

АЛГОРИТМ ФОРМИРОВАНИЯ СВЕРТОЧНЫХ АНСАМБЛЕЙ В НЕРЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЯХ ДЛЯ ОБРАБОТКИ ПРОГНОЗОВ ВРЕМЕННЫХ РЯДОВ В АВТОМАТИЧЕСКИХ ТОРГОВЫХ СИСТЕМАХ

Горелов Михаил Андреевич

МГТУ им. Баумана
misha_gorelov@mail.ru

AN ALGORITHM FOR THE FORMATION OF CONVOLUTIONAL ENSEMBLES IN NON-RECURRENT NEURAL NETWORKS FOR PROCESSING TIME SERIES FORECASTS IN AUTOMATIC TRADING SYSTEMS

M. Gorelov

Summary. The article deals with the main features of convolutional neural networks, their structure and application features. An algorithm for forming convolutional ensembles for processing time series forecasts based on the averaging of subnetwork predicates is proposed.

Keywords: neural network, time series, recurrent neural network, forecast.

Аннотация. В статье рассмотрены основные особенности нерекуррентных нейронных сетей сверточного типа, их структура и особенности применения. Предложен алгоритм формирования сверточных ансамблей для обработки прогнозов временных рядов на базе усреднения предикатов подсетей.

Ключевые слова: нейронная сеть, временной ряд, рекуррентная нейронная сеть, прогноз.

Оценка финансовых временных рядов, предсказания будущих изменений цен и определение паттернов движения рыночных показателей активно развивающаяся область исследований в области искусственных нейронных сетей (ИНС) [12]. Они находят широкое применение в различных подходах, описанных в литературе [1] и основанных на интеллектуальном анализе текста и обработке естественного языка при анализе и использовании различной информации для построения нелинейной прогнозной модели для точного прогнозирования цен [2]. Исходя из предложенных методов в таких подходах распространено применение ансамблей сверточных нейронных сетей [3] для достижения более высокого уровня совпадений в результатах оценки финансовых временных рядов.

Сверточная нейронная сеть или Convolutional Neural Network (CNN) — ИНС со специальной архитектурой, разработанной в 1988 году. CNN обладают рядом особенностей, которые эквивалентны некоторым особенностям зрительной коры, что позволяет ИНС приобрести значительные возможности в распознавании паттернов и распознавании зрительных образов во многих случаях достигающих возможностей человека [4]. Преимущество использования CNN для временных рядов состоит в том, что они могут эффективно извлекать знания и изучать

внутреннее представление непосредственно из необработанных данных временных рядов, и им не требуются специальные знания из области приложения для фильтрации входных функций [5]. Поэтому можно сказать, что CNN обладают большим потенциалом для выявления сложных закономерностей, скрытых в данных временных рядов.

Для того чтобы рассмотреть различные подходы формирования сверточных ансамблей необходимо проанализировать типовую архитектуру сверточной нейронной сети.

Входной слой — непосредственное представление входных данных пригодных для обработки следующими слоями сети — формула (1)

$$I_n = W \times H \times D, \quad (1)$$

где I_n — размерность входного слоя сети, W — длительность эпохи временного ряда, H — допустимые амплитуды значений предсказуемой величины, D — количество зависимых слоев временного ряда.

Сверточный слой — основной слой рассматриваемой ИНС, предназначен для выделения признаков

