

УДК 53.082.55, 519.688

АЛГОРИТМ ОБРАБОТКИ И АНАЛИЗА СПЕКТРОВ КОМБИНАЦИОННОГО РАССЕЯНИЯ С ПРИМЕНЕНИЕМ НЕЙРОННЫХ СЕТЕЙ

Е. В. Дьячков¹, М. А. Казарян², А. В. Обходский¹, Е. В. Обходская¹,
А. С. Попов¹, В. И. Сачков¹

В работе рассмотрено решение задачи обработки большого объема данных при анализе спектров комбинационного рассеяния газовой смеси. В основе алгоритма применяется искусственная нейронная сеть. Определены условия применения нейронных сетей при решении практических задач анализа спектров в реальном масштабе времени, в том числе для дистанционного поиска тяжелых углеводородов. Оценка быстродействия алгоритма проводилась с применением средств вычислительной техники с последовательной и параллельной обработкой данных.

Ключевые слова: спектр комбинационного рассеяния, газовая смесь, обработка данных, нейронная сеть, программное обеспечение, параллельные вычисления.

Введение. Эффективное решение задачи достоверного обнаружения месторождений углеводородов возможно путем одновременного применения нескольких взаимодополняющих методов поиска, в том числе метода дистанционного лазерного зондирования поверхностного слоя Земли, который является на сегодняшний день одним из перспективных методов, и пассивной сейсморазведки [1].

Одной из проблем применения метода дистанционного геохимического анализа месторождений углеводородов является необходимость обработки большого объема данных, получаемых в результате измерения спектров комбинационного рассеяния с помощью лидарных комплексов. От скорости обработки и достоверности получаемых данных во многом зависит эффективность дистанционного геохимического поиска. Задача

¹ Национальный исследовательский Томский государственный университет, 634050 Россия, Томск, Ленина пр-т, 36.

² ФИАН, 119991 Россия, Москва, Ленинский пр-т, 53; e-mail: kazar@sci.lebedev.ru.

анализа спектров комбинационного рассеяния с визуализацией выходных данных сложно поддается формализации, в связи с этим одним из исследуемых вопросов являлся анализ применимости искусственных нейронных сетей для обработки спектрометрической информации. Такой подход к решению проблемы поиска месторождений углеводородов с использованием метода лазерного дистанционного зондирования позволяет исключить человеческий фактор и автоматизировать процесс обработки данных.

Алгоритм обработки данных. Значительное ускорение в работе программного обеспечения возможно за счет распараллеливания алгоритмов нейронной сети. Так, вычисления в задачах анализа спектров комбинационного рассеяния могут выполняться быстрее с применением вычислительных мощностей нескольких процессоров.

Анализируемые спектры комбинационного рассеяния, регистрируемые с помощью лидара, представлялись в виде отдельных файлов, содержащих массивы значений интенсивностей в области антистоксовых частот с варьируемым шагом по частоте.

Алгоритм обработки спектральных данных был реализован в виде программного обеспечения в среде Nsight Eclipse Edition, работающей под управлением операционной системы Linux Ubuntu и поддерживающей разработку программ для программно-аппаратной архитектуры параллельных вычислений CUDA.

С учетом формата входных и выходных данных, для решения задачи анализа спектров комбинационного рассеяния использовалась нейронная сеть прямого распространения сигнала, входной слой которой содержит число нейронов, соответствующее количеству значений в анализируемых спектрах. Количество значений интенсивности во входных файлах обучаемой выборки и обрабатываемых спектров при этом является одинаковым. При проведении исследований не применялись рекуррентные нейронные сети, так как исследуемые наборы спектров не зависят друг от друга и предыдущее состояние сети не должно влиять на результат их распознавания.

Количество нейронов в выходном слое зависит от количества классифицируемых объектов. Значения нейронов выходного слоя лежали в диапазоне от 0 до 1. Значение, близкое к 1 на выходных нейронах, означает соответствие обрабатываемого спектра тому или иному спектру действующих месторождений, с которыми происходит сравнение. Опытным путём было установлено, что для достижения оптимальных результатов используемая нейронная сеть должна иметь один скрытый слой, количество нейронов в котором может меняться в зависимости от количества нейронов во входном слое.

При выборе алгоритма обучения нейронной сети учитывались ее топология, модель входных данных и возможный способ обучения [2]. Для обучения нейронной сети

применялся способ обучения с учителем, предполагающий подачу на вход сети выборки из обучающих спектров комбинационного рассеяния газовой смеси, полученных на действующих месторождениях. Поскольку при проведении исследований применялась нейронная сеть прямого распространения, подходящим решением для ее обучения является использование алгоритма многослойного персептрона – алгоритма обратного распространения ошибки [3].

