Глубокое обучение в задачах классификации временных рядов

О. Ю. Бахтеев, М. С. Попова, В. В. Стрижов

Московский физико-технический институт

Решается задача классификации временных рядов акселерометра мобильного телефона. Строится сеть глубокого обучения. Структура сети — композиция ограниченной машины Больцмана, автокодировщика и двухслойной нейросети. Анализируется зависимость ошибки классификации от числа параметров и размера обучающей выборки. Работа посвящена построению сети глубокого обучения и оптимизации ее параметров с помощью вычислительных мощностей графического ускорителя на основе сервиса облачных вычислений Amazon Web Services.

Эксперимент проводится с использованием Theano — библиотеки для вычислений для языка Python. Theano активно используется для построения моделей глубокого обучения, а также для построения более специализированных библиотек. Функции вычислений Theano компилируются, что позволяет выполнять вычисления за приемлемое время. Другой отличительной особенностью Theano является возможность применения графического процессора с использованием архитектуры параллельных вычислений CUDA. Также в экспериментальном режиме доступна возможность применить интерфейс OpenCL.

**Постановка задачи.**

Пусть задана выборка

|  |  |
| --- | --- |

состоящая из множества пар объект – класс, **x***i* ∈ ℝ*n*. Каждый объект **x** принадлежит одному из *Z* классов с меткой *yi* ∈ **Y** = {1*,*…*,Z*}.

Моделью классификации или сетью глубокого обучения **f** назовем суперпозицию функций

|  |  |
| --- | --- |

где ***μ****k,k* ∈{1*,*…*,* 3}*,* — модели, параметрическое семейство вектор-функции; **w** — вектор параметров моделей; *r*-ю компоненту вектора **f**(**x***,***w**) будем интерпретировать как вероятность отнесения объекта **x***i* к классу с меткой *r*.

Требуется минимизировать функцию ошибки *S* на обучающей выборке:

|  |  |
| --- | --- |

где *S* — сумма отрицательных логарифмов правдоподобия по всем объектам выборки:

|  |  |
| --- | --- |

Сеть глубокого обучения состоит из трех основных компонент: ограниченной машины Больцмана ***μ***1, автокодировщика ***μ***2 и двухслойной нейросети с softmax-классификатором ***μ***3.

**Общий план запуска алгоритма.**

Для запуска алгоритма на сервере AWS требуется зарегистрироваться на AWS, сконфигурировать экземпляр AWS (instance), скопировать код и данные проекта на сервер и подключиться к серверу по протоколу SSH. После этого требуется зайти в папку с проектом, запустить вычислительный эксперимент, скопировать полученные результаты эксперимента на локальный компьютер и уничтожить экземпляр.

**Вычислительный эксперимент.**

Набор данных содержит записи движений для 6 классов переменной длины по трем координатам акселерометра мобильного телефона. Из каждой записи использовались первые 200 сегментов. Пример данных из выборки WISDM приведен на рис. [1](#x1-30011). Т. к. выборка не сбалансирована, в нее добавлялись повторы записей классов, содержащих количество записей, меньшее чем у большего класса.

