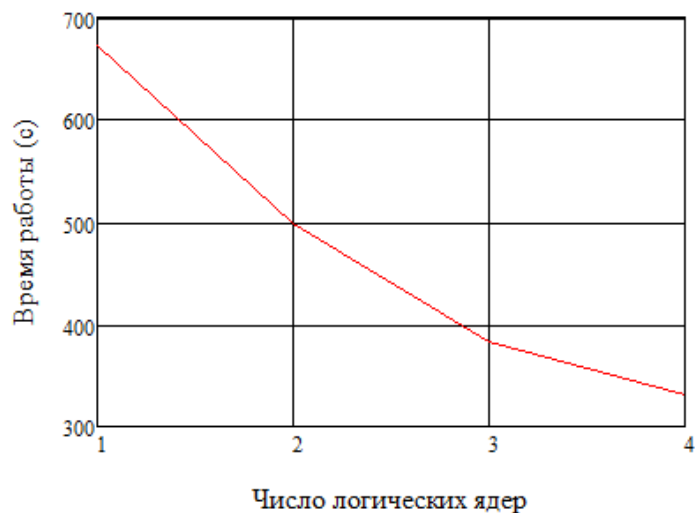


Рисунок 2. Время работы алгоритма

Таким образом, был разработан и программно реализован распараллеленный самоконфигурируемый алгоритм генетического программирования. Представленная схема распараллеливания позволяет без снижения эффективности увеличивать скорость работы программной системы, что является важным в случае с ресурсоемкими вычислениями.