

П. І. Бідюк¹, Є. В. Гуць¹, В. В. Гавриленко², Н. В. Рудоман²

¹ Національний технічний університет України "КПІ ім. Ігоря Сікорського", Київ, Україна

² Національний транспортний університет, Київ, Україна

ПРОГНОЗУВАННЯ ЦІН АКЦІЙ З ВИКОРИСТАННЯМ РЕКУРЕНТНОЇ НЕЙРОННОЇ МЕРЕЖІ LSTM

Анотація. Дослідження проведено для ознайомлення зі структурою та принципом роботи рекурентної нейронної мережі LSTM (Long short-term memory) та аналізу можливості її використання для прогнозування цін акцій однієї з великих технологічних компаній. В роботі описано теоретичний матеріал, що стосується рекурентних нейронних мереж та мережі LSTM. На прикладі статистичних даних акцій компанії Apple було продемонстровано роботу обраного методу та обчислено оцінки якості прогнозу RMSE, MAE, MAPE і оцінено точність короткострокового прогнозу. Результати дослідження показали, що рекурентні нейронні мережі можна застосовувати для прогнозування часових рядів і при цьому отримувати результат з високою точністю. У подальших дослідженнях будуть запропоновані інші види нейронних мереж та оцінка їх роботи на фінансових даних.

Ключові слова: прогнозування цін, рекурентна нейронна мережа, LSTM, оцінка якості прогнозу, точність короткострокового прогнозу, прогнозування часових рядів.

Вступ

Сьогодні ми спостерігаємо накопичення великих обсягів інформації, які потребують відповідної обробки для прийняття певних рішень. Проблеми математичного моделювання стають все більше актуальними для ефективної організації управління певними суб'єктами господарювання та економічними спільнотами через те, що якість прийнятих рішень значною мірою залежить від якості прогнозування їх наслідків. Саме тому рішення, прийняті сьогодні, повинні ґрунтуватися на достовірних оцінках можливого розвитку процесів та подій в майбутньому [1].

У період технічного прогресу спостерігається ріст впливу великих корпорацій, які впевнено займають лідируючі місця на світовому ринку та мають мільярдні прибутки. Ціноутворення акцій світових гігантів є дуже цікавим для вивчення процесом, який, як правило, є нестационарним, оскільки містить тренд та змінну у часі дисперсію.

Американський сервіс «Yahoo! Finance» надає досить детальні дані по темам, пов'язаним з бізнесом, фінансами та економікою.

Довідкова інформація включає в себе котирування і рейтинги цінних паперів, прес-релізи та фінансові звіти компаній [2].

Для пошуку закономірностей в масивах даних використовуються методи інтелектуального аналізу даних. Сьогодні для побудови моделі широко використовуються нейронні мережі.

Цікавим підвидом є рекурентні нейронні мережі, які містять зворотні зв'язки та можуть зберігати інформацію. Ця властивість дає змогу прогнозувати інформацію, спираючись на контекст, що дало змогу ефективно вирішувати цілий ряд задач: розпізнавання мовлення, моделювання голосу, переклад, розпізнавання зображень, прогнозування часових рядів та інші.

Дане дослідження присвячене побудові і навчанню нейронної мережі LSTM та аналізу її використання для прогнозування цін акцій однієї з великих технологічних компаній.

Постановка задачі. У дослідженні ставляться такі задачі: – аналіз теоретичного матеріалу щодо функціонування рекурентних нейронних мереж; – розроблення структури мережі на мові Python; – демонстрація функціонування нейронної мережі на вибірці фінансових даних; – оцінювання якості прогнозу та аналіз отриманих результатів.

Аналіз літератури. Задачі прогнозування є надзвичайно популярними у самих різноманітних галузях діяльності, включаючи бізнес, економіку, фінанси, кліматологію, управління технологічними системами, технічними об'єктами і т. ін. Серед сучасних досліджень, спрямованих на високоякісне розв'язання задач моделювання і прогнозування необхідно вказати на [3–5]. У цих роботах розв'язуються задачі моделювання і короткострокового прогнозування процесів різної природи на основі регресійного та байєсівського підходів до аналізу даних. При цьому увага концентрується на стаціонарних та нестационарних фінансових та економічних процесах з детермінованими трендами, які з достатньою точністю можна описати детермінованими функціями, наприклад, поліномами, експонентами, гармонічними функціями і т. ін. Також пропонуються алгоритми моделювання і прогнозування вибраних процесів із застосуванням адаптації. Аналіз вказаних та багатьох інших літературних джерел свідчить, що сьогодні існує запит на нові підходи до математичного моделювання і високоякісного прогнозування різноманітних процесів. Тому створення адекватних моделей та оцінювання прогнозів на їх основі, особливо для складних нелінійних нестационарних процесів є високо актуальною задачею.

