УДК 519.688

Д.М. ЧАБАНЕНКО, асп. Черкаського національного університету ім. Богдана Хмельницького.

ВИЯВЛЕННЯ КОРОТКОЧАСНОЇ ТА ДОВГОТРИВАЛОЇ ПАМ'ЯТІ ТА ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ МЕТОДАМИ СКЛАДНИХ ЛАНЦЮГІВ МАРКОВА

В статті запропоновано нові методи прогнозування часових рядів, суттєвими відмінностями яких є використання складних ланцюгів Маркова з пам'яттю. Проведено експериментальну роботу та зроблені висновки щодо можливості прогнозування часових рядів запропонованими методами. Основну увагу приділено обґрунтуванню кроків алгоритму та аналізу отриманих результатів прогнозування. Також запропоновано методику оцінювання можливої похибки прогнозів на основі аналізу дисперсії процесу, що моделюється. Наведені результати прогнозування індексів фондових ринків, таких як Dow Jones та інших. Іл.: 1. Бібліогр.: 17 назв.

Ключові слова: прогнозування, часові ряди, складні ланцюги Маркова, індекси фондових ринків.

Постановка проблеми та аналіз літератури. Прогнозування часових рядів – важлива, до кінця не розв'язана на сьогодні проблема. Незважаючи на велику кількість публікацій, все ще залишаються нерозв'язані питання, зв'язані з природою рядів, що аналізуються, видокремленням систематичних рис, наявних в ряді, від випадкових та інші. Насьогодні популярним є поділ трейдерів на фундаменталістів і техніцистів (чартистів) [1]. Ідея фундаменталістів базується на існуванні справедливої ціни акцій, навколо якої коливаються котирування. Прихильники фундаментального аналізу використовують багатовимірний регресійний аналіз [2] для дослідження впливу різноманітних зовнішніх факторів на справедливу ціну, а отже і на спектр наявних цін на ринку. Відхилення від фундаментальної ціни вважається випадковими та не беруться до уваги. На відміну від фундаменталістів, прихильники технічного аналізу базуються на положенні, що історія руху ціни містить усю необхідну інформацію для дослідження системи, отже їх моделі використовують трендові [3] або авторегресійні (AR) моделі [4] для прогнозування майбутніх значень ціни. Варто згадати розширення регресійних і авторегресійних моделей методом групового врахування аргументів та його узагальнення для нечітких даних у роботі [5]. У класичних авторегресійних моделях використовуються лінійні залежності між лаговими змінними та прогнозним значенням. Але така функція також може бути складною та нелінійною [6], побудованою за нейромережевим принципом: так звані NARX та NARMAX-моделі [7], які є нейромережевим аналогами моделей AR та ARMA. На даний час в задачах обробки нелінійних часових рядів найбільше розповсюдження отримали три типи рекурентних нейронних мереж: мережі Вільямса-Зіпсера [8], Елмана [9] та Джордона [10]. Нейронні мережі безперечно є одним з перспективних шляхів у задачах прогнозування, хоча їх практичне використання у бізнесі зв'язане з труднощами, так як необхідна попередня обробка даних для вибору входів системи, специфікація структури моделі, довготривале навчання та наявні помилки в найвідповідальніші моменти розвернення тренду [11]. Тому задача пошуку, конструювання та ідентифікації адекватних прогностичних моделей, стійких до шуму, є актуальною.

Ланцюги Маркова, як прості так і складні, відносяться до моделей випадкових процесів, в яких наявна пам'ять. Марківські процеси займають проміжне положення між випадковими процесами, в яких ймовірності наступного стану не залежать від передісторії, та детермінованими процесами, в яких майбутнє повністю визначається передісторією. Прості ланцюги Маркова широко розглянуті в літературі, наприклад [12], займають значне місце в теорії випадкових процесів. Що стосується складних, або ланцюгів Маркова з пам'яттю, то їх розгляд у літературних джерелах є менш популярним, через складнощі аналізу їх властивостей. Ефективним підходом є зведення складних ланцюгів Маркова до простих за допомогою узагальнення формалізації стану досліджуваної системи [13].