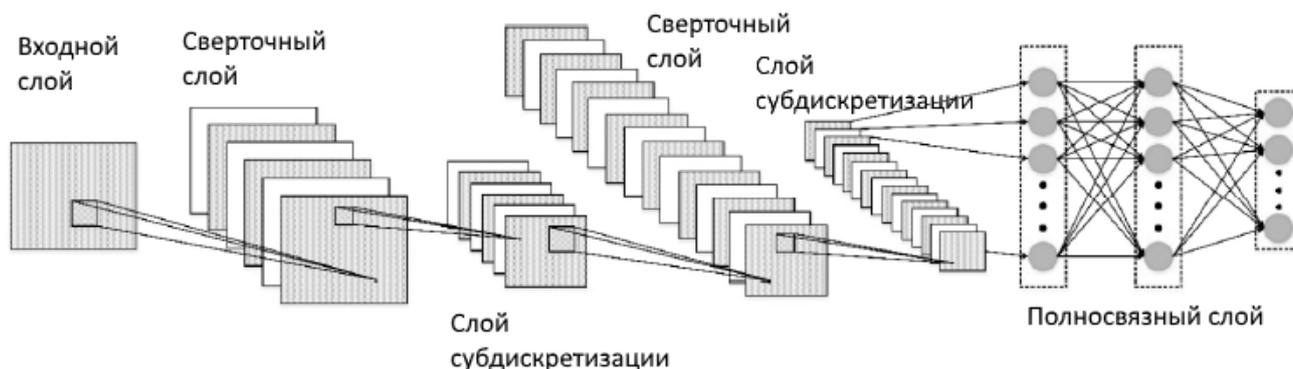


Рис. 1. Базовая архитектура CNN

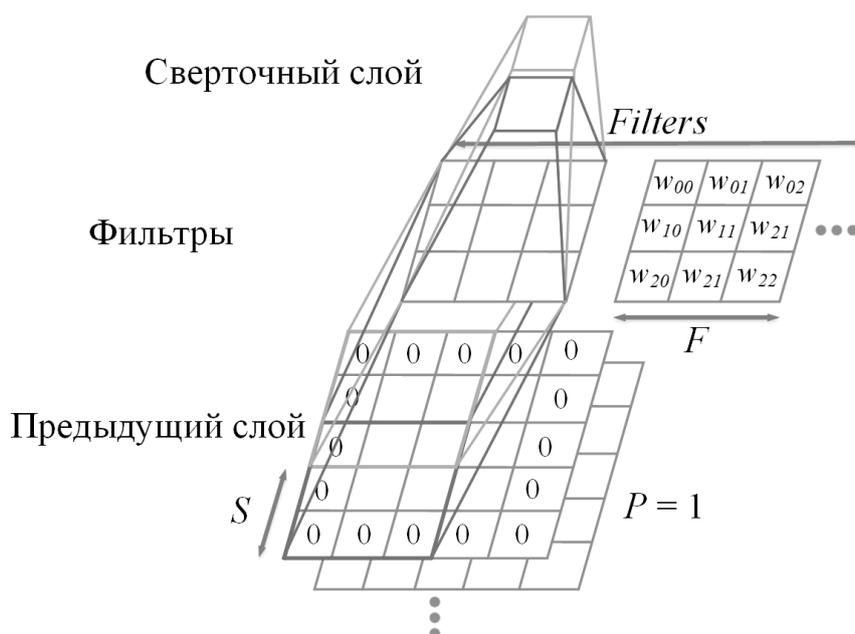


Рис. 2. Формирование сверточного слоя нейронной сети

и их преобразования с целью использования в более глубоких слоях сети и получения более сложных признаков.

$$y_i = \sum_i f(x_i \times \omega_i + b_i), i = 1, 2, \dots, k \quad (2)$$

где w_i и b_i вес и смещение соответственно, $f()$ — функция активации, \times — операция свертки.

Состоит из двух и более мерных матриц весов связей нейронов предыдущего слоя с нейронами сверточного слоя. Каждый такой слой применяет операцию свертки на входные данные. В связи со схожестью данной операции с операцией фильтрации данных, такие слои так же часто называют фильтрами. Количество и тип таких фильтров определяет то какое количество и какие признаки необходимо выявить во временном

ряду. Саму операцию свертки можно рассматривать как применение и скольжение одномерного (временного) фильтра по временному ряду [6]. Более того, поскольку выходом свертки является новый фильтрованный временной ряд, применение нескольких сверток подразумевает создание многомерного временного ряда, размерность которого равна количеству используемых фильтров в слое. Обоснование этой стратегии заключается в том, что применение нескольких сверток приводит к генерации нескольких отличительных признаков, которые обычно улучшают производительность модели. На практике доказано, что этот вид слоя очень эффективен, и наложение различных сверточных слоев позволяет более глубоким слоям изучать высокоуровневые или более абстрактные функции, а слои, расположенные рядом с входом, позволяют изучать низкоуровневые функции.