Невирішені питання. Основними завданнями сучасних підходів до моделювання і прогнозування

процесів різної природи є побудова адекватних моделей складних за структурою нелінійних нестационарних процесів, які забезпечать отримання високоякісних оцінок прогнозів. При цьому необхідно забезпечити врахування можливих невизначеностей даних і структури моделі з метою підвищення якості проміжних та остаточних результатів обчислень.

Мета статті. Метою дослідження є аналіз можливості застосування рекурентних нейронних мереж до моделювання і короткострокового прогнозування нелінійних нестационарних фінансових процесів. Для розв'язання задачі будуть використані рекурентні мережі які містять зворотні зв'язки та можуть зберігати інформацію.

Виклад основного матеріалу

Рекурентні нейронні мережі та мережа LSTM.

Людський мозок працює так, що людина думає послідовно, тобто не починає думати кожний наступний момент спочатку. Читаючи будь-який текст, ми розуміємо зміст та значення кожного слова, відштовхуючись від контексту. Традиційні нейронні мережі не володіють цією властивістю і це їх головний недолік. Уявимо, наприклад, що нам стоїть задача класифікувати події, що відбуваються у фільмі. Незрозуміло, чи могла б зробити це звичайна мережа, не беручи до уваги попередні події у фільмі.

Для вирішення розглянутої проблеми були придумані рекурентні нейронні мережі (Recurrent Neural Networks, RNN), які містять зворотні зв'язки та можуть зберігати інформацію. На рис.1 показано схематична будова таких мереж. Фрагмент А отримує на вхід X_t і повертає h_t . Наявність оберненого зв'язку дає змогу передавати інформацію від одного шару мережі до іншого [6–8].

Для кращого розуміння рекурентну нейронну мережу можна зобразити у розгорнутому, більш звичному вигляді, що не містить циклів (рис. 2). у даному поданні елементи є копією однієї і тієї ж самої мережі, кожна з яких передає інформацію наступній копії.

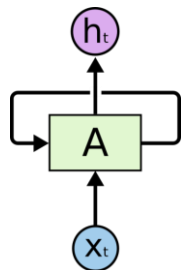


Рис. 1. Рекурентна нейронна мережа

Останні кілька років РНМ користуються великим попитом для вирішення цілого ряду задач: розпізнавання мовлення, моделювання голосу, переклад, розпізнавання зображень, прогнозування часових рядів та інші.

В більшості випадків використовуються модифіковані мережі, як, наприклад, LSTM (Long short-term memory). Про них більш детально описано далі в роботі.

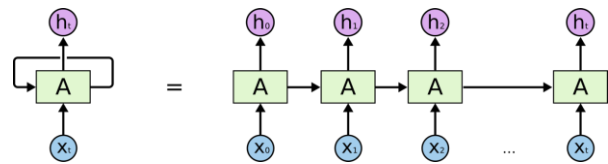


Рис. 2. Рекурентна нейронна мережа в розгорнутому вигляді

Незважаючи на всі переваги РНМ, існує кілька проблем, пов'язаних з їх використанням. Іноді для виконання задачі нам потрібно тільки недавня інформація. Розглянемо, наприклад, мережу, яка прогнозує наступне слово на основі попередніх. У випадку, коли прогнозоване слово очевидне і явно впливає із контексту, відстань між актуальною інформацією і місцем, де вона знадобилась, невелика (рис. 3), тому можна з легкістю навчити РНМ, використовуючи інформацію із минулого.