Ідея використання ланцюгів Маркова у задачах прогнозування часових рядів, розглядається у роботах [14], де пропонується імовірнісна модель зв'язку прихованого марківського процесу з явними значеннями величин, які спостерігаються та породжуються вищезгаданим марківським процесом. В роботах пропонується алгоритми оцінювання параметрів марківського процесу (ймовірностей переходів, початкової ймовірності стану системи) та параметрів зв'язку явних станів та прихованих. Алгоритми є ітераційними і базуються на максимізації адекватності моделі, яка будується, даним, на яких відбувається навчання. Суттєвим недоліком вищенаведеного підходу, на нашу думку, є використання простого ланцюгу Маркова, який не враховує передісторію. В роботах [15] пропонується модернізація підходу прихованих марківських моделей для ланцюгів Маркова вищих порядків, які враховують пам'ять ряду та мають більші прогностичні можливості. Але визначення прогнозної ціни як математичного сподівання прогнозного розподілу ймовірностей не дає можливості розрізнити невизначеності та біфуркації.

В роботах [16] пропонується метод прогнозування часових рядів на основі складних ланцюгів Маркова. Однією з переваг запропонованого алгоритму є використання ієрархії часових приростів, яка дозволяє здійснити прогнозування на різних частотних рівнях, та процедури "склеювання", яка дає можливість поєднати прогнози з різною дискретизацією значень, у один ряд прогнозу. Серед недоліків можна назвати надавання великого значення переходам з максимальною ймовірністю, ігноруючи при цьому менш ймовірні прогнозні сценарії, ймовірність яких відрізняється незначно, але характер кривої є суттєво іншим. Такі ситуації практично відсутні при прогнозуванні рядів з високою детермінованістю, таких як регулярні коливання, ряди, породжені моделями детермінованого хаосу та деякі інші. Переважна більшість реальних рядів має невизначеності, які спричинюють похибки прогнозів та, на жаль, не розроблені методи їх оцінювання з ціллю здійснення передпрогнозного аналізу та оцінювання можливості їх прогнозування методами СЛМ та оцінювання можливих похибок прогнозів.

Ціль статті, постановка задачі. Задано часовий ряд $\{y_i\}$ з проміжком дискретизації Δt. Побудувати ряд-продовження вихідного часового ряду, який проходить через найбільш імовірні точки майбутнього продовження ряду, виявлені на основі методу складних ланцюгів Маркова (СЛМ). Розробити методи оцінювання можливих похибок прогнозування.

Основна частина. В роботі [16] запропонована методика прогнозування часових рядів на основі СЛМ.

Ланцюги Маркова першого порядку дають можливість змоделювати досліджуваний процес та визначити ймовірності станів через n кроків

$$P_{t+n}=A_n P_t , \tag{1}$$

де P_{t+n} – ймовірності станів в момент часу $t+n$; P_t – початковий розподіл (вектор) ймовірностей в момент часу t; A_n – матриця ймовірностей переходів між станами.

У задачах прогнозування часових рядів важливим питанням залишається кодування станів у дискретному ланцюгу Маркова. Зв'язування станів найбільш доцільно здійснити з абсолютними (або відносними) змінами вхідного ряду:

$$R_t = (p_t - p_{t-\Delta t}) \text{ або } R_t = (p_t - p_{t-\Delta t})/p_t , \tag{2}$$

де R_t – прибутковість ряду в дискретний момент часу; p_t – ціна в момент часу t; $p_{t-\Delta t}$ – попереднє значення ціни в момент часу $t - \Delta t$.

Відомо, що математичне сподівання ряду R_t дорівнює нулю, а дисперсія має специфічну властивість, яка називається кластеризація волатильності. Дисперсія раптово зростає через певні неочікувані різкі зміни досліджуваного ряду, що можливо пов'язане з кризовими явищами.