Обучение нейронной сети осуществлялось путем подачи на ее входы серии обучающих примеров. Каждый образец проходил обработку внутри структуры нейронной сети, далее вычислялся выходной сигнал сети, который сравнивался с соответствующим значением целевого вектора, представляющего собой требуемый выход.

Для программной реализации искусственной нейронной сети и решения поставленной задачи использовалась встраиваемая программная библиотека FANN (Fast Artificial Neural Network), обеспечивающая широкий набор настроек и выбор топологий нейронных сетей [4].

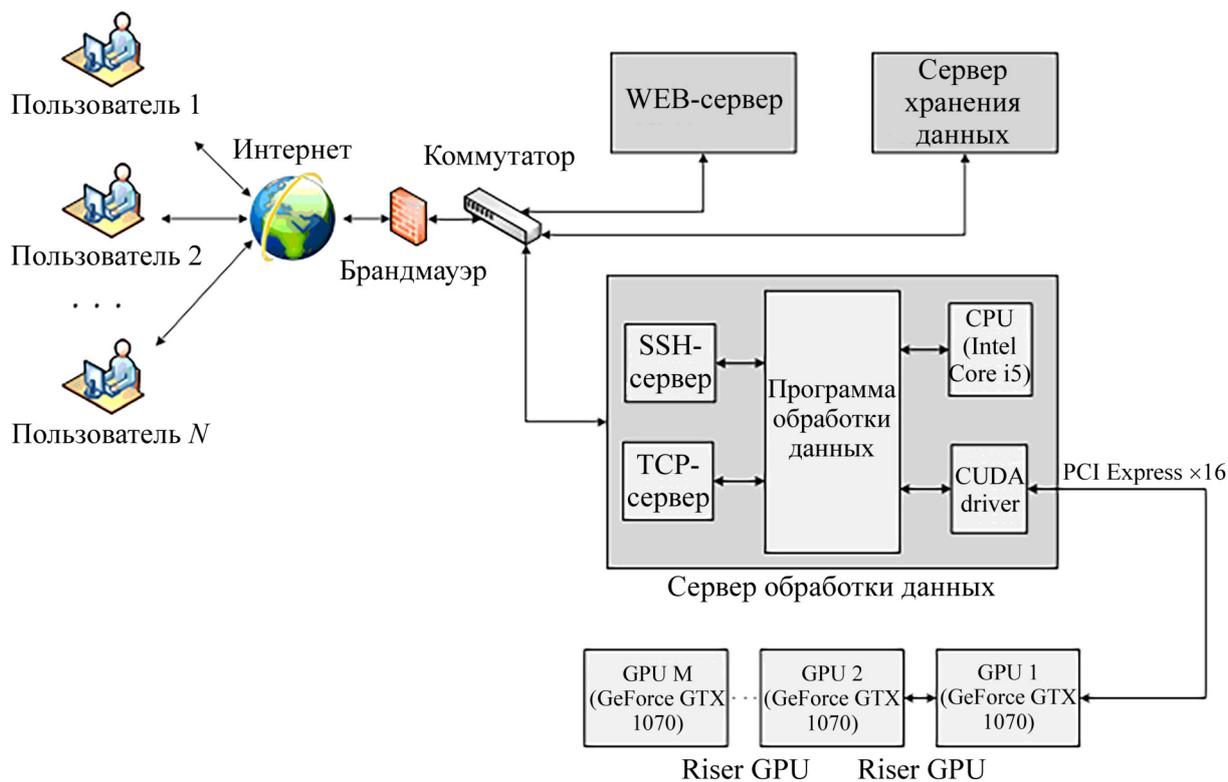


Рис. 1: Структура аппаратно-программного комплекса.

В результате проведения серии предварительных опытов по обучению нейронной сети в качестве функции активации был выбран гиперболический тангенс, применяемый для решения широкого круга задач и обеспечивающий, в данном случае, минимальное время и наилучшее качество обучения [5].

Экспериментальный стенд. В ходе проведения исследований образцовые и требующие обработки спектры сохранялись в базе данных аппаратно-программного комплекса (рис. 1) [6].

Загрузку входных файлов, содержащих образцовые и требующие обработки спектры, в программу обработки осуществляет пользователь посредством браузера своего ПК. Для этого пользователь подключается к WEB-серверу, на котором расположена соответствующая WEB-страница и встроенные PHP-скрипты TCP- и SSH-клиентов. Все файлы переносятся из рабочего каталога сайта на WEB-сервер в каталог программы на сервере обработки данных. Для соединения WEB-сервера с программой обработки используется потоковый сокет.