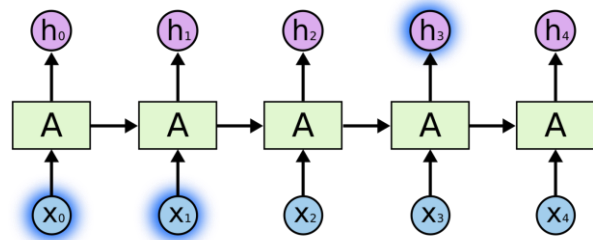


Рис. 3. РНМ з невеликою відстанню між вузлами

Але можуть виникнути випадки, коли нам потрібно більше контексту. Допустимо, нам потрібно передбачити останнє слово у тексті «Я живу в Україні. Я впевнено розмовляю українською». Найближчі слова підказують, що останнім словом буде мова, але щоб визначити яка, нам потрібно дістатись слова «Україні» із більш далекого минулого. Таким чином, різниця між активною інформацією і точкою її застосування може стати дуже великою (рис. 4). В результаті, з мірою росту розмірів мережі втрачається властивість зв'язувати інформацію.

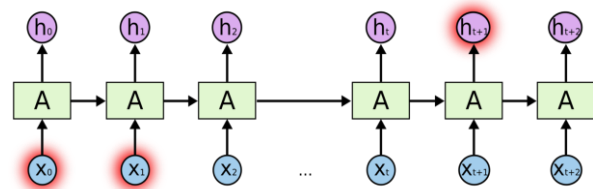


Рис. 4. РНМ з великою відстанню між вузлами

Теоретично, такої проблеми виникати не повинно. Людина може обережно підбирати параметри мережі для розв'язування задачі такого типу. На практиці виявилось, що навчити РНМ таким параметрам неможливо. Це питання досліджували Зеп Хохрайтер (1991) та Іошуа Бенджіо (1994). Вони знайшли доведення того, що це зробити вкрай важко.

Long short-term memory (Довга короткострокова пам'ять), скорочено LSTM – підвид рекурентних нейронних мереж, у якому реалізовано можливість навчання довгострокових зв'язків.

Будь-яка рекурентна нейронна мережа має вигляд ланцюга модулів нейронної мережі, що повторюються. У звичайній РНМ модуль має просту будову і складається, наприклад, з одного шару із функцією активації tanh (рис. 5).

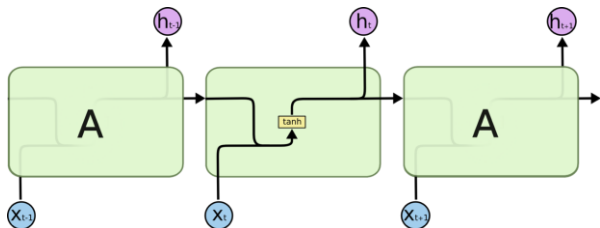


Рис. 5. Ланцюг звичайної РНМ, що складається із одного шару

Структура LSTM також є ланцюгом, але модулі мають більш складну структуру. Замість одного шару, як у звичайній РНМ, модуль LSTM містить чотири шари, які взаємодіють між собою (рис. 6).

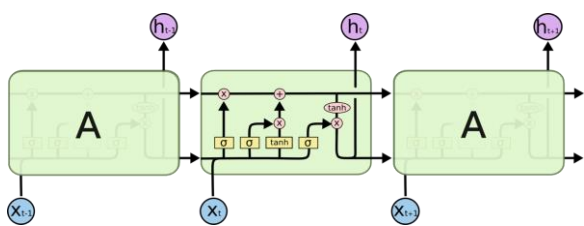


Рис. 6. Ланцюг LSTM, що складається із чотирьох шарів

Основним компонентом LSTM є стан комірки (cell state) – горизонтальна лінія у верхній частині схеми (рис. 7). Вона проходить через всю комірку і бере участь лише у деяких лінійних перетвореннях. Інформація також може не зазнавати жодних перетворень, проходячи через неї.

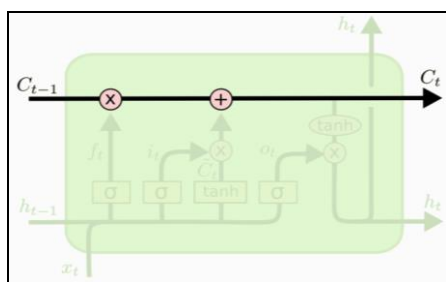


Рис. 7. Стан комірки

Проте LSTM може видаляти інформацію із стану комірки, цей процес регулюється фільтрами (gates) (рис. 8). Фільтри дозволяють пропускати інформацію на підставі деяких умов. Вони складаються з шару сигмоїдальної нейронної мережі і операції поелементного множення.