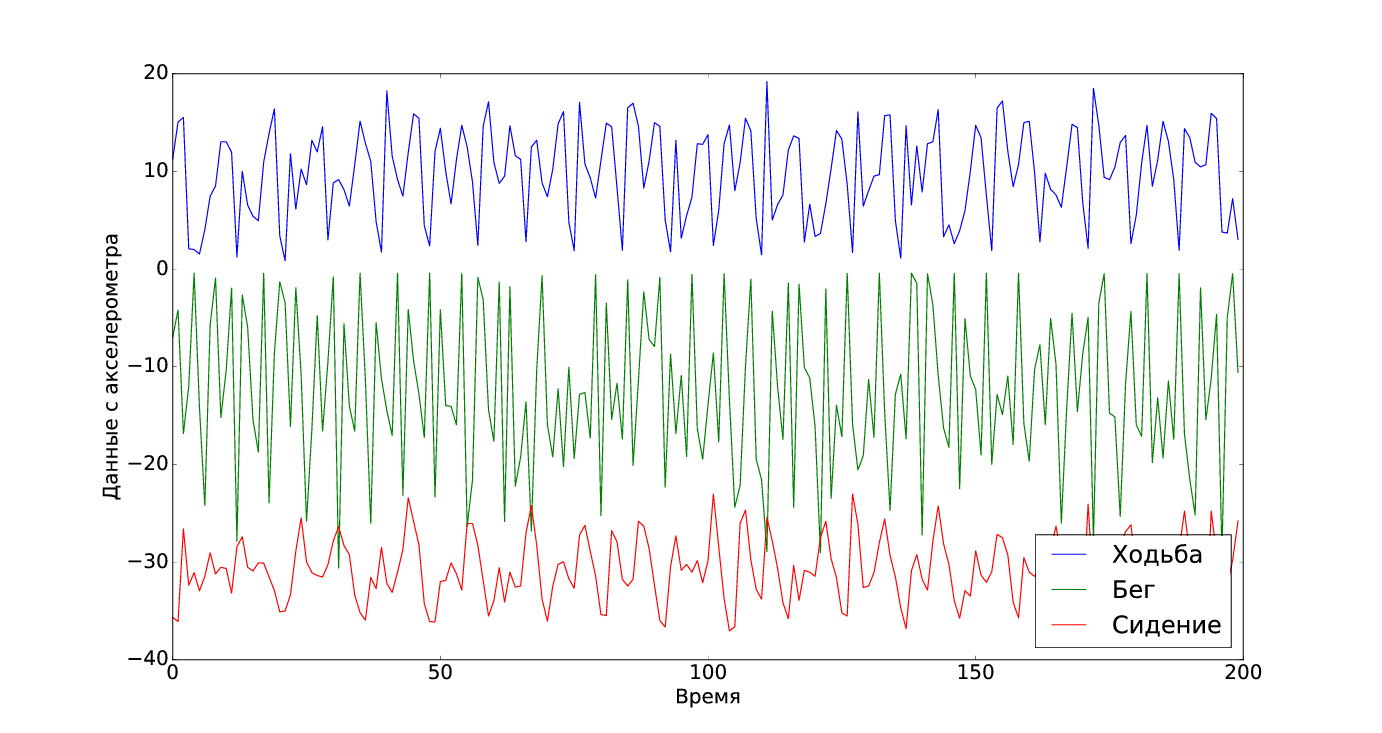