Розподіл щільності ймовірностей прибутковості дає змогу здійснити класифікацію станів системи, згідно зі значеннями прибутковостей. Визначення меж кожного класу дає можливість задавати оператор перетворення дійсних значень прибутковостей у натуральні номери дискретних станів

$$R_t \rightarrow s_t \tag{3}$$

Для дискретної послідовності станів $\{S_t\}$ $t = 1, \ldots, n$ оцінюються ймовірності переходів між станами (коефіцієнти a_{ij} матриці перехідних ймовірностей A):

$$a_{ij} = P(s_{t-1} = i,\ s_t = j), \quad i, j = 1, \ldots, s_{max}. \tag{4}$$

Дана матриця ймовірностей і буде використовуватись в прогнозуванні. В роботі [17] нами запропонована технологія, згідно якої на кожному наступному кроці вибирається стан з максимальною ймовірністю, який у подальшому прогнозуванні вважається тим, що уже відбувся. Таким чином ігноруються стани, які мають ймовірності, нижчі за максимальну. Для подолання можливих помилок, які з цим пов'язані, пропонується система правил, які дозволяють уточнити майбутній стан, базуючись на кластер ній організації функції розподілу щільності ймовірності.

При використанні системи станів (3), пов'язаних з прибутковостями, пряме використання формули (1) не є корисним, оскільки прогнозується значення ряду, а не стан ланцюга Маркова. Очевидно, що кожна послідовність станів $\{S_i\}$ довжиною n відповідає певному рівню величини, що прогнозується $y(t_{last} + n\Delta t)$ через n кроків з вибраною дискретизацією Δt. Ймовірність певної послідовності станів (шляху до точки $y(t_{last} + n\Delta t)$) можна визначити за формулою (1), таким чином, для усіх можливих послідовностей $\{S_i\}$ $i = 1, \ldots, n$ та відповідних шляхів значень величини, що прогнозується, можна поставити у відповідність ймовірність того чи іншого сценарію прогнозу.

В роботах [16, 17] нами пропонувалось розглядати лише сценарії з найбільшою ймовірністю, ігноруючи при цьому менш ймовірні. Також, приймаючи на кожному кроці сценарій (наступний стан) з максимальною ймовірністю, ми виходили з припущення, що кроки з меншою ймовірністю для моменту часу t також не матимуть пріоритет в наступні моменти часу $t+i$. Для перевірки вищезазначених положень можна порівняти прогноз за технологією СЛМ та довірчі інтервали відхилення інтервального прогнозу чи за технологією СЛМ від діапазону запропонованого розширеного алгоритму.

Основна мета даного дослідження полягає у оцінці достовірності прогнозів, отриманих методикою [16, 17] та виявлення критеріїв застосування даної методики до рядів того чи іншого походження.

Отже, ймовірність сценарію прогнозу $S = \{s_t\}$ $i = 1, ..., n$ визначається за формулою:

$$P_S = \prod_{t=1}^{n_{progn}} P(s_t) = \prod_{t=1}^{n_{progn}} P(s_{t-1}) \cdot P(s_t \mid s_{t-1}), \tag{5}$$

де s_t – узагальнений стан в момент t; $P(s_t)$ – ймовірність стану s_t в момент t, яку можна визначити з (1) як відповідну компоненту вектору P_t.

Оскільки під сценарієм ми розуміємо певну траєкторію динаміки показника, що прогнозується, то обернене перетворення станів, зв'язаних з приростом ряду, у прогнозні значення ряду може бути здійснене:

$$y_t = y_{t-1} + R_t \tag{6}$$

для абсолютних та відносних приростів відповідно, де R_t – спрогнозовані прирости, які відповідають прогнозній послідовності станів ланцюга Маркова.

Різні траєкторії можуть проходити через спільні точки. Можна обчислити ймовірність того, що через n кроків показник матиме певне значення (зупиниться в певній точці). Ймовірності значень на основі ймовірностей сценаріїв можна обчислити так:

$$P(y_t = y) = \frac{\sum_{i=1}^{Ns} - sign\left(\left|y_t^i - y\right| - \delta\right)}{Ns} = F(t, y). \quad (7)$$

Таким чином, для кожного дискретного моменту часу t маємо функцію розподілу ймовірностей $F(y)$. Узагальнюючи отриманий розподіл, переходимо до математичного сподівання та стандартного відхилення прогнозу, а також виділити емпіричну моду та квантилі.

