КОМП'ЮТЕРНІ НАУКИ

УДК 5.51.517 https://doi.org/10.31713/vt4202313

Кінда В. В., аспірант (Рівненський державний гуманітарний університет, м. Рівне)

ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ МЕТОДАМИ НЕЧІТКОЇ ЛОГІКИ ШЛЯХОМ ЗАДАННЯ УНІВЕРСАЛЬНОЇ МНОЖИНИ СКІНЧЕННИМИ РІЗНИЦЯМИ ВИЩИХ ПОРЯДКІВ

Проведено дослідження прогнозних моделей, що базуються на застосування нечіткої логіки для часових рядів. Наведено модифікований варіант задання універсальної множини та її розбиття для проведення подальшої фазифікації. Реалізовано відповідний метод у середовищі розробки високого рівня Python у вигляді програмного продукту та окремий модуль для перевірки якості його роботи. Здійснено апробацію роботи програмного продукту, проаналізовано одержані результати та розглянуто напрямки подальших перспективних досліджень.

Ключові слова: **часовий ряд; нечітка множина; скінченні різниці; екстраполяція; прогноз.**

Актуальність теми. Аналіз часових рядів являється важливим інструментом у різних сферах діяльності – бізнесі, науці, економіці і тд. При цьому, методи аналізу даних допомагають краще зрозуміти природу процесів, виявити приховані закономірності, тренд та сезонності. На сьогодні, для прогнозу часових рядів часто використовуються методи машинного навчання — регресійні моделі, нейронні мережі різних архітектур. Проте суттєвими викликами вищезгаданих методів являється відсутність достатньої кількості даних для навчання моделей, що в свою чергу, доволі сильно може впливати на точність прогнозу та їх обчислювальну ресурсоємність. З іншого боку, перевагою методів пірамідальної екстраполяції, прогнозування часових рядів на основі нечітких множин дає змогу працювати із малими вибірками та швидко робити прогноз, що дозволяє використовувати дані моделі як одні із базових для використання в фінансових часових рядах при роботі на валютних біржах, криптомаркетах, фінансових маркетах. Вищезгадані переваги обумовлюють важливість та перспективність подальших досліджень в даній області.

Аналіз досліджень. Дослідження Заде [1] в області нечіткої логіки набули широкого застосування в методах оптимізації, теорії прийняття рішень, теорії керування, штучного інтелекту. Одним із напрямів подальших робіт в даній області стали предиктивні моделі засобами нечіткої логіки. До них можна віднести класи нечітких нейронних мереж, МГУА – метод групового врахування аргументів, алгоритмів та методів прогнозування засобами нечіткої логіки. Згодом в роботі [2] була визначена та формалізована теорія застосування нечітких множин для часових рядів. Описано алгоритм побудови точкового прогнозу. Наступні роботи [3–5] здебільшого пов'язані із різними варіантами вибору функцій належності та задання універсальної множини.

Викладення основного матеріалу. Пропонується застосовувати різниці першого роду для побудови універсальної множини, що в свою чергу дає якісний числовий прогноз. В даній роботі запропонована модифікація методу Стівенсона Портера шляхом використання скінченних різниць вищих порядків. Фізичний зміст першого порядку відповідає швидкості, а другого – прискоренню. Тому саме їх дослідження та порівняння із вже існуючими методами і модифікаціями становить суттєвий інтерес та значимість.

Далі алгоритм складається із наступних кроків.

Крок 1. Нехай дано часовий ряд із елементами $y_i, i = \overline{1, n}$.

Пропонується моделювання часового ряду шляхом використання скінченних різниць другого порядку. $V_i = y_i - y_{i-1}, i = \overline{2, n}$, $B_j = V_j - V_{j-1}, j = \overline{3, n}$

Проводимо задання універсальної множини як інтервал який вміщує всі елементи ряду B_i. $U = [minB_i, maxB_i], i = \overline{3, n}$. Кількість субінтервалів розбиття визначаємо згідно[9], $m = 7$.

