

УДК 681.3, 004.6

Степаненко А.В., студент магістратури,
Хлевний А.О., асистент кафедри технологій управління, к.т.н.
Київський національний університет ім. Тараса Шевченка

ВИКОРИСТАННЯ АЛГОРИТМУ ЧАСОВИХ РЯДІВ MICROSOFT ДЛЯ ПРОГНОЗУВАННЯ ПРОДАЖІВ

Степаненко А.В., Хлевний А.О. Використання алгоритму часових рядів Microsoft для прогнозування продажів. У статті розглядається прогнозування продажів продукції за відомою історією продажів та без урахування додаткових факторів алгоритмом часових рядів Microsoft. У роботі аналізуються переваги та недоліки зазначеного алгоритму, робляться висновки щодо доцільності його застосування у прогнозуванні. Практична реалізація алгоритму часових рядів Microsoft довела його ефективність при прогнозуванні продажів на основі історичних даних.

Ключові слова: прогнозування, алгоритм часових рядів Microsoft, ARIMA, ARTXP

Степаненко А.В. Использование алгоритма временных рядов Microsoft для прогнозирования продаж. В статье рассматривается прогнозирование продаж продукции по известной истории продаж и без учета дополнительных факторов алгоритмом временных рядов Microsoft. В работе анализируются преимущества и недостатки указанного алгоритма, делаются выводы о целесообразности его применения в прогнозировании. Практическая реализация алгоритма временных рядов Microsoft доказала его эффективность при прогнозировании продаж на основе исторических данных.

Ключевые слова: прогнозирование, алгоритм временных рядов Microsoft, ARIMA, ARTXP

Stepanenko A.V. Microsoft Time Series algorithm application for sales forecasting. The article deals with the forecasting of product sales by known sales history without additional factors with the help of Microsoft Time Series algorithm. The study is concentrated on the analysis of the advantages and drawbacks of the given algorithm. Authors give the conclusions about the expediency of Microsoft Time Series algorithm application in product sales forecasting. The outcomes of practical Microsoft Time Series algorithm leveraging proved its effectiveness in forecasting sales based on historical data.

Key words: forecasting, ARIMA, ARTXP, Microsoft Time Series algorithm

Постановка наукової проблеми. На сучасному етапі можливість прогнозувати майбутнє, спираючись лише на попередні дані, веде до стратегічних переваг, та може бути ключем до успіху організації. Прогнозування алгоритмом часових рядів Microsoft – це прогнозування тимчасової змінної, що перетворюється на важливий інструмент у вирішенні проблем, оскільки дозволяє моделювати складні системи, де мета полягає в прогнозуванні поведінки системи, а не в роботі системи. Дійсно, в останні кілька десятиліть все більше уваги приділяється цій проблемі. Розробки в таких галузях як статистика та комп'ютерні науки стали підґрунтям для виникнення методів часових рядів Microsoft, які замінили застарілі методи, основані в першу чергу на інтуїції.

Метою даної статті є проаналізувати переваги та недоліки використання алгоритму часових рядів Microsoft для прогнозування продажів продукції за відомою історією продажів без впливу інших факторів та показати його практичне застосування. Такий вид завдання найчастіше зустрічається в роботі компаній, коли ми не можемо використовувати факторний аналіз і повинні спиратися виключно на історичні дані. Програмні рішення будуються для того, щоб впоратися з величезною кількістю даних.

Аналіз досліджень. У роботі Чугуєвої І.О.[1] спостерігається, що загалом всі прогностичні методи можна розділити на дві основні групи: статистичні та структурні методи відповідно. До першої групи належать прогностичні моделі, які використовують статистику для прогнозування результатів [3] - кінцева модель представлена аналітичною формулою. Друга група складається з покрокових методів, де кінцевий алгоритм виглядає як певний набір етапів. У наших дослідженнях статистичні підходи представлені ARIMA (реалізовано в Microsoft Time Series). Microsoft SQL Server [4] забезпечує регресійні алгоритми, оптимізовані для прогнозування безперервних значень.

