

УДК 004.85

**ПРОГНОЗИРОВАНИЕ КОТИРОВОК ЦЕННЫХ БУМАГ МЕТОДАМИ
ЛИНЕЙНОЙ РЕГРЕССИИ, ДЕРЕВА РЕШЕНИЙ И С ПОМОЩЬЮ
МНОГОСЛОЙНОЙ НЕЙРОННОЙ СЕТИ**

**Ким Наталья Георгиевна
Хлебородова Лидия Дмитриевна**
студенты

Научный руководитель: **Осипов Геннадий Сергеевич**
д.т.н., зав. кафедрой Информатики
ФГБОУ ВО «Сахалинский государственный университет»

Аннотация: Разработан программный комплекс, позволяющий решать задачу прогнозирования котировок ценных бумаг на основании обучающей выборки, содержащей более тысячи наблюдений.

Апробировались и сравнивались методы линейной регрессии, дерева решений (decision tree) и искусственной нейронной сети.

Ключевые слова: прогноз котировок ценных бумаг, линейная регрессия, дерево решений, нейронная сеть.

**PREDICTING STOCK PRICES USING LINEAR REGRESSION,
DECISION TREE, AND MULTI-LAYER NEURAL NETWORK**

**Kim Natalia Georgievna
Khleborodova Lydia Dmitrievna**
Scientific adviser: **Osipov Gennady Sergeevich**

Abstract: A software package has been developed that allows solving the problem of predicting securities quotes based on a training sample containing more than a thousand observations.

Linear regression, decision tree and artificial neural network methods were tested and compared.

Key words: stock price forecast, linear regression, decision tree, neural network.

1. Постановка задачи

Объектом и предметом исследования является проблема построения функции f регрессии (авторегрессии) вида [1]:

$$f[\mathbf{x}(\tau), y(\tau)] \Rightarrow y(t+1) | t \geq 0$$

$\tau=0,t$

где $\mathbf{x}(\cdot)$ – вектор внутренних аргументов;

$y(\cdot)$ – прогнозируемая величина.

Целью – отработка методологии синтеза искомой функции f методом извлечения новых знаний из базы данных с помощью методов линейной регрессии, дерева решений и обучаемой многослойной искусственной нейронной сети.

В качестве исследовательского инструментария выбрана система символьной математики Wolfram Mathematica [2], которая является одной из наиболее продвинутых унифицированных платформ компьютерного моделирования, решения задач машинного обучения и разработок в области синтеза систем искусственного интеллекта.

2. Обучающая выборка

На рисунке 1 представлен оператор ввода импорта данных, а также начальный и конечный фрагменты обучающей выборки из 1003 записей.

```
data = First @ Import["Test.xlsx", "Dataset", "HeaderLines" -> 1]
```

Date	Open	High	Low	Close	Amount	Volume	Close+
Fri 27 Jul 2018 00:00:00	243.99	244.74	241.06	243.25	2541960.	617617305.	243.73
Sat 28 Jul 2018 00:00:00	244.8	246.9	243.5	243.73	2966460.	727341923.	242.65
Sun 29 Jul 2018 00:00:00	243.52	244.72	242.11	242.65	2712340.	659072971.	240.82
Mon 30 Jul 2018 00:00:00	242.0	242.64	240.03	240.82	2103330.	507203050.	240.87
Tue 31 Jul 2018 00:00:00	241.75	243.5	239.15	240.87	2594210.	626355853.	239.1
Wed 21 Apr 2021 00:00:00	318.0	319.55	316.6	318.0	3446200.	1097277610.	315.9
Thu 22 Apr 2021 00:00:00	319.0	319.85	310.75	315.9	6157270.	1939588227.	315.1
Fri 23 Apr 2021 00:00:00	315.15	317.7	312.75	315.1	4402740.	1388166557.	318.5
Sat 24 Apr 2021 00:00:00	314.75	318.5	313.8	318.5	2654440.	839845251.	317.85


Рис. 1. Фрагменты обучающей выборки

В исследуемой задаче $\mathbf{x} = (Open, High, Low, Close, Amount, Volume)$, $y = (Close, Close +)$ – текущая и прогнозная цена закрытия.