Крок 2. Знаходимо кумулятивну частоту потрапляння B_i в субінтервали розбиття універсальної множини. Проводимо розбиття трьох найбільш частотних субінтервалів у відношенні (4, 3, 2) для трьох інтервалів із найбільшою кумулятивною частотою. Оновлюємо кількість субінтервалів *m* відповідно.

Крок 3. Визначаємо нечіткі множини $X_j, j = \overline{1, m}$ для кожного отриманого субінтервалу із трикутною функцією належності для якої задаємо значення нижньої межі, верхньої та центру. Для заданого ряду проводимо його фазифікацію.

Крок 4. Проводимо дефазифікацію нечітких даних за допомогою наступної формули:

$$b_j = \{ \begin{array}{l} \dfrac{1 + 0.5}{\dfrac{1}{a_1} + \dfrac{0.5}{a_2}}, j = 1; \\ \dfrac{0.5 + 1 + 0.5}{\dfrac{0.5}{a_{j-1}} + \dfrac{1}{a_j} + \dfrac{0.5}{a_{j+1}}}, 2 \le j \le m - 1; \\ \dfrac{0.5 + 1}{\dfrac{0.5}{a_{j-1}} + \dfrac{1}{a_j}}, j = m; \end{array}$$

де a_j є центрами трикутних функцій належності для відповідних нечітких множин X_j.

Крок 5. Визначення прогнозних значень часового ряду використовуючи зворотну формулу переходу до попередніх значень із першого кроку.

Критерій оцінки якості моделі. Для перевірки якості побудованого прогнозу застосовують загальновідомі статистичні критерії MSE (Mean Squared Error), AFER (Average Forecasting Error Rate), MAPE (Mean Absolute Percentage Error). В нашому дослідженні будемо використовувати метрики MSE, AFER та MAPE, які задаються наступними формулами:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - y_i)^2}{n},$$

$$AFER = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - y_i|}{y_i},$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - y_i)^2}{y_i},$$

де y_i – реальне значення часового ряду; $\hat{y}_i$ – прогнозне значення.

Таблиця 1

Прогнозні результати роботи запропонованого методу

Year	Enrollmets	Forecast	$(\hat{y}_i - y_i)^2$	$\frac{\lvert\hat{y}_i - y_i\rvert}{y_i}$	$\frac{(\hat{y}_i - y_i)^2}{y_i}$
1971	13055	—	—	—	—
1972	13563	—	—	—	—
1973	13867	13963.29	9272.50	0.006944	0.668674
1974	14696	14746.26	2525.70	0.003420	0.171863
1975	15460	15416.59	1884.05	0.002808	0.121866
1976	15311	15516.52	42238.03	0.013423	2.758672
1977	15603	15542.08	3711.56	0.003905	0.237875
1978	15861	15786.59	5536.21	0.004691	0.349045
1979	16807	16694.26	12711.13	0.006708	0.756300
1980	16919	17045.52	16007.04	0.007478	0.946099
1981	16388	16323.52	4157.81	0.003935	0.253710
1982	15433	15514.95	6715.09	0.005310	0.435112
1983	15497	15242.71	64661.85	0.016409	4.172540
1984	15145	15218.95	5467.96	0.004883	0.361040
1985	15163	15173.08	101.55	0.000665	0.006697
1986	15984	15945.71	1465.89	0.002395	0.091710
1987	16859	17015.62	24530.43	0.009290	1.455035
1988	18150	18114.08	1290.43	0.001979	0.071098
1989	18970	18981.68	136.47	0.000616	0.007194
1990	19328	19330.68	7.19	0.000139	0.000372
1991	19337	19578.29	58222.72	0.012478	3.010949
1992	18876	18886.68	114.11	0.000566	0.006045
			MSE	**AFER**	**MAPE**
			13037.89	**0.005402**	**0.794095**