У даній роботі ми попередньо аналізуємо підхід до вирішення завдання прогнозування продажів. Актуальним завданням є оцінка ефективності маркетингових кампаній з метою реклами та просування продукту на ринок. Ці заходи є засобом збільшення продажів. Отже, якщо ми можемо передбачити продажі якогось продукту заздалегідь, а потім порівняти його з фактичними даними, ми можемо зафіксувати прибутковість проведених маркетингових кампаній.

Виклад основного матеріалу й обґрунтування отриманих результатів. Методи прогнозування дають прогноз на основі статистичних взаємозв'язків між вхідними стовпцями в наборі даних. Головною

ідеєю є правильне тлумачення поточних даних для отримання об'єктивних законів. Вхідні дані представлені набором даних. У проведеній експериментальній роботі нами розглядаються три основні категорії: назва продукту, продажі та розглянутий період. Щоб полегшити сприйняття моделі прогнозування, ми досліджуємо один продукт. Методи прогнозування застосовуються до набору даних для отримання ймовірного прогнозу на майбутній період.

Алгоритм часових рядів Microsoft

Алгоритм часових рядів Microsoft (Microsoft Time Series) забезпечує алгоритми регресії, які оптимізовані для прогнозування безперервних значень, таких як продаж продуктів з часом. В той час, як інші алгоритми Microsoft, такі як дерева рішень, вимагають додаткових стовпців нової інформації, як вхідні дані для прогнозування тенденції, а модель часових рядів цього не вимагає. Модель часових рядів може передбачати тенденції, що ґрунтуються тільки на оригінальному наборі даних, який використовується для створення моделі. Можна також додавати нові дані до моделі, коли ви робите прогноз і автоматично включаєте нові дані до аналізу тенденцій [4].

Алгоритм часових рядів Microsoft містить два окремих алгоритми для аналізу:

1) алгоритм ARTXP (для короткострокового прогнозу, для регресії дерев з перехресним прогнозом) [2], який був введений в SQL Server 2005, оптимізований для прогнозування наступного ймовірного значення в серії;

2) алгоритм ARIMA був доданий в SQL Server 2008 для підвищення точності довгострокового прогнозування.

За замовчуванням Analysis Services використовує кожен алгоритм окремо для навчання моделі, а потім поєднує результати, щоб отримати найкращий прогноз для змінної кількості прогнозів. Також можна обрати лише один з алгоритмів на основі ваших даних та вимог до прогнозування [4].

Microsoft Research розробила оригінальний алгоритм ARTXP, який використовувався в SQL Server 2005, засновуючи реалізацію на алгоритмі дерева рішень Microsoft. Таким чином, алгоритм ARTXP може бути описаний як модель авторегресії дерева для представлення даних періодичних часових рядів. Цей алгоритм пов'язує змінну кількість минулих елементів з кожним поточним елементом, який передбачається [4].

Алгоритм ARIMA був доданий до алгоритму часових рядів Microsoft в SQL Server 2008 для поліпшення довгострокового прогнозування. Це реалізація процесу обчислення авторегресійних інтегрованих ковзних середніх (рухомих середніх). Методологія ARIMA дає змогу визначати залежності в спостереженнях, що виконуються послідовно в часі, і може включати випадкові потрясіння як частина моделі. Метод ARIMA також підтримує мультиплікативну сезонність [4].

За замовчуванням, алгоритм часових рядів Microsoft використовує обидва методи, ARTXP і ARIMA, і поєднує результати для поліпшення точності прогнозування. Якщо необхідно використати тільки конкретний метод, можна встановити параметри алгоритму для використання тільки ARTXP або тільки ARIMA, або керувати комбінуванням результатів алгоритмів [4].

Переваги та недоліки алгоритму часових рядів Microsoft

Беручи до уваги, що цей алгоритм є «закритим джерелом» і поставляється «out of the box», він готовий до використання і ми можемо швидко розпочати його застосування до вихідних даних. Але обмеження полягає в тому, що ми можемо маніпулювати тільки дозволеними значеннями параметрів, і це визначає його плюси та мінуси. Як конкретна реалізація концепції методу дерев рішень корпорації Майкрософт на високому рівні, цей підхід успадковує також переваги та недоліки методу дерев рішень.