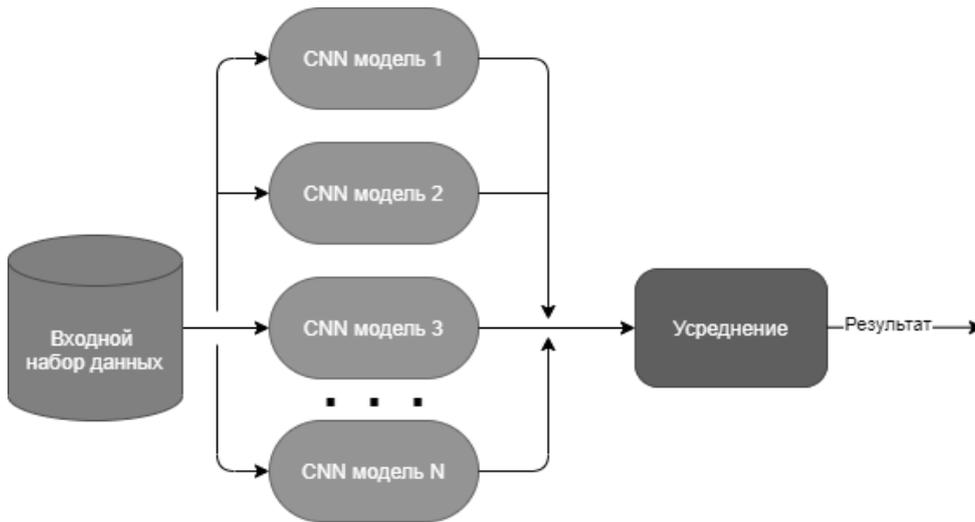


Рис. 3. Обобщенная структура сверточного ансамбля созданного по стратегии усреднения

Объединение результатов сверточных слоев позволяет устранить ограничения, когда карты объектов созданные такими слоями повторяют входной ряд. Результаты свертки фильтров объединяют в скользящее окно, уменьшая длину ряда чтобы достичь некоторой инвариантности трансляции обученных функций. Такое объединение часто выделяют в отдельный слой, который называется — **слой субдискретизации**.

$$z = \frac{1}{N} \sum_{(i,j) \in S} x_{i,j}, i, j = 1, 2, \dots, p \tag{3}$$

где p — ширина сверточного слоя, а S — квадратная матрица шириной p , N — общее количество элементов, $x_{i,j}$ — выходной весовой коэффициент сверточного фильтра.

Помимо цели сокращения вероятности быстрого переобучения, данный слой способствует сокращению вычислительных затрат и расходу памяти. Наиболее важным параметром данного слоя является ширина сверточного окна и функция субдискретизации, которая определяет какие данные берутся из результатов фильтров.

Полносвязный слой, основное назначение данного слоя — преобразование сигналов, полученных на сверточных уровнях сети к одномерному виду и выделение признаков на одномерном уровне. Данный слой также может использоваться в качестве последнего (выходного) слоя.

$$Y = \sum_i f(\omega z + b) \tag{4}$$

где Y и z — выходной вектор и входные признаки, а ω и b вес и смещение соответственно.

При беглой оценке может показаться что при достаточном увеличении количества фильтров и слоев субдискретизации можно достичь необходимых результатов в оценке данных временного ряда, однако, результаты, производимые такими сетями, зависят от множества параметров, так, для CNN обученной методом градиентного спуска итоговая нелинейная модель анализа временного ряда будет зависеть от порядка данных в каждой эпохе, инициализации, скорости обучения и т.д.