Таблиця 2

Порівняльна таблиця результатів прогнозних моделей

Year	Enroll mets	Song Chissom [9]	Song Chissom [10]	Jilani Burney [5]	Jilani Burney & Ardil [6]	Jilani Burney & Ardil [7]	Stevenson & Porter [8]	Proposed Method
1971	13055	—	—	—	14464	13579	–	—
1972	13563	14000	—	—	14464	13798	13410	—
1973	13867	14000	—	—	14464	13798	13932	13963.29
1974	14696	14000	—	14730	14710	14452	14664	14746.26
1975	15460	15500	14700	15615	15606	15373	15423	15416.59
1976	15311	16000	14800	15614	15606	15373	15847	15516.52
1977	15603	16000	15400	15611	15606	15623	15580	15542.08
1978	15861	16000	15500	15611	15606	15883	15877	15786.59
1979	16807	16000	15500	16484	16470	17079	16773	16694.26
1980	16919	16813	16800	16476	16470	17079	16897	17045.52
1981	16388	16813	16200	16469	16470	16497	16341	16323.52
1982	15433	16789	16400	15609	15606	15737	15671	15514.95
1983	15497	16000	16800	15614	15606	15737	15507	15242.71
1984	15145	16000	16400	15612	15606	15024	15200	15218.95
1985	15163	16000	15500	15609	15606	15024	15218	15173.08
1986	15984	16000	15500	15606	15606	15883	16035	15945.71
1987	16859	16000	15500	16477	16470	17079	16903	17015.62
1988	18150	16813	16800	18482	18473	17991	17953	18114.08
1989	18970	19000	19300	18481	18473	18802	18879	18981.68
1990	19328	19000	17800	19158	19155	18994	19303	19330.68
1991	19337	19000	19300	19155	19155	18994	19432	19578.29
1992	18876	—	19600	18475	18473	18916	18966	18886.68
MSE	**—**	**775687.0**	**407507.0**	**82269.0**	**227194.0**	**41426.0**	**21575.0**	**13037.89**
AFER	**—**	**4.38**	**3.11**	**1.41**	**2.39**	**1.02**	**0.57**	**0.54**
MAPE	**—**	**26.24**	**44.60**	**5.25**	**8.43**	**3.06**	**1.32**	**0.79**

Набір даних був взятий із робіт [5–10] (The enrollments of the University of Alabama) для перевірки якості запропонованої моделі. Запропонована модель дає кращі результати по обраних метриках (табл. 1, 2) у порівнянні із моделями [5–10].

Подальші дослідження полягають в побудові алгоритму нечіткого прогнозування для мультиваріантних часових рядів та розробці алгоритму оптимізації параметрів при дефазифікації.

1. Zadeh L. A. Fuzzy sets. *Information and control*. 1965. Vol. 8. Pp. 338–353. **2.** Song Q., Chissom B. S. Fuzzy time series and its models. *Fuzzy Sets and Systems*. 1993. Vol. 54. Pp. 269–277. **3.** Huarng K. Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems*. 2001. Vol. 12. Pp. 387–394. **4.** J. R. Hwang, S. M. Chen, C. H. Lee. Handling forecasting problems using fuzzy time series. *Fuzzy Sets and Systems*. 1998. Vol. 100. Pp. 217–228. **5.** T. A. Jilani, S. M. A. Burney. M-factor high order fuzzy time series forecasting for road accident data. *IEEE-IFSA 2007* : World Congress, Cancun, Mexico, June 18–21. Forthcoming in Book series Advances in Soft Computing, Springer-Verlag, 2007. **6.** T. A. Jilani, S. M. A. Burney, C. Ardil. Fuzzy Metric Approach for Fuzzy Time Series Forecasting based on Frequency Density Based Partitioning. *Proceedings of World Academy of Science. Engineering and Technology*. 2007. Vol. 23. Pp. 333–338. **7.** T. A. Jilani, S. M. A. Burney, C. Ardil. Multivariate high order fuzzy time series forecasting for car road accidents. *International Journal of Computational Intelligence*. 2007. Vol. 4, No. 1. Pp. 15–20. **8.** Stevenson M., J. E. Porter. Fuzzy time series forecasting using percentage change as the universe of discourse. *Proceedings of World Academy of Science, Engineering and Technology*. 2009. Vol. 55. Pp. 154–157. **9.** Q. Song, B. S. Chissom. Forecasting enrollments with fuzzy time series. Part I. *Fuzzy Sets and Systems*. Vol. 54. Pp. 1–9. **10.** Q. Song, B. S. Chissom. Forecasting enrollments with fuzzy time series. Part II. *Fuzzy Sets and Systems*. 1994. Vol. 62. Pp. 1–8.