Якщо проаналізувати дві вбудовані моделі всередині цього підходу, а саме ARIMA і ARTXP, то переваги і недоліки можуть бути представлені в наступному вигляді:

Плюси моделі ARIMA:

1. Добре для довгострокового прогнозування.
2. Підтримує мультиплікативну сезонність.
3. Поєднує авторегресію.
4. Відома модель, перевірена з плином часу.

Мінуси моделі ARIMA:

1. Неможливо запуснути моделювання політики.
2. Традиційні методи ідентифікації моделі суб'єктивні.

3. Не підтримує крос-прогноз.

Плюси моделі ARTXP:

1. Підтримує крос-прогноз.
2. Дає відповідні прогнози в найближчому майбутньому.
3. Краще дивитися на різні змінні, що впливають один на одного з плином часу.

Мінуси моделі ARTXP:

1. Менш стабільний.
2. Не підходить для довгострокового прогнозування.
3. Висока складність розрахунків для довгострокових прогнозів (вимагає набагато більше часу).

У нашому дослідженні ми використали короткостроковий та довгостроковий прогнози. *Короткостроковий прогноз* являє собою прогноз на кілька кроків вперед, тобто здійснюється побудова прогнозу не більше ніж на 3% від обсягу спостережень або на 1-3 кроки вперед. Для побудови даного прогнозу ми використали алгоритм ARTXP. *Довгостроковий прогноз* - це прогноз більш ніж на 5% від обсягу спостережень, або на 7 і більше кроків вперед. Також під цим типом прогнозу розуміють прогноз на один або половину сезонного циклу. Для побудови даного прогнозу ми використали алгоритм ARIMA.

Після застосування алгоритму ми отримали наступні результати (рис.1).

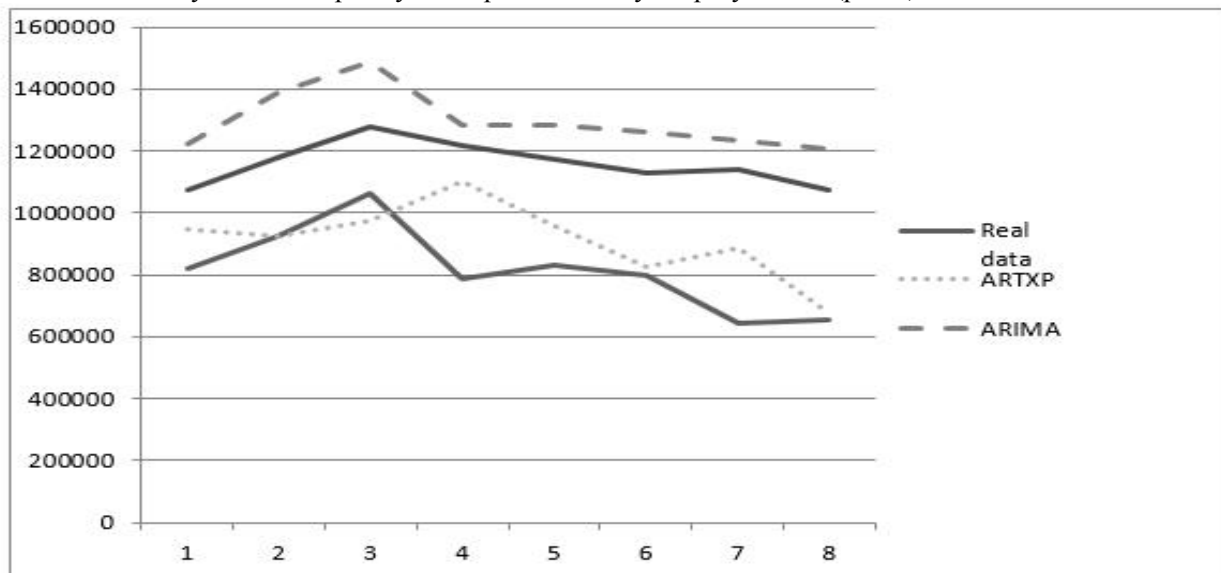


Рис. 1. Результати роботи алгоритмів для першого продукту

На першому сірою лінією відмічено графік продажів, крапами – прогноз з використанням алгоритму ARTXP, пунктирною лінією – прогноз з використанням алгоритму ARIMA, чорною лінією – прогноз з використанням суміші алгоритмів. Вісь ординат відображає суму продажів товарів у тис. гривень, а вісь абсцис відображає місяці. Графік (рис.1) демонструє досить дивний результат, оскільки ARTXP показав найкращий результат на усьому проміжку. Проте цей метод повинен показувати найліпший результат на короткострокових прогнозах. Пояснити це ми не можемо, оскільки ми не володіємо додатковою інформацією про продукти в базі.

