

МОДЕЛЬ РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ ДЛЯ НЕЙРОСЕТОВОЙ ДИАГНОСТИКИ И МОНИТОРИНГА БОЛЬНОГО

Для удаленного процесса нейросетевой диагностики и мониторинга пациента разработана модель обработки больших данных (медико-диагностические данные измерительных приборов, изображения исследуемого объекта, информация о пользователе веб-портала [1,2]) в распределенной среде с помощью концепции MapReduce.

В качестве инструментария выбрана многослойная нейронная сеть (MLP), которая обучает входные данные с использованием механизма обратного распространения.

Данные, подаваемые на вход нейронной сети, делятся на блоки определенного размера, в зависимости от количества доступных вычислителей. Запускается функция Map, все операции которой выполняются параллельно и независимо друг от друга. На выходе этой функции вычисляются промежуточные значения выходов нейронов первого слоя. Далее следует распределение полученных промежуточных результатов, основной целью которого является балансировка нагрузки вычислительного кластера.

Затем запускается функция Reduce, связанная с передаваемым от функции Map ключом. Результатом работы этой функции будут вычисленные выходы первого слоя.

Чередование процессов Mapping и Reducing осуществляется до получения окончательного результата. Графовая модель нейрообработки данных на основе MapReduce представлена на рис. 1. Вершина графа соответствуют независимые параллельные операции, а направленным дугам – передачи данных между вычислителями (результаты операций Map передаются как аргументы операциям Reduce).

На рис. 1. используются следующие

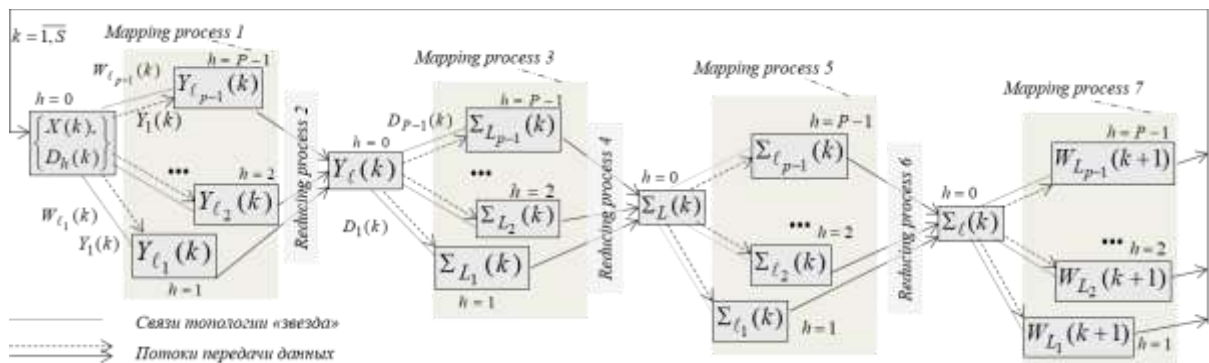


Рис. 1. Модель MapReduce параллельной нейрообработки данных в распределенной среде с топологией «звезда»

обозначения: $X(k) = [x^1(k), \dots, x^{n_L}(k)]^T$ – K входных сигналов, и соответствующие им целевые признаки $D(k) = [d^1(k), \dots, d^{n_L}(k)]^T$ ($k = \overline{1, K}$); $W_\ell(k) = [W_{\ell_1}(k), W_{\ell_2}(k), \dots, W_{\ell_{p-1}}(k)]$ – матрицы весов; $Y_\ell(k) = [y_\ell^1(k), \dots, y_\ell^{n_\ell}(k)]^T$ – вектор выходов нейронов; $\Sigma_L(k)$ – вектор ошибок реакции нейронов для выходного L-го слоя и $\Sigma_\ell(k)$ – для скрытых слоев $\ell = \overline{L-1, 2}$; P – количество вычислителей.

Для оценки эффективности алгоритмов создан экспериментальный кластер Hadoop (версия 2.7.5), состоящий из 5 компьютеров (Intel Core 2 Quad CPU Q8200 @2.33GHz, ОС Microsoft Windows 10) с физической топологией «звезда»: 4 узла являются Datanodes, один – Namenode. Задачи управлялись менеджером YARN. При установке серверов кластера выдвигаются очень жесткие требования к памяти и к жесткому диску. С этим связано ограничение на размер обрабатываемых данных.

Список литературы

1. N. G. Axak, "Development of multi-agent system of neural network diagnostics and remote monitoring of patient", *Eastern-European Journal of Enterprise Technologies*, vol. 4, no. 9(82), pp. 4-11, 2016.
2. Н. Г. Аксак "Мультиагентная система нейросетевой диагностики и удаленного мониторинга пациента", *Інформаційні технології: проблеми та перспективи: монографія*, В. С. Пономаренко, Ред. Харків, Україна: Вид. Рोजко С. Г., 2017, с. 325-340.