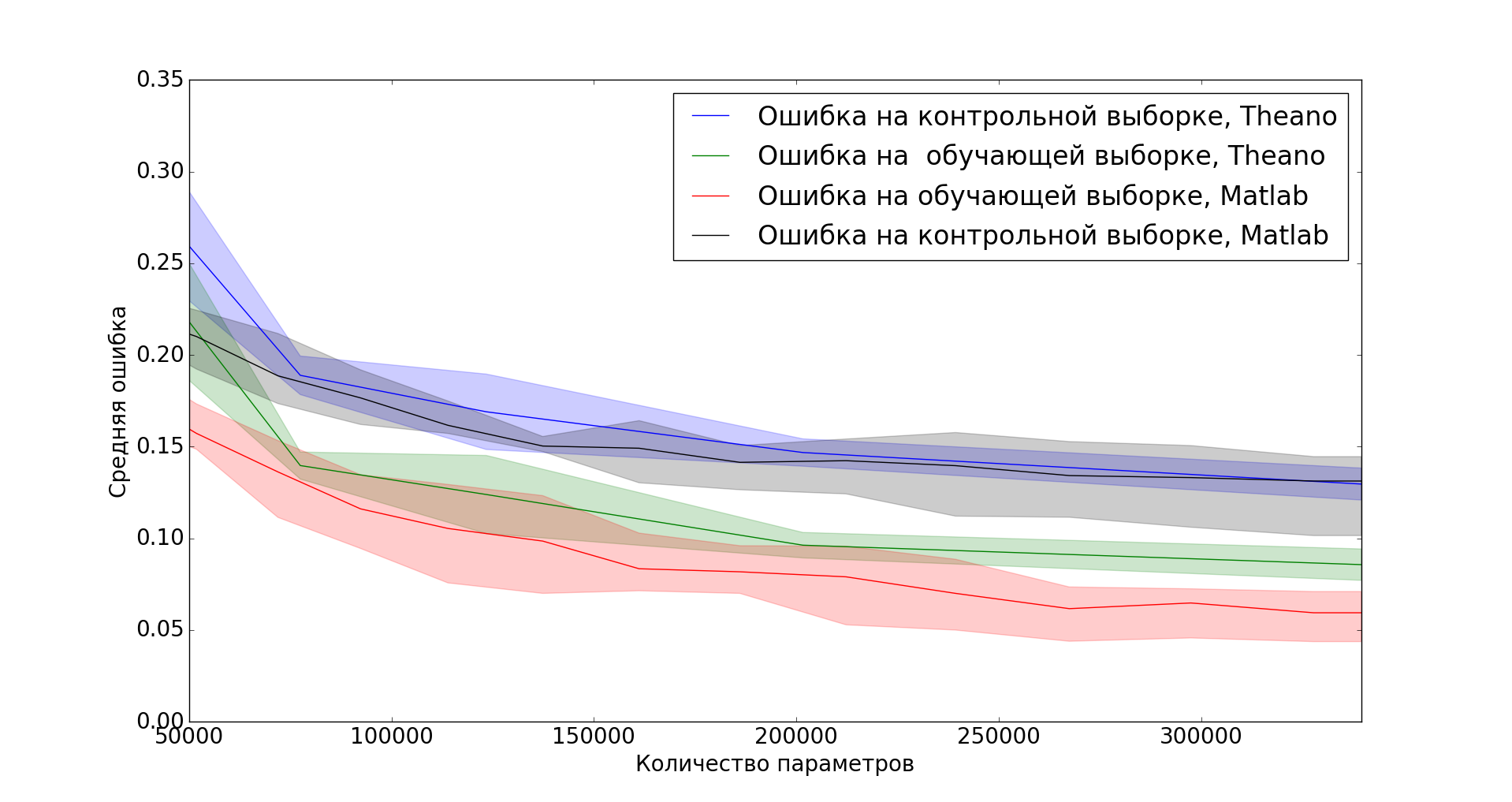