В связи с этим более широкое распространение нашел подход, применяющий объединение таких сетей, обученных независимо, для получения агрегированного прогноза как элегантное решение для уменьшения обобщенной ошибки [7–9].

Основной принцип любой ансамблевой стратегии — взвесить несколько моделей и объединить их индивидуальные прогнозы для улучшения характеристик прогнозирования; в то время как ключевым моментом для эффективности ансамбля является то, что его компоненты должны характеризоваться точностью и разнообразием в своих предсказаниях [10]. Таким образом, комбинация прогнозов нескольких моделей добавляет смещение, которое, в свою очередь, противодействует дисперсии одной обученной модели. Следовательно, уменьшая дисперсию прогнозов, ансамбль моделей может работать лучше, чем любая наилучшая модель.

В литературе было предложено несколько стратегий для проектирования и разработки ансамбля регрессионных моделей. Наиболее эффективные и широко используемые стратегии: усреднение по ансамблю, сбор и суммирование.

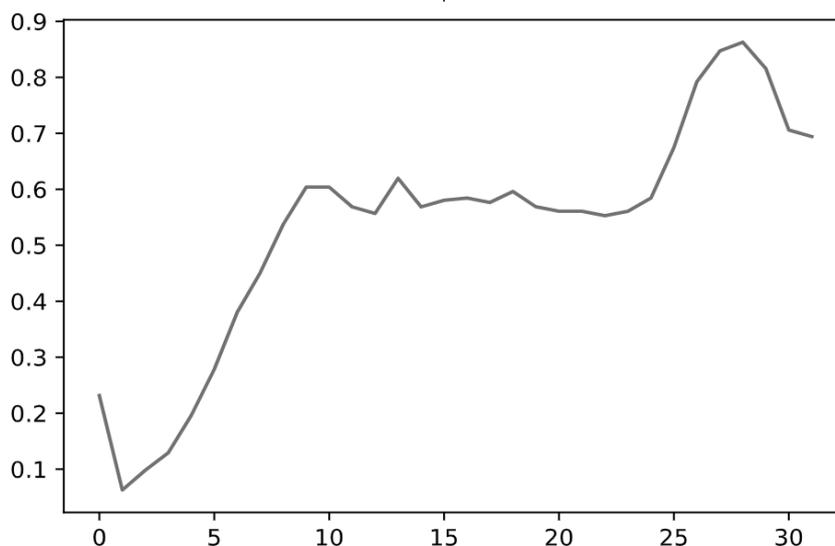


Рис. 4. Исходные данные для обучения

В рамках данной статьи для создания алгоритма рассмотрим подробнее стратегию усреднения выходных результатов по ансамблю.

При использовании данной стратегии общий принцип строится на том что каждая сверточная ИНС должна быть обучена с применением различных входных параметров, но общем входном наборе данных, иными словами при различной инициализации. Результатом такого обучения будет предиктор ϕ_k — коэффициент, характеризующий будущее изменение контролируемого параметра. Таким образом при использовании подобного метода ансамблевого обучения мы получаем множество единичных предикторов с целью получения наилучшего ϕ_a что и будет являться целью работы алгоритма по созданию сверточного ансамбля [11].

Алгоритм реализации ансамбля сверточных нейронных сетей по стратегии усреднения предикатов принимает следующий вид:

Определение основных параметров применяемых CNN моделей 1-N:

1. Размерность сверточного слоя;
 - a. Размерность слоя субдискретизации;
 - b. Уровень сети — количество пар скрытых слоев;
 - c. Определение количества подсетей;
 Данные параметры должны быть общими для всех подсетей ансамбля.
 - a. Формирование параметров входных слоев сетей;
 - b. Определение и нормализация размеров входного набора данных;
2. Ввод глубины слоя, обычно равному количеству взаимосвязей в наборе данных;

