Н.А. Костин, А.А. Белоусов

РЕАЛИЗАЦИЯ И ИССЛЕДОВАНИЕ РЕШЕНИЯ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ ДИНАМИКИ КУРСА КРИПТОВАЛЮТ С ПОМОЩЬЮ МОДЕЛЕЙ АВТОРЕГРЕССИИ.

(Самарский университет)

Задача прогнозирования временных рядов долгое время актуальна для самых различных приложений. Многочисленные исследования данной темы показали, что поиск индивидуального алгоритма, который лучше всего работает для всех возможных сценариев, безнадежен. Поэтому вместо того, чтобы стремиться к разработке единого улучшенного алгоритма, текущие исследовательские усилия сместились в сторону более глубокого понимания причин, по которым метод прогнозирования может работать хорошо в одних условиях, а в других терпеть неудачу.

Цель данной работы заключается в реализации и исследовании решения задачи прогнозирования динамики курса криптовалют с помощью моделей авторегрессии.

Методология, реализованная в данной работе, заключается в прогнозировании динамики курса с использованием авторегрессионного интегрированного скользящего среднего (модель ARIMA) и сравнение результатов ее работы с прогнозами алгоритма долгосрочной кратковременной памяти (LSTM) [1].

Посредством исследовательского анализа наши усилия были направлены на выявление связей между рядом переменных в данном наборе данных криптовалюты и на оценку наличия корреляций, которые могли бы помочь понять природу цены Биткойн [2]. Впоследствии, посредством применения различных методов, в работе ставилась задача рассмотреть следующий вопрос исследования: «эффективны ли модели авторегрессии в прогнозировании цены криптовалюты».

Поскольку предсказать точную цену очень сложно, мы упростили проблему. Мы пытаемся предсказать, будет ли цена расти, уменьшаться или оставаться неизменной в определенных пределах. То есть формально задачу можно описать следующим уравнением:

$$m[t]=\begin{cases} Down, if\ \frac{p[t]}{p[t-1]} < T^- \\ Stay, if\ \frac{p[t]}{p[t-1]} \geq T^-\ and\ \frac{p[t]}{p[t-1]} \leq T^+ \\ Up, if\ \frac{p[t]}{p[t-1]} > T^+ \end{cases}$$

В данной формуле t является хронологическим индексом сделки, m[t] — это движение цены Биткойна в момент времени t, а p[t] — это цена Биткойна в долларах США в момент времени t. T- и T+ — это нижний и верхний порог соответственно.

Описание разработанного алгоритма для модели ARIMA.

Во-первых, нам необходимо проверить стационарность исходного временного ряда в уровнях (с трендом и без тренда) и то же самое для первых различий. Для этого был использован расширенный тест Дики – Фуллера (ADF) [3].

Анализируя результаты, мы заключили, что цена биткойна не является стационарной по уровню (так как мы не отвергаем нуль, то есть существование единичного корня), но является стационарной в первом отклонении (поскольку мы отвергаем нулевое существование единичного корня), благодаря сравнению статистики ADF со всеми критическими значениями.

Далее мы построили графики автокорреляционной функции (ACF) и частичной автокорреляционной функции (PACF) и проанализировали их и первые отклонения этого временного ряда соответственно. Из анализа коррелограммы мы выяснили, что p = 1 или p = 2 будет порядком авторегрессионной части модели ARIMA.

Для выбора наиболее подходящей модели ARIMA, нами была использована оценка информационным критерием Акаике (AIC) для каждой комбинации p, d, q. В нашем случае минимальный критерий AIC является оптимальным и гласит, что наиболее подходящей моделью для нашего набора данных является ARIMA (1,1,1) [4].

Авторегрессия первого порядка (AR (1)), как и процесс скользящего среднего (MA (1)) являются статистически значимыми, поскольку их коэффициенты отличаются от нуля. Данный вывод мы можем сделать благодаря анализу связанных p-значений (приблизительно равных нулю, то есть ниже, чем любой уровень значимости).

Описание алгоритма и функционирования элементов модели LSTM.

Во-первых, элемент забывания соглашается с тем, какая информация должна храниться в обновлении до следующего шага. Информация из предыдущего скрытого состояния и текущий ввод объединяются с помощью сигмоидальной функции. Значения находятся в диапазоне от 0 до 1. Если значения ближе к 0, то они забываются, а если к 1, то сохраняются.

Во-вторых, необходимо передать предыдущее скрытое состояние и текущий входной элемент в сигмовидную функцию для преобразования значений между 0 и 1, которые означают либо не важность, либо важность, соответственно.