REFERENCES:

1. Zadeh L. A. Fuzzy sets. *Information and control*. 1965. Vol. 8. Pp. 338–353. **2.** Song Q., Chissom B. S. Fuzzy time series and its models. *Fuzzy Sets and Systems*. 1993. Vol. 54. Pp. 269–277. **3.** Huarng K. Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems*. 2001. Vol. 12. Pp. 387–394. **4.** J. R. Hwang, S. M. Chen, C. H. Lee. Handling forecasting problems using fuzzy time series. *Fuzzy Sets and Systems*. 1998. Vol. 100. Pp. 217–228. **5.** T. A. Jilani, S. M. A. Burney. M-factor high order fuzzy time series forecasting for road accident data. *IEEE-IFSA 2007* : World Congress, Cancun, Mexico, June 18–21. Forthcoming in Book series Advances in Soft Computing, Springer-Verlag, 2007. **6.** T. A. Jilani, S. M. A. Burney, C. Ardil. Fuzzy Metric Approach for Fuzzy Time Series Forecasting based on Frequency Density Based Partitioning. *Proceedings of World Academy of Science. Engineering and Technology*. 2007. Vol. 23. Pp. 333–338. **7.** T. A. Jilani, S. M. A. Burney, C. Ardil. Multivariate high order fuzzy time series forecasting for car road accidents. *International Journal of Computational Intelligence*. 2007. Vol. 4, No. 1. Pp. 15–20. **8.** Stevenson M., J. E. Porter. Fuzzy time series forecasting using percentage change as the universe of discourse. *Proceedings of World Academy of Science, Engineering and Technology*. 2009. Vol. 55. Pp. 154–157. **9.** Q. Song, B. S. Chissom. Forecasting enrollments with fuzzy time series. Part I. *Fuzzy Sets and Systems*. Vol. 54. Pp. 1–9. **10.** Q. Song, B. S. Chissom. Forecasting enrollments with fuzzy time series. Part II. *Fuzzy Sets and Systems*. 1994. Vol. 62. Pp. 1–8.

Kinda V. V., Post-graduate Student (Rivne State University for the Humanities, Rivne)

FORECASTING OF TIME SERIES USING FUZZY LOGIC METHODS BY SPECIFYING A UNIVERSAL SET BY HIGHER ORDER FINITE DIFFERENCES

Time series analysis plays a major role in such fields of activity as business, science, economics, etc. One of the most difficult task here is to make a “proper” forecast with the high accuracy. There are dozens of forecasting models and techniques which have issue which consists of the lack of a sufficient amount of data for training algorithms mostly. Machine learning methods – regression models, neural networks with various architecture can be included here. On the other hand, the advantage of pyramidal extrapolation methods, forecasting time series based on fuzzy sets allows us to work with small data samples which gives us an ability for making a simultaneous forecast. Therefore we can use them at the very beginning of time series analysis as one of the basic ones using in financial time series. Especially working with series in fields such as currency exchanges, crypto and stock markets. It is proposed to use finite differences of various order to set a universe of discourse (UoD) in fuzzy time series analysis. This paper propose a modification of Stevenson Porter's method by using finite differences of higher

orders. Performance evaluation of the algorithm was calculated by mean square error (MSE), absolute forecast error rate (AFER) and mean absolute percentage error (MAPE) metrics. The proposed model gives better results according to the selected metrics (Tables 1, 2) in comparison with other models such as Song&Chissom, Jilani&Burney and Stevenson&Porter algorithms and their variations. Further research consists of in the construction of a fuzzy forecasting algorithm for multivariate time series and the application of appropriate algorithms for finite differences of different orders for a univariate time series and comparison with pyramidal extrapolation methods by defined metrics.

***Keywords*: time series; fuzzy set; finite differences; extrapolation; forecast.**