3. Ввод условий инициализации для каждой используемой CNN ансамбля;
4. Последовательная подготовка сверточных слоев подсетей;
5. Последовательная подготовка слоев субдискретизации подсетей;
6. Повтор пунктов 4–5 для каждого уровня сетей (см 1.с);
7. Выбор функции активации полносвязного слоя сетей;
8. Принятие решения о необходимости формирования отдельного выходного слоя для формирования предиктора подсети;
9. Независимое обучение подсетей;
10. Агрегация выходных предикторов в блоке усреднения;
11. Формирования итогового предиктора;
12. Оценка результата и ошибки;
13. Принятие решения о необходимых корректировках;

Рассмотрим результаты работы сверточного ансамбля нереккурентной нейронной сети сформированного с использованием данного алгоритма.

В качестве временного ряда использовались данные из открытых источников. Внешний вид среза, по одной эпохе и одному признаку, временного ряда, использовавшегося для обучения приведен на рис. 4.

Определим основные параметры общие для всех моделей CNN ансамбля:

1. Размерность сверточного слоя $F = 2$, Размерность слоя субдискретизации $p = 1$, уровень сети

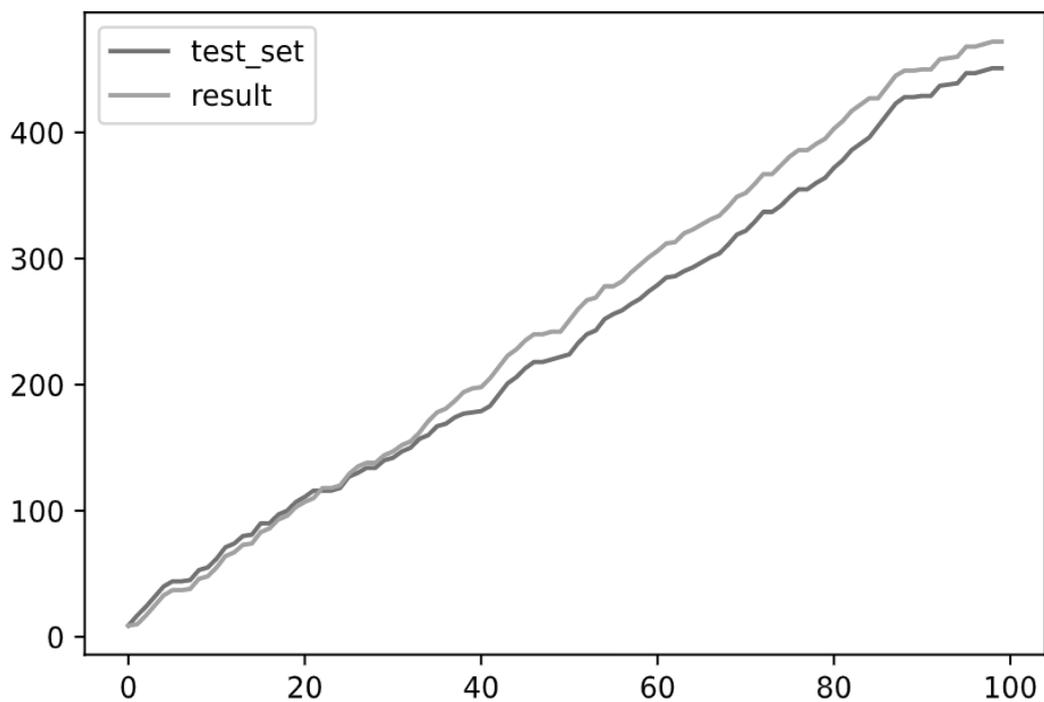


Рис. 5. Сверточная нейронная сеть. Сравнение тестовой последовательности и предсказания выхода нейросетью. Подсеть #1. $s = 0.2417$