Массивы входных данных формировались двумя способами: зашумлением эталонных наборов спектров комбинационного рассеяния и “моделированием” спектров месторождений, непохожих ни на один из эталонных спектров.

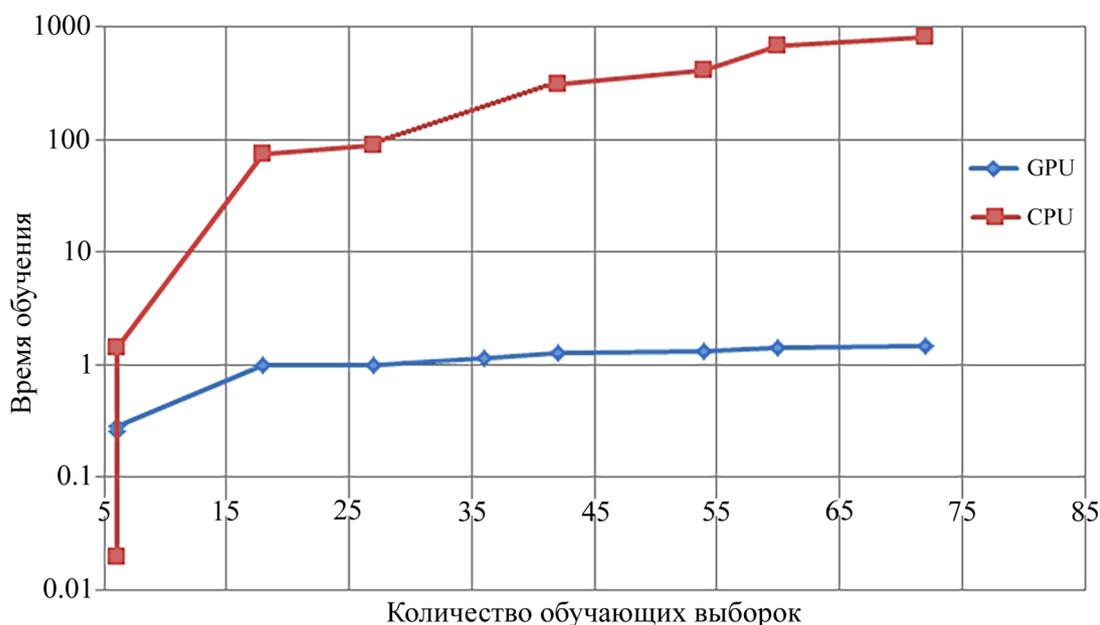


Рис. 2: Время обучения нейронной сети на процессорах с последовательной (CPU) и параллельной (GPU) обработкой данных.

Результаты исследований. Проверка правильности функционирования алгоритма анализа спектров комбинационного рассеяния оценивалась по количеству спектров, которые были верно классифицированы по степени подобия эталонным в результате обработки.

После этого нейронная сеть проверялась на сорока тестовых выборках, и в 37 из 40 случаях (93%) нейронная сеть выдавала правильный результат.

Результаты исследований показали, что использование для расчётов вычислительных средств с параллельной обработкой данных при малых объёмах обучающей выборки является нецелесообразным, так как основные затраты времени уходят на запись данных в глобальную память процессора. Однако при обработке относительно больших массивов данных наблюдается значительное увеличение скорости как обучения нейронной сети, так и анализа спектров.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации, уникальный идентификатор работ (проекта) RFMEFI57817X0237.

Л И Т Е Р А Т У Р А

- [1] В. А. Хабаров, Д. В. Попов, *Технические науки – от теории к практике* **42**, 59 (2015).
- [2] Д. А. Сорокоумова, О. Н. Корелин, А. В. Сорокоумов, *Труды Нижегородского государственного технического университета: Информатика и системы управления* **3**, 77 (2015).
- [3] E. Yoshidaa, K. Shizumaa, S. Endoa, T. Okab, *Nuclear Instruments and Methods in Physics Research* **484**, 557 (2002).
- [4] Fast Artificial Neural Network Library [Электронный ресурс]. <http://leenissen.dk/fann>.
- [5] Г. И. Рудой, *Выбор функции активации при прогнозировании нейронными сетями* (М., Вычислительный центр им. А. А. Дородницына РАН, 2011).
- [6] А. М. Захаров, М. А. Казарян, А. В. Обходский, А. С. Попов, *Краткие сообщения по физике ФИАН* **43**(12), 56 (2016).

Поступила в редакцию 24 июля 2018 г.