3. Построение линейной регрессии

На рисунке 2 приведен фрагмент текста в среде Wolfram Mathematica для обучения и тестирования (построения прогноза на требуемый отрезок времени) при использовании линейной регрессии

```
(*Строим модель линейной регрессии*)
LR = Predict[data[[All, 2 ;;]] → "Close+", Method → "LinearRegression"]
      |предсказать |всё |метод

PredictorFunction [  Input type: Mixed (number: 6)
                    Method: LinearRegression
                    Number of training examples: 1003 ]


(*Используем модель для прогноза*)
LR[{319, 322.5, 317, 317.85, 3491430, 1117478344}]
```

Рис. 2. Прогноз методом линейной регрессии

4. Построение модели дерева решений

На рисунке 3 приведены аналогичные по функционалу фрагменты текста при использовании в качестве метода предсказания метода дерева решений

```
(*Строим модель дерева решений*)
DT = Predict[data[[All, 2 ;;]] → "Close+", Method → "DecisionTree"]
      |предсказать |всё |метод

PredictorFunction [  Input type: Mixed (number: 6)
                    Method: DecisionTree
                    Number of training examples: 1003 ]

(*Используем модель для прогноза*)
DT[{319, 322.5, 317, 317.85, 3491430, 1117478344}]
```

Рис. 3. Прогноз методом дерева решений

5. Сравнение методов линейной регрессии и дерева решений

Графическая интерпретация сравнения методов линейной регрессии и дерева решения при решении задачи интерполирования и прогнозирования приведена на рисунке 4. Для наглядности выведены только последние 250 (финальных) данных

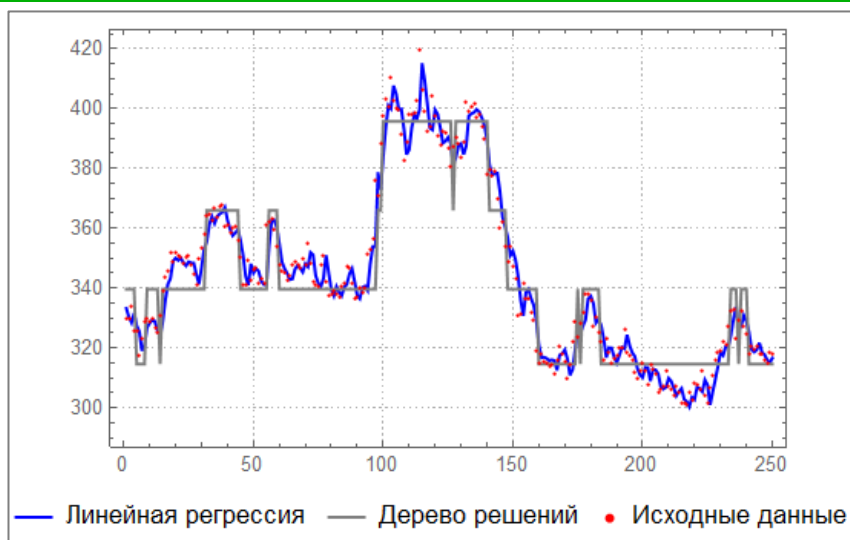


Рис. 4. Сравнение результатов обучения

6. Использование нейронной сети

За последнее время нейронные сети стали одной из базовых парадигм компьютерного моделирования и составляют основу синтеза систем искусственного интеллекта.

В качестве примера на рисунке 5 представлен фрагмент текста по созданию в среде Wolfram Mathematica многослойной сети прямого распространения с шестью нейронами во входном слое (в соответствии с размерностью вектора внутренних аргументов $x(\cdot)$) и одним нейроном в выходном слое (в системе одна прогнозируемая величина $Close +$).

```

Создаем нейронную сеть с 6 нейронами во входном слое, одним скрытым слоем,
содержащим 2 нейрона и одним нейроном в выходном слое.
В качестве функции активации используется сигмоида

chain = NetChain[{
  [нейронная сеть]
  LinearLayer["Input" -> Length[maxdata], "Output" -> 2],
  [линейный слой] [длина]
  LogisticSigmoid,
  [логистическая функция]
  LinearLayer["Input" -> 2, "Output" -> 1],
  [линейный слой]
  LogisticSigmoid,
  [логистическая функция]
  ElementwiseLayer[({# * maxtarget &})]
  [слой, где заданная функция выполняется для каждого элемента]
}]

NetChain [
  uninitialized
  Input vector (size: 6)
  1 LinearLayer vector (size: 2)
  2 LogisticSigmoid vector (size: 2)
  3 LinearLayer vector (size: 1)
  4 LogisticSigmoid vector (size: 1)
  5 420. x vector (size: 1)
  Output vector (size: 1)
]
    
```