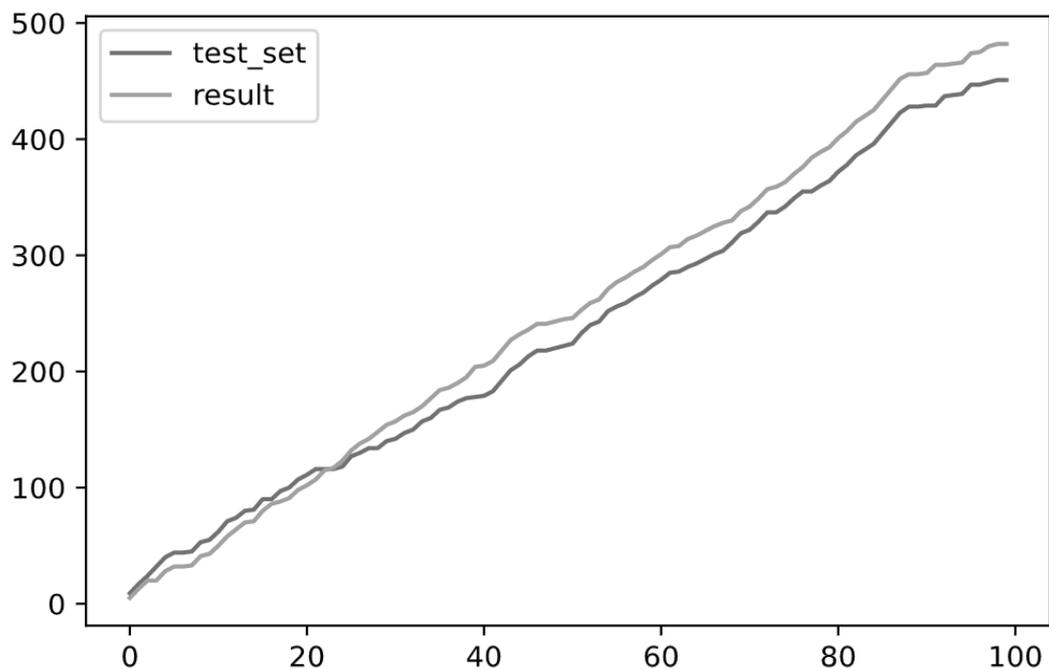


Рис. 6. Сверточная нейронная сеть. Сравнение тестовой последовательности и предсказания выхода нейросетью. Подсеть #2. $s = 0.2815$