Експериментальні результати. Апробація алгоритму була виконана на рядах фінансово-економічної динаміки, зокрема на фондових індексах Dow Jones, S&P 500 та інших. Прогнозування здійснювалось з обчисленням прибутковостей при різних проміжках дискретизації Δt. На рис. 1 зображена множина найбільш ймовірних сценаріїв динаміки індексу S&P 500 для дискретизації Δt = 1. Прогноз здійснювався на 8 днів (точки 50 – 58 на графіках).

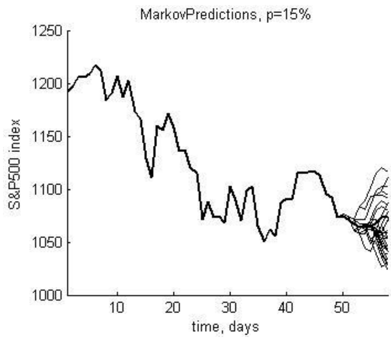

Рис. 1. Прогнозні сценарії індексу S&P 500.

На графіку рис. 1 показані прогнозні сценарії, які мають найбільшу ймовірність. Товстою суцільною лінією зображений прогноз за технологією СЛМ, який відповідає одному найбільш ймовірному сценарію.

Дані результати свідчать про високий рівень невизначеності індексу фондового ринку S&P 500, хоча з великою впевненістю прогнози показують падіння протягом перших двох днів. Для порівняння прогнозувались детерміновані сигнали, такі як синусоїди, як приклад періодичного сигналу, та динамічного ряду моделі "хижак-жертва".

Для рядів з високою детермінованістю алгоритм породжує значно менше прогнозних сценаріїв, та їх траєкторії є близькими одна до одної.

Висновки. В даній роботі запропоновано метод прогнозування часових рядів на основі виявлення пам'яті в часовому ряді. Головною перевагою запропонованого методу використання складних ланцюгів Маркова з пам'яттю, на відміну від простих ланцюгів Маркова. Також варто зазначити більш швидке, порівняно з нейромережевими методами, навчання моделі, а також можливість оцінки можливої похибки прогнозу за допомогою оцінки невизначеності даних та моделі на основі оцінювання дисперсії прогнозу. Результати показують, що прогноз за технологією СЛМ знаходиться у межах середньоквадратичних відхилень розподілу та близько до модальних значень, що означає, що запропонований метод дає оцінку точності методу СЛМ.