Сигмоїдальний шар повертає числа від нуля до одиниці, які позначають, яку частку кожного блоку інформації слід пропустити далі по мережі. Нуль в даному випадку означає "не пропускати нічого", одиниця – "пропустити все". У LSTM існують три таких фільтра, що дозволяють захищати і контролювати стан комірки.

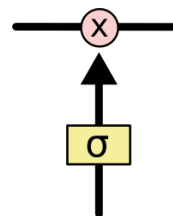


Рис. 8. Фільтр РНМ

Першим кроком роботи мережі слід визначити, яка інформація у стані комірки є непотрібною і її можна викинути. Це рішення приймає сигмоїдальний шар, що називається «шаром фільтра забування» (англ. forget gate level). Його функція полягає в тому, щоб приймати на вхід h_{t-1} та x_t і повертати значення від 0 до 1 для кожного числа із стану комірки C_{t-1} , де 1 означає «повністю зберегти», а 0 – «повністю забути» (рис. 9).

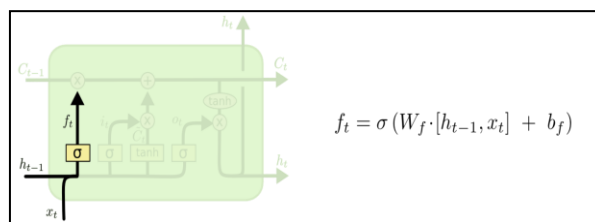


Рис. 9. Фільтр забування

Наступним кроком слід вирішити яка нова інформація буде зберігатись. Даний етап складається із двох частин. Спочатку шар під назвою «шар вхідного фільтра» (англ. input layer gate) визначає які значення слід оновити. Після чого шар tanh буде вектор значень нових кандидатів які додадуться в стан комірки (рис.10).

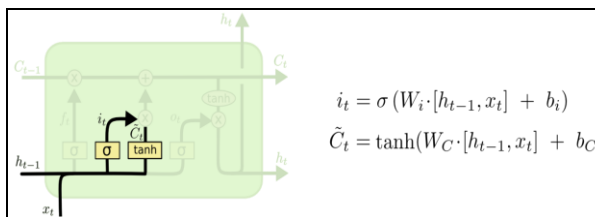


Рис. 10. Шари вхідного фільтра та tanh

Після цього потрібно замінити старе значення стану комірки на нове. Для цього слід помножити старе значення на f_t , щоб забути те, що ми викинули, і додати значення нового кандидата (рис. 10) – це кандидати на нові значення, помножені на число, що показує, наскільки ми хочемо оновити кожне із значень стану (рис. 11).

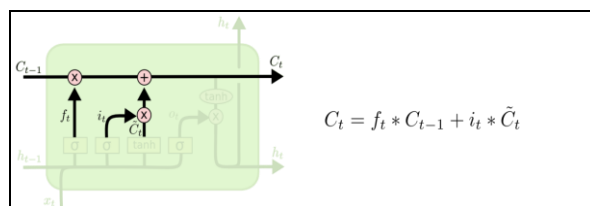


Рис. 11. Крок заміни старого значення комірки на нове

Останнім кроком потрібно вирішити, яку інформацію передати на вихід. Вихідні дані базуються на нашому стані комірки із застосуванням деяких фільтрів. Сигмоїдальний шар вирішує яку інформацію із стану комірки ми будемо виводити. Потім значення стану комірки переходять через шар \tanh , щоб отримати на виході значення в діапазоні від -1 до 1, і перемножуються з вихідними значеннями сигмоїдального шару, що дає змогу виводити лише потрібну інформацію (рис. 12).

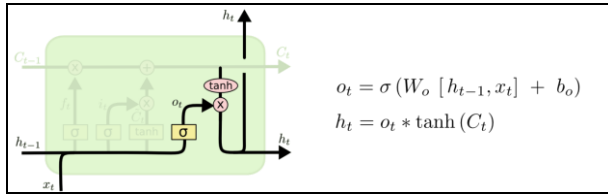


Рис. 12. Формування вихідного значення мережі

Приклад застосування мережі LSTM на фінансових даних. У цьому дослідженні будуть прогнозуватись дані закриття торгів компанії Apple. Обраний датасет містить дані торгів за останні 5 років. Вигляд знайдених даних подано на рис. 13.