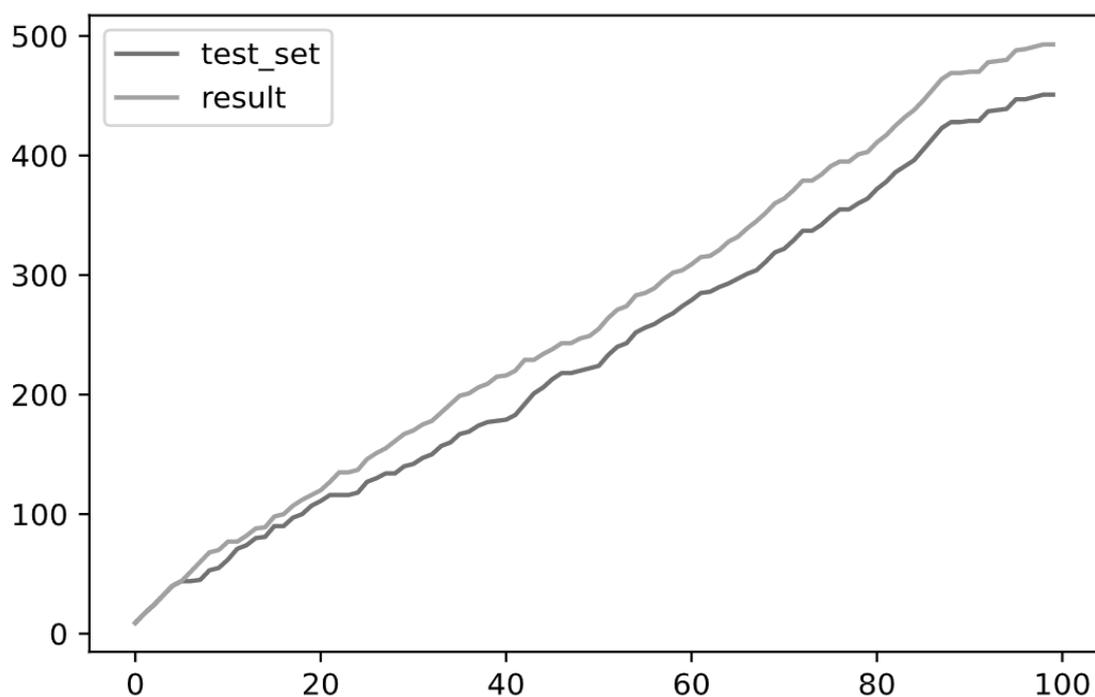


Рис. 7. Сверточная нейронная сеть. Сравнение тестовой последовательности и предсказания выхода нейросетью. Подсеть #3. $s = 0.3287$

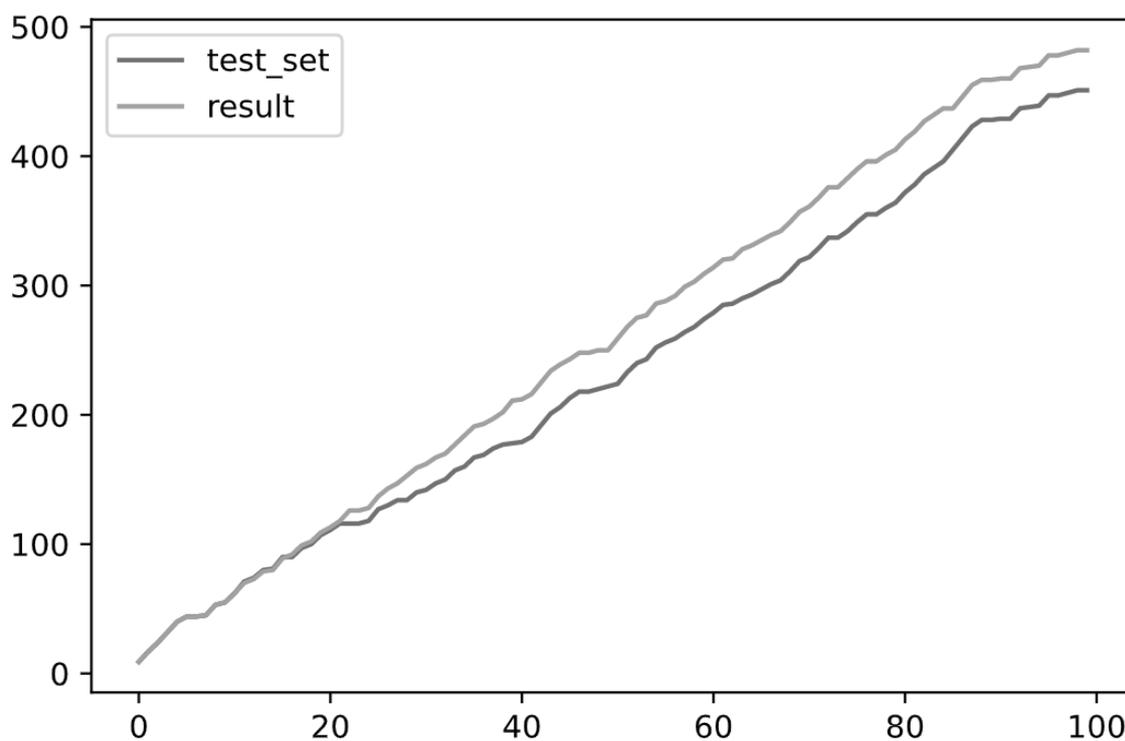


Рис. 8. Ансамбль сверточных нейронных сетей. Сравнение тестовой последовательности и предсказания выхода нейросетью. $s = 0.2051$

принимая равным $n = 1$, а количество подсетей $N = 3$.

