A NEW FUZZY TIME SERIES MODEL BASED ON HEDGE ALGEBRA TO FORECAST BITCOIN

Hoang Tung

Dong Nai University

Email: tungaptechbd@gmail.com

(Received: 1/10/2023, Revised: 14/11/2023, Accepted for publication: 18/12/2023)

ABSTRACT

The fuzzy time series model has become a research topic attracting attention because of its practical value in the field of time series forecasting, specifically, it is useful for time series with small observations or the one of strong fluctuations. This paper introduces a fuzzy time series model based on hedge algebra with a new formula for calculating forecasting values. The Bitcoin time series is employed for testing the model's performance. Experimental results show that the new model gives better forecasting results than the ARIMA model, which has been popular for a long time.

Keywords: *Time series, Forecasting, Fuzzy time series, Hedge Algebras, ARIMA, Bitcoin*

1. Introduction

In practice, there are many time series that do not have a large enough number of observations. This may be because this time series is newly formed or because it has not been collected in the past. Besides, there are also many time series that fluctuate very strongly, and their historical value quickly becomes obsolete, making the number of meaningful observations for the forecast not much.

(Wang, 2011), (Arumugam & Anithakumari, 2013), and (Senthamarai & Sakthivel, 2014) show that, in many time series with a small number of observations, the fuzzy time series model often gives quite good forecasting results, even better than the ARIMA model, which is given for good forecasting results.

Besides fuzzy sets, Hedge Algebra is another approach used for developing fuzzy time series models. Fuzzy time series models following this approach show quite good forecasting power, comparing experimental results, they give better forecasting results than many models using the fuzzy set approach.

(Tung et al., 2016) is the first study that presents a fuzzy time series model based on Hedge Algebra. According to the approach of this study, the values of time series that need forecasting, *c(t)*, will be quantified by the fuzzy linguistic terms (terms) forming the fuzzy time series *f(t)*. Then, these fuzzy terms, instead of being quantified by fuzzy sets, are quantified by Hedge Algebra. Specifically, each term is quantified by a fuzzy interval and a semantic core. Each such fuzzy interval is treated as an interval over the universe of discourse of *c(t)*.

Continuing this research direction, (Tung et al., 2016) propose using Hedge Algebra including only two hedges to generate qualitative terms, instead of having to search for suitable Hedge Algebras. As a result, this study introduced a new way of generating terms that provide more reasonable intervals, contributing to improving forecasting quality. Besides, this study also proposes to use the average value of historical values over the intervals to calculate the forecasting value.

(Tung & Thuan, 2019) introduces how to use difference series to improve the forecasting quality of HA based fuzzy time series model. Accordingly, instead of forecasting the *c(t)*, the difference series, *vc(t)*, is forecasted. This study argues that *vc(t)* carries more information than *c(t)* so may better reflect the motion law of *c(t)*. In addition, the study also proposes a new way of generating terms compared to previous studies. Experimental results show that this study gives quite positive results.

(Thuan & Tung, 2020) applies groups of relationships over time to calculate forecasting value. (Thuan & Tung, 2020) applies the PSO algorithm to optimize the parameters of Hedge Algebra. The forecasting results on some time series of these studies show that the Hedge Algebra approach to building fuzzy time