Список літератури: **1.** *Samanidou E.* Agent-based Models of Financial Markets [електронний ресурс] / *E. Samanidou, E. Zschischang, D. Stauffer, T. Lux* // Arxiv:physics/0701140v1. – 2007. – Режим доступу: http://arxiv.org/ **2.** *Четыркин Е.М.* Статистические методы прогнозирования / *Е.М. Четыркин*. – М.: Статистика, 1977. – 200 с. **3.** *Yaffee Robert A.* Introduction to time series analysis and forecasting / *Robert A. Yaffee, Money McGee* // New York: Academic press inc. – 2000. – 528 p. **4.** *Бокс Дж.* Анализ временных рядов. Прогноз и управление / *Дж. Бокс, Т. Дженкинс*. – М.: Мир, 1974. – 242 с. **5.** *Зайченко Ю.П.* Нечеткие модели и методы в интеллектуальных системах: учеб. пособие для иностр. студ. вузов, направления "Компьютерные науки" / *Ю.П. Зайченко*; [М.З. Згуровский (общ.ред.)]. – К.: Слово, 2008. – 344 с. **6.** *Петерс Э.* Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка / *Э.Петерс*. – М.: Мир, 2000. – 333 с. **7.** *Mandic D.P.* Recurrent Neural Networks for Prediction / *D.P. Mandic, J.A. Chambers*. – Chichester: John Wiley&Sons, 2001. – 285 p. **8.** *Williams R.J.* A Learning Algorithm for Continually Running Fully Recurrent Neural Networks / *R.J. Williams, D. Zipser* // NeuralComputation. – 1989. – № l. – P. 270-280. **9.** *Elman J.L.* Finding structure in time / *J.L. Elman* // Cognitive Science. – 1990. – № 14. – P. 179-211. **10.** *Jordan M.* Constrained supervised learning / *M. Jordan* // Journal of Mathematical Psychology. – 1992. – № 36. – P. 396-452. **11.** *Біліловець О.С.* Нейромережеве прогнозування в сфері електронної комерції / *О.С. Біліловець* // Економіко-математичне моделювання соціально-економічних систем Збірник наукових праць МННЦ ІТіС, Вип. 13. – К.: 2008. **12.** *Романовский В.И.* Дискретные цепи Маркова / *В.И. Романовский*. – М.-Л.: Гостехиздат, 1949. **13.** *Сапцин В.М.* Опыт применения генетически сложных цепей Маркова для нейросетевой технологии прогнозирования / *В.М. Сапцин* // Вісник Криворізького економічного інституту КНЕУ. – Кривий Ріг, КЕІ КНЕУ, 2009. – Вип. 2 (18). – С. 56-66. **14.** *Zhang Y.* Prediction of Financial Time Series with Hidden Markov Models / *Y. Zhang* // School of Computer Science. – Vol. Master of Applied Science: Simon Fraser University, 2004. – 102 p., **15.** *Rabiner L.R.* A tutorial on hidden markov models and selected applications in speech recognition / *L.R. Rabiner* // In Proceedings of IEEE. – 1989. – Vol. 77. – P. 257-286. **16.** *Чабаненко Д.М.* Алгоритм прогнозування фінансових часових рядів на основі складних ланцюгів Маркова / *Д.М. Чабаненко* // Вісник Черкаського університету. – 2010. – Вип 173. – С. 90-102. **17.** *Soloviev V.* Prediction of financial time series with the technology of high-order Markov chains [електронний ресурс] / *V. Soloviev, V. Saptsin, D. Chabanenko* // Working Group on Physics of Socio-economic Systems (AGSOE). – Drezden, 2009. – Режим доступу: http://www.dpg-verhandlungen.de/2009/dresden/agsoe.pdf.

Стаття представлена д.ф.-м.н., проф. Соловйовим В.М.

УДК 519.688

Выявление короткой и длинной памяти и прогнозивание временных рядов методами сложных цепей Маркова / Чабаненко Д.Н. // Вестник НТУ "ХПИ". Тематический выпуск: Информатика и моделювание. – Харьков: НТУ "ХПИ". – 2010. – № 31. – С. 184 – 190.

В статье предложены новые методы прогнозирования временных рядов, существенным отличием которых является использование сложных цепей Маркова с памятью. Проведена экспериментальная работа и сделаны выводы о возможности прогнозирования временных рядов предложенными методами. Основное внимание уделяется обоснованию шагов алгоритма и анализу полученных результатов прогнозирования. Также предложено методику оценки возможной погрешности прогнозов на основе анализа дисперсии моделируемого процесса. Приведены результаты прогнозирования индексов фондовых рынков, таких як Dow Jones, S&P 500 и других. Ил.: 1. Библиогр.: 17 назв.

Ключевые слова: прогнозирование, временные ряды, сложные цепи Маркова, финансовые рынки, индексы фондовых рынков.

UDC 519.688

Identification of short- and long-range memory and time series forecasting with the methods of complex Markov chains / Chabanenko D.M. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2010. – №. 31. – P. 184 – 190.

The article deals with new time series forecasting methods, the main distinctions of which is the usage of memory-based complex Markov chains. The experimental work was done and the conclusions are made about possibilities of time series forecasting with the proposed methods. The main attention is paid to the algorithm explanation and the forecasting results analysis. Also the methods of prediction errors evaluation is proposed, based on the analysis of predicted signal's variance. The forecasting results for financial market's indices are reviewed, such as Dow Jones, S&P 500 and others. Figs: 1. Refs: 17 titles.

Key words: forecasting, time series, complex Markov chains, financial markets, indices for financial market's.

Поступила в редакцію 30.05.2010