Основні статистичні характеристики колонки «Close» подано у табл. 1.

Створимо і навчимо рекурентну нейронну мережу LSTM і перевіримо її роботу на даних закриття акцій компанії. Створена мережа складається з 4 шарів (рис. 14), використовує оптимізатор Adam та середньоквадратичну помилку, як функцію втрат.

Обраний датасет розділено на 2 частини, з яких 80% тренувальна вибірка, 20% – валідаційна. Результати роботи нейронної мережі зображено на рис. 15, де спочатку графіка позначено тренувальну вибірку, надалі – валідаційну, та наприкінці позначені спрогнозовані значення.

Date	High	Low	Open	Close	Volume	Adj Close
2015-12-07	29.965000	29.452499	29.745001	29.570000	128336800.0	27.173285
2015-12-08	29.650000	29.215000	29.379999	29.557501	137238000.0	27.161795
2015-12-09	29.422501	28.770000	29.410000	28.905001	185445600.0	26.562178
2015-12-10	29.235001	28.877501	29.010000	29.042500	116850800.0	26.688534
2015-12-11	28.847500	28.212500	28.797501	28.295000	187544800.0	26.001623
...
2020-12-01	123.470001	120.010002	121.010002	122.720001	128166800.0	122.720001
2020-12-02	123.370003	120.889999	122.019997	123.080002	89004200.0	123.080002
2020-12-03	123.779999	122.209999	123.519997	122.940002	78967600.0	122.940002
2020-12-04	122.860001	121.519997	122.599998	122.250000	78133200.0	122.250000
2020-12-07	124.570000	122.250000	122.309998	123.750000	78742725.0	123.750000

Рис. 13. Вигляд даних для компанії Apple

```

Model: "sequential"
Layer (type)                Output Shape                Param #
=====
lstm (LSTM)                  (None, 60, 50)             10400
lstm_1 (LSTM)                (None, 50)                  20200
dense (Dense)                 (None, 25)                  1275
dense_1 (Dense)               (None, 1)                   26
=====
    
```

Рис. 14. Структура мережі

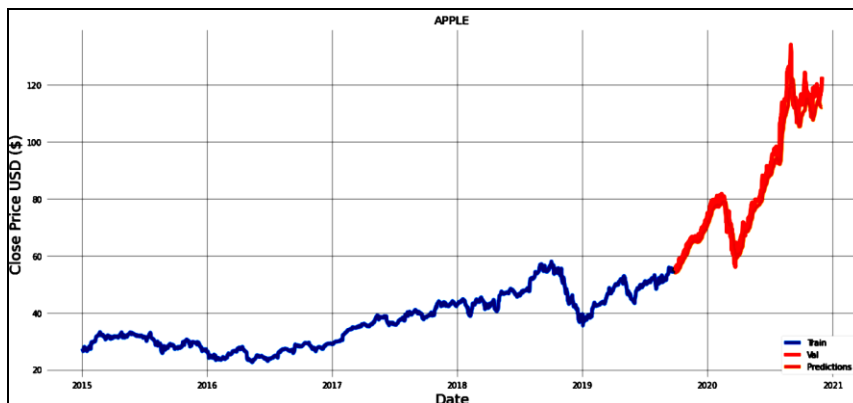


Рис. 15. Результати прогнозування на даних компанії Apple

Таблиця 1 – Статистичні характеристики колонки «Close»

Характеристика			
Максимум	Мінімум	Середнє	Медіана
133.948898	20.994093	48.579072	41.933216

Таблиця 2 – Оцінки якості прогнозу

Оцінка			Точність прогнозу, %
RMSE	MAE	MAPE	
3,916	2,838	3,173	96,827

Аналіз отриманих результатів. Із графіків видно, що мережа досить добре справляється із задачею прогнозування даних.

Значення, отримані шляхом прогнозування, є досить близькими до реальних. Але лише однієї візуальної оцінки недостатньо для виведення висновку щодо роботи моделі та порівняння з іншими моделями, для цього слід поррахувати оцінки якості моделі, такі, як RMSE, MAE, MAPE [6, 9] (табл. 2).