Рис. 1: Пример временных рядов из выборки WISDM, проекция на первую координату. Проекции, соответствующие сидению и бегу были сдвинуты для наглядности.

Рис. 2: Зависимость ошибки от числа нейронов

Для оценки зависимости качества классификации от размера обучающей выборки была проведена кроссвалидация с фиксированным количеством объектов в обучающей выборке (25% исходной выборки) и переменным размером обучающей выборки. Число нейронов было установлено как 364:224:112. При проведении процедуры скользящего контроля для каждого отсчета было произведено пять запусков. График зависимости ошибки классификации от размера обучающей выборки представлен на рис. [3](#x1-30033).

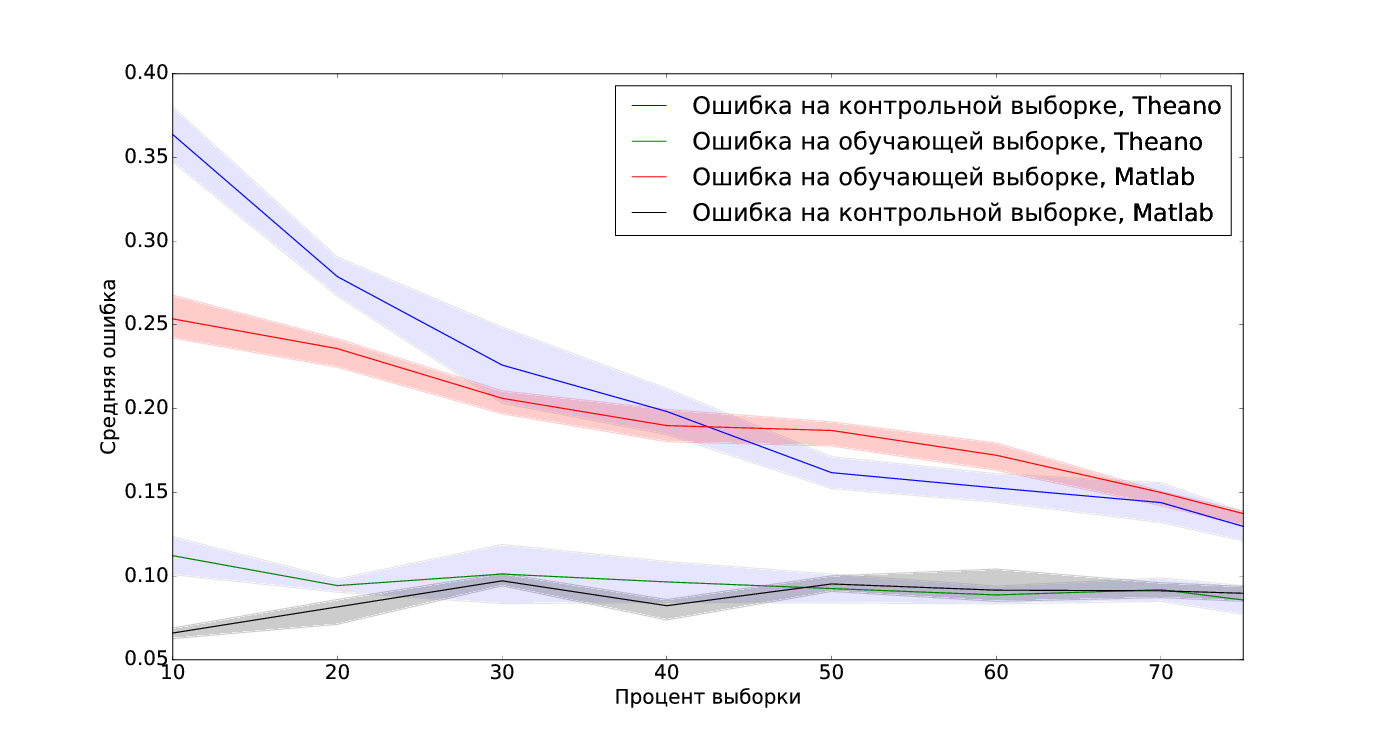


Рис. 3: Зависимость ошибки от размера обучающей выборки

Для исследования скорости работы процесса обучения нейросети в зависимости от конфигурации Theano был сделан следующий эксперимент: проводилось обучение двухслойной нейросети на основе подсчитанных заранее параметров ограниченной машины Больцмана и автокодировщика. Обучение проходило за 100 итераций. При обучении алгоритм запускался параллельно с *n* разными стартовыми позициями, *n* ∈{1*,*…*,* 4}*.* Число нейронов было установлено как 300:200:100. Запуск осуществлялся со следующими конфигурациями Theano:

* вычисление на центральном процессоре, задействовано одно ядро;
* вычисление на центральном процессоре, задействовано четыре ядра;
* вычисление на центральном процессоре, задействовано восемь ядер;
* вычисление на графическом процессоре.

Результаты эксперимента приведены на рис. [4](#x1-30044). Как видно из графика, вычисление с использованием CUDA показывает значительное ускорение по сравнению с вычислением на центральном процессоре. Детальное описание структуры нейросети можно найти в статье *Бахтеев О. Ю., Попова М.С., Стрижов В.В.* Системы и средства глубокого обучения в задачах классификации // Системы и средства информатики. — 2016. — № 2.

Рис. 4: Результаты эксперимента по исследованию скорости процесса обучения