Рис. 5. Синтез нейронной сети

Рисунок 6 демонстрирует текст и заключительный момент обучения нейронной сети с заданными основными параметрами – обучающей и тестовой выборкой, количеством эпох обучения.

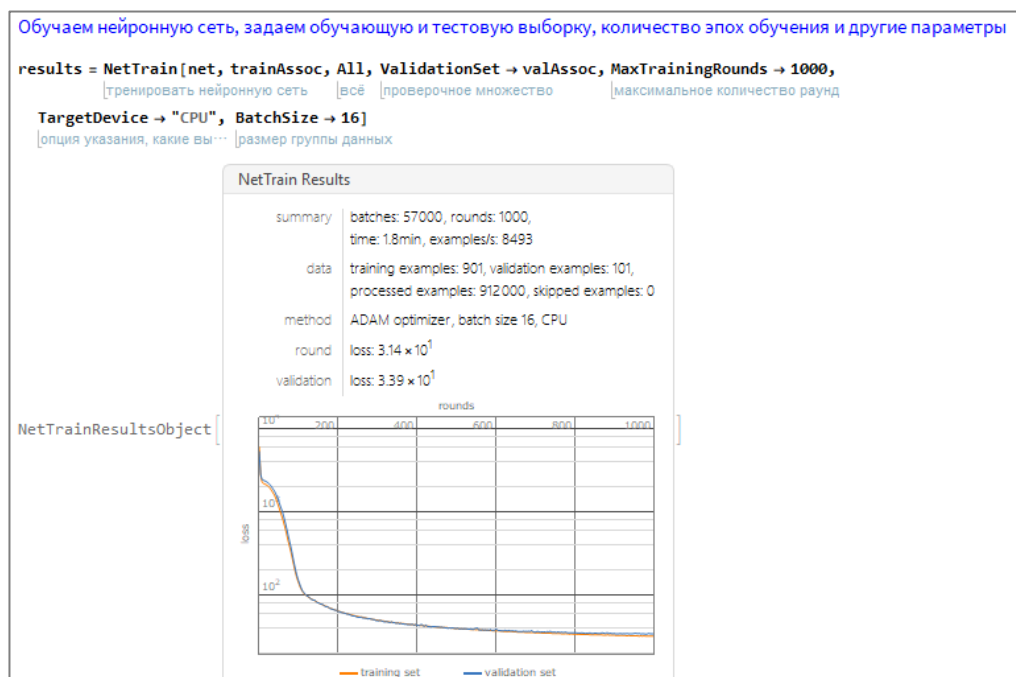


Рис. 6. Обучение нейронной сети

Выводы

Разработанное программное обеспечение предназначено для исследования и прогнозирования комбинированных стохастических временных рядов, обладающих сезонными колебаниям, постепенным усвоением инсайдерской информации, подверженным финансовым пузырям, взлетам и падениям.

Используемые методы машинного обучения позволяют получить прогноз с относительной ошибкой предсказания не более одного процента.

Список литературы

1. Осипов Г.С., Вашакидзе Н.С., Филиппова Г.В. Основы прогнозирования финансовых временных рядов на базе *NeuroXL Predictor* // Постулат. 2017. № 7. С. 26.

2. Stephen Wolfram. An Elementary Introduction to the Wolfram Language. URL: <https://www.wolfram.com/language/elementary-introduction/2nd-ed/> (Дата обращения 07.05.2021).

© Н.Г. Ким, Л.Д. Хлебородова, Г.С. Осипов 2021