Затем также необходимо передать скрытое состояние и текущее входное состояние в функцию tanh, чтобы сжать значения между -1 и 1. После умножения обоих выходов сигмовидный выход будет решать, какая информация важна. В этот момент существуют условия для умножения состояния ячейки на вектор забывания. Это может привести к падению некоторых значений, если мы умножим их на значения, близкие к нулю. Затем он принимает выходной сигнал от входа и делает точечное дополнение, чтобы обновить состояние ячейки с важной информацией, которая дает нам новое состояние ячейки. Наконец, выходной элемент решает, каким должно быть следующее скрытое состояние. Скрытое состояние содержит информацию из предыдущих входных состояний. Необходимо передать предыдущее скрытое состояние и текущий ввод в сигмовидную функцию, а измененное состояние ячейки – в функцию tanh. После необходимо умножить их, чтобы решить, какую информацию будет хранить скрытое состояние. Выходной элемент является скрытым состоянием.

Сравнение полученных результатов.

Результаты показывают, как модели LSTM и ARIMA работают при прогнозировании цен на биткойн. Каждая модель прогнозирует цену биткойна с точки зрения значения ошибки прогнозирования как для RMSE, так и для MAE [5, 6].

Используя Ft в качестве значения прогноза, At в качестве фактическое значение и n в качестве количества временных шагов, MAE и RMSE [6] могут быть определены следующим образом в уравнениях (1) и (2):

$$MAE = \frac{\sum_{t-1}^{n} |A_t - F_t|}{n}; \qquad (1)$$

$$RMSE = \sqrt{\frac{\sum_{t-1}^{n} (A_t - F_t)^2}{n}}. \qquad (2)$$

RMSE с использованием моделей ARIMA и LSTM составляет 4725 и 361 соответственно, что дает улучшение в среднем на 4364 ошибки, достигнутых с использованием LSTM. В то же время, MAE показывает среднее снижение ошибок на 3867 от ARIMA до модели LSTM. Эти результаты ясно показывают, что модель LSTM улучшила прогнозы ARIMA в среднем на 92% и 94%, согласно RMSE и MAE. Детали результатов можно наблюдать в таблице 1.

Таблица 1 – RMSE и MSE ошибки прогнозирования для моделей ARIMA и LSTM

Предсказание	RMSE		% Редукции	MAE		% Редукции
	ARIMA	LSTM		ARIMA	LSTM	
Цена биткоина	4725	361	92%	4106	239	94%

Для обеих моделей временные ряды были разделены на данные обучения и тестирования - 83,5% и 16,5% соответственно, чтобы сравнить их. В случае модели ARIMA наилучшей подходящей моделью была ARIMA (1,1,1) с использованием минимальных критериев AIC. Эта модель дает правильный прогноз направления, однако прогнозные значения постоянно ниже реальных наблюдений.

Несмотря на то, что эта работа дает ожидаемые результаты, следует отметить некоторые ограничения. Во-первых, размер данных является относительно коротким, и алгоритм LSTM работает лучше в случаях более длинных временных рядов. Во-вторых, цены на биткойны не корректируются сезонно. Временной ряд цен на биткойны представляет сезонность в несколько месяцев, и даже в выходные дни пользователи, как правило, совершают больше транзакций, чем в будние дни. После устранения сезонности временные ряды будут чище, и будут получены более высокие показатели прогнозирования. В-третьих, ежедневные данные создают большую волатильность для ряда, чем можно было бы пожелать. Следовательно, еженедельные данные также могут помочь повысить точность прогнозов.

Чрезвычайная волатильность, связанная с биткойном, представляет определенные проблемы для этой криптовалюты. В данной работе было обращено внимание на потенциальные преимущества использования методов глубокого обучения для прогнозирования цен на криптовалюту. Результаты исследования показывают, что прогнозы LSTM имеют лучшую точность с точки зрения ошибок прогнозирования: RMSE и MAE, они показывают улучшение в среднем на 92% и 94%, соответственно, относительно модели ARIMA.

Литература

1. Namin S. S. Forecasting Economics and Financial Time Series: ARIMA vs. LSTM [Текст] / S. S. Namin, A. S. Namin // Unpublished thesis of KTH. – 2018. – Vol. 1(1). – P. 1-22.
2. Aalborg H.A. What Can Explain the Price, Volatility and Trading Volume of Bitcoin? [Текст] / H.A. Aalborg, P. Molnár, J.E. de Vries // Finance Research Letters. – 2019. – Vol. 29(1). – P. 255-265.
3. Dickey D.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root [Текст] / D.A. Dickey, W.A. Fuller // Journal of the American Statistical Association. – 1979. – Vol. 74(336a). – P. 427-431.
4. A Statistical Analysis of Cryptocurrencies [Текст] / S. Chan, J. Chu, S. Nadarajah, J. Osterrieder // Journal of Risk and Financial Management. – 2017. – Vol. 10(2). – P. 1-23.
5. Anticipating Cryptocurrency Prices Using Machine Learning [Текст] / L. Alessandretti, A. ElBahrawy, L.M. Aiello, A. Baronchelli // Complexity: Hoboken. – 2018. – Vol. 2018(1). – P. 1-16.
6. Chai T. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments Against Avoiding RMSE in the Literature [Текст] / T. Chai, R.R. Draxler // Geoscientific Model Development. – 2014. – Vol. 7(1). – P. 1247-1250.