

В. Ю. Кузьмин (ООО Вай2Гео). Сравнение многослойного перцептрона и LSTM сетей для прогнозирования моментов вероятностных распределений.

В работе проведено сравнение различных архитектур нейронных сетей в задаче построения прогнозов значений моментов конечных нормальных смесей (математического ожидания, дисперсии, коэффициентов асимметрии и эксцесса), полученных с помощью аппроксимации методом скользящего разделения смесей [1].

Методология поиска оптимальных гиперпараметров нейронных сетей была ранее апробирована при рассмотрении задачи отнесения прогнозируемого значения к одному из диапазонов возможных значений [2], в данной работе при применении схожей методологии осуществлен переход к построению непрерывных прогнозов.

В качестве входных данных для нейросетей используется последовательность моментов, выходными являются прогнозы значений на несколько следующих шагов. Рассмотрен набор конфигураций топологий сети, в частности, с 1, 2 и 3 скрытыми слоями и различным количеством нейронов. Каждая конфигурация была обучена в режиме многослойного перцептрона (функция активации ReLU) и как рекуррентная LSTM сеть [3]. Качество построения прогноза сравнивалось в стандартных метриках MSE (средней квадратической погрешности) и MAE (средняя абсолютная ошибка). Кроме того, анализировались временные затраты на обучение.

Показано, что LSTM-сети значительно улучшают показатели по MSE/MAE-метрикам — примерно на порядок по сравнению с перцептроном. При этом на высокопроизводительном оборудовании с использованием GPU-ускорения многослойные перцептроны обучаются на 1–2 порядка быстрее, что позволяет приблизиться к построению прогнозов в реальном времени. Анализ топологий показал, что увеличение количества нейронов в слоях, равно как и увеличение количества слоев не всегда приводит к улучшению качества прогноза — наилучшие результаты получены на одном скрытом слое для перцептрона и на двух слоях для LSTM-сети. Для LSTM сети на выбранных конфигурациях были сравнены функции активации `softsign` и `tanh`, рассмотрен ряд оптимизаторов (`adam`, `adadelta` и пр.). Показано, что с учетом примененной методологии для всех рядов моментов наилучшая точность достигается при применении оптимизатора `adam` с функцией активации `tanh`.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 18-29-03100).

СПИСОК ЛИТЕРАТУРЫ

1. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. М.: Изд-во Московского ун-та, 2011, 512 с.
2. *Gorshenin A. K., Kuzmin V. Yu.* Improved architecture of feedforward neural networks to increase accuracy of predictions for moments of finite normal mixtures. — Pattern Recognition and Image Analysis, 2019, v. 29, № 1, p. 79–88.
3. *Greff K., Srivastava R. K., Koutnik J., Steunebrink B. R., Schmidhuber J.* LSTM: A Search Space Odyssey. — IEEE Transactions on Neural Networks and Learning Systems, 2017, v. 28, is. 10, p. 2222–2232.