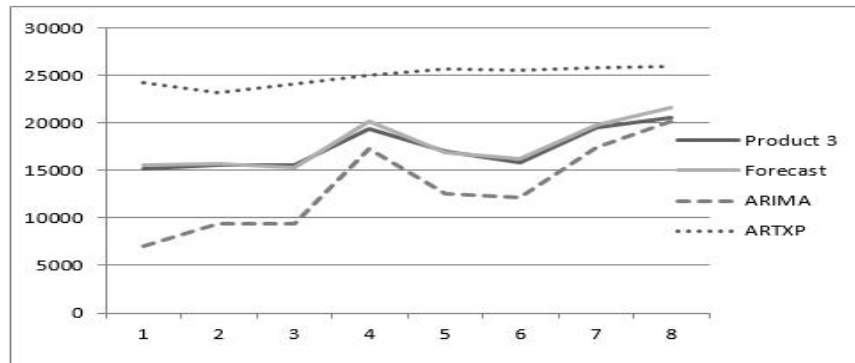


Рис. 2. Результат роботи алгоритмів для другого продукту

На графіку рис.2 видно, що алгоритм ARIMA підвищує свою точність при довгостроковому прогнозуванні, тоді як ARTXP показує найліпший результат на короткостроковому прогнозі. Проте змішаний алгоритм знову показує кращі результати на усьому проміжку. Такий самий результат показало прогнозування щодо інших продуктів. Мінімальне відхилення прогнозу від реальних показників становить 2%, а максимальне – 20%. В середньому відхилення значень в прогнозі від реальних даних становить 10-12%.