Висновки

У роботі було розглянуто принцип роботи рекурентної нейронної мережі LSTM та перевірено її роботу на значній за обсягом тестовій вибірці. На

прикладі даних акцій компанії Apple було продемонстровано роботу обраного методу.

За результатами отриманого прогнозу обчислено оцінки RMSE, MAE, MAPE.

Отримали точність на рівні 96,8%, що є досить гарним показником.

Можна зробити висновок, що рекурентні нейронні мережі можна застосовувати для прогнозування часових рядів і при цьому отримувати результат з високою точністю.

В подальшому можна розглянути роботу і інших видів нейронних мережі для прогнозування фінансових рядів, а також порівняти їх роботу із звичними регресійними моделями.

СПИСОК ЛІТЕРАТУРИ

1. Бідюк П.І., Романенко В.Д., Тимошук О.Л. Аналіз часових рядів (навчальний посібник). – Київ: Політехніка, 2010. – 317 с.
2. Дані акцій компанії Apple на ресурсі Yahoo Finance. URL: <https://finance.yahoo.com/quote/AAPL>.
3. Robert C.R. Bayesian Choice. – New York: Springer, 2007. – 617 p.
4. Poop S-H. A Practical Guide to Forecasting Financial Market Volatility. – Hoboken, NJ: John Wiley & Sons, Ltd. 2005. – 238 p.
5. Tsay R.S. Analysis of Financial Time Series. – Hoboken, NJ: John Wiley & Sons, Ltd. 2010. – 715 p.
6. Klaus Greff, Rupesh Kumar Srivastava, Jan Koutnuk, Bas R. Steunebrink. LSTM: A Search Space Odyssey. 2015. 10 p.
7. Felix A. Gers, Jorgen Schmidhuber, Fred Cummins. Learning to Forget: Continual Prediction with LSTM. Neural Computation 12 (10), 2000. P. 2451–2471
8. Understanding LSTM Networks URL: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
9. Бідюк П.І., Половцев О.В. Аналіз та моделювання економічних процесів перехідного періоду. – Київ: НТУУ «КПІ», 1999. – 230 с.

Received (Надійшла) 30.06.2021

Accepted for publication (Прийнята до друку) 25.08.2021

Forecasting stock prices using the recurrent neural network LSTM

P. Bidyuk, Y. Huts, V. Gavrilenko, N. Rudoman

Abstract. Modern financial processes exhibit complex nonlinear structure and non-stationary behavior over short and long time intervals. Such behavior can be explained by influence of multiple external stochastic disturbances and complicated nature of the price-forming processes at stock exchange. The traditional approach to modeling and forecasting behavior of such processes is based upon application of regression analysis and respective mathematical models. However, this approach to modeling is not always successful due to complicated structure of the time series under study and limited possibilities of the regression models to describe adequately nonlinear and non-stationary behavior of the processes selected for analysis. That is why it is more appropriate to use as alternative approach neural networks that exhibit very often superior characteristics of modeling and forecasting in comparison to regression models. This study was conducted to get familiar with the structure and principle of operation of the recurrent neural network LSTM (Long short-term memory) and analyze the possibility for its use to forecast the stock prices of one of the major world known technology companies. The theoretical material concerning recurrent neural networks and LSTM network was described in the paper in a volume providing the possibility for understanding basic principles of operation and practical application of the networks. Using as example the Apple stocks statistical data, the functioning of the chosen modeling method was demonstrated and the estimates of the quality of estimated forecasts such as RMSE, MAE, MAPE were calculated; and the accuracy of the forecast was estimated. The results of the simulation achieved showed that recurrent neural networks can be used to predict time series behavior and obtain the results with necessary high accuracy. Further research will suggest application of other types of neural networks, static and dynamic Bayesian networks, and evaluation of their performance on financial data exhibiting nonlinear and non-stationary behavior. Also it is planned to construct specialized decision support system based upon system analysis principles that would contain all necessary functions for further enhancement of modeling and forecasting results.

Keywords: forecasting stock prices, recurrent neural network, LSTM, assessment of forecast quality, accuracy of short-term forecast, time series forecasting.