2. Произведем подготовку входных данных. Принимаем глубину входного слоя $D = 1$, проводим нормализацию данных.
3. Произведем подготовку всех подсетей ансамбля исходя из указанных параметров из п. 1.
4. В качестве функции активации полносвязного слоя принимаем сигмоидную функцию:

$$AF_{fc} = \frac{1}{1 + e^{-x}}$$

где x — сигнал нейрона полносвязного слоя, полученный в результате взвешенного суммирования входных сигналов нейрона.

5. Усреднение будем проводить без весовых модификаций входных значений.

Далее итеративно, изменяя наиболее важные параметры сетей — F и n получим семейство результирующих графиков иллюстрирующих результаты работы отдельных нейросетей входящих в состав ансамбля (рис. 5–7) и ансамбля в целом (рис. 8). Для улучшения читаемости приводятся срезы графиков в наихудших отрезках. Суммарная ошибка предиктора, s , приведена на графиках.

Таким образом итеративное применение данного алгоритма с учетом коррекций, производимых в подсетях ансамбля, позволяет получить результат предсказания выходной величины временного ряда с наименьшей ошибкой относительно одиночной сверточной нейросети. Однако, формирование обучение подобных нейросетей может потребовать дополнительные временные ресурсы.

ЛИТЕРАТУРА

1. S. Mehtab, J. Sen, S. Dasgupta. Analysis and Forecasting of Financial Time Series Using CNN and LSTM-Based Deep Learning Models. Режим доступа: <https://arxiv.org/abs/2011.08011> (дата обращения: 17.12.2020).
2. S. Mehtab and J. Sen, "A robust predictive model for stock price prediction using deep learning and natural language processing", In Proc. of the 7th International Conference on Business Analytics and Intelligence, Bangalore, India, 2019.
3. S. Mehtab and J. Sen, "Stock price prediction using convolutional neural network on a multivariate time series", In Proc. of the 3rd National Conference on Machine Learning and Artificial Intelligence (NCMLAI'20), New Delhi, India, 2020.
4. Livieris, I.E.; Pintelas, E.; Stavroyiannis, S.; Pintelas, P. Ensemble Deep Learning Models for Forecasting Cryptocurrency Time-Series. *Algorithms* 2020, 13, 121.
5. Livieris, I.E.; Pintelas, E.; Pintelas, P. A CNN-LSTM model for gold price time-series forecasting. *Neural Comput. Appl.* doi:10.1007/s00521-020-04867-x. Available online: <https://link.springer.com/article/10.1007/s00521-020-04867-x> (accessed on 20.12.2020).
6. Michelucci, U. *Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection*; Springer: Berlin/Heidelberg, Germany, 2019.
7. Rokach, L. *Ensemble Learning: Pattern Classification Using Ensemble Methods*; World Scientific Publishing Co Pte Ltd.: Singapore, 2019.
8. Lior, R. *Ensemble Learning: Pattern Classification Using Ensemble Methods*; World Scientific: Singapore, 2019; Volume 85.
9. Zhou, Z. H. *Ensemble Methods: Foundations and Algorithms*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2012.
10. Bian, S.; Wang, W. On diversity and accuracy of homogeneous and heterogeneous ensembles. *Int. J. Hybrid Intell. Syst.* 2007, 4, 103–128.
11. Chen, L. and G. Shakhnarovich. "Learning Ensembles of Convolutional Neural Networks." (2014).
12. Осовский С. Нейронные сети для обработки информации / Пер. с польского И. Д. Рудинского. — М.: Финансы и статистика, 2002. — 344 с.: ил.

© Горелов Михаил Андреевич (misha_gorelov@mail.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»