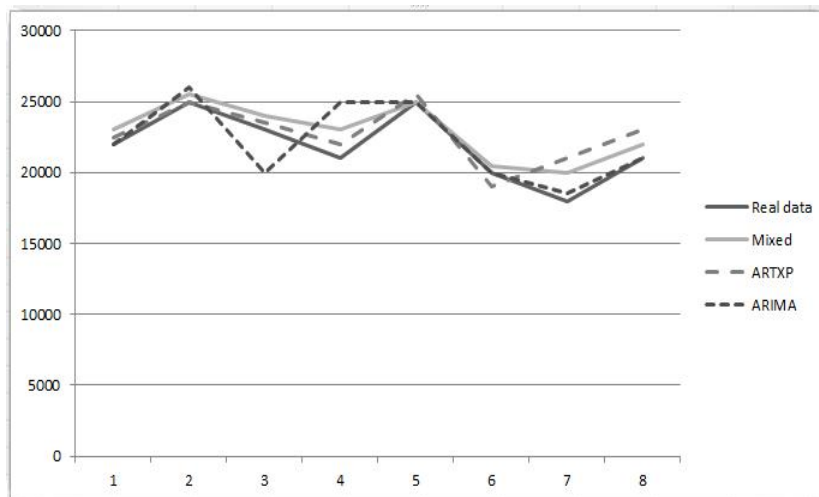


Рис.3 Результат роботи алгоритмів для третього продукту

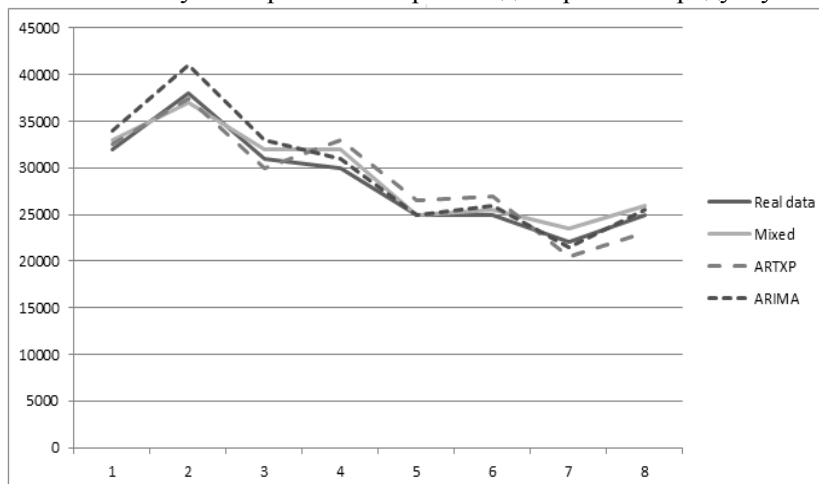


Рис.4 Результат роботи алгоритмів для четвертого продукту

Для графіків на рис.3 та рис.4 було використано дані продажів за найбільший проміжок часу, і результат аналізу підтвердив, що використання ARIMA доцільне для довгострокових прогнозів, оскільки на довгому проміжку часу він показує найбільшу точність, а ARTXP показує найліпшу точність на короткому проміжку часу. Використання обох методів одночасно показує середній результат і може бути використаним коли немає чіткого розуміння який саме прогноз необхідний, короткостроковий чи довгостроковий.

Висновки та перспективи подальшого дослідження. Практична реалізація методу часових рядів та його окремих алгоритмів довела, що алгоритм ARIMA дає найкращі результати при довгострокових прогнозуваннях. Алгоритм ARTXP дає найкращі результати при короткотривалих прогнозах, а змішаний алгоритм демонструє точність на усьому проміжку прогнозування. Таким чином практична реалізація алгоритму часових рядів Microsoft довела його ефективність при прогнозуванні продажів на основі історичних даних. В подальшій роботі спробуємо проаналізувати та порівняти прогнозування продажів методом дерев рішень та генетичним алгоритмом.

1. Чугуева И.А. Модель прогнозирования временных рядов по выборке максимального подобия. Дисертация на соискание ученой степени кандидата технических наук.: 05.13.18 / И.А. Чугуева. – М., 2012. – 153 с.
2. Алгоритм временных рядов (Майкрософт). Режим доступа: <http://docs.microsoft.com/ru-ru/sql/analysis-services/data-mining/microsoft-time-series-algorithm?view=sql-server-2017>.
3. Geisser S. Predictive Inference: An Introduction. Monographs on Statistics and Applied Probability / S. Geisser. – NY: Chapman & Hall, 1993. – 265 p.
4. Microsoft Developer Network, Available from: <www.msdn.com>.
5. TreePlan Software. Introduction to Decision Trees. Available from: <<http://treeplan.com/chapters/introduction-to-decision-trees.pdf>>.
6. Pandey A.K. Early Software Reliability Prediction: A Fuzzy Logic Approach / A.K. Pandey, N.K. Goyal. – New Dehli: Springer India, 2013. – 153 p.
7. Xu G. Applied Data Mining / Guandong Xu, Yu Zong, Zhenglu Yang.–CRC Press, 2013.–284 p.
8. Gordon L. Using Classification and Regression Trees (CART) in SAS Enterprise Miner For Applications in Public Health / Leonard Gordon // SAS Global Forum 2013. Paper 089-2013. – 2013. – 8 p.
9. Loh W.-Y. Classification And Regression Trees / W.-Y. Loh // WIREs Data Mining and Knowledge Discovery.–2011.–Vol.1.–pp.14-23.
10. Lytvynenko T.I. Problem of data analysis and forecasting using decision trees method / T.I. Lytvynenko // Problems in programming. – 2016. – Vol.2-3. – pp.220-226.
11. Wu X. Top 10 Algorithms In Data Mining / Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, Dan Steinberg // Knowledge and Information Systems. – 2008. – Vol.14, Iss.1. – pp.1-37.
12. Skiena S. The Algorithm Design Manual / Steven S. Skiena. – London: Springer-Verlag, 2008. – 730 p.
13. Bright Hub Project Management. Disadvantages to Using Decision Trees. Available from: <<http://www.brighthubpm.com/project-planning/106005-disadvantages-to-using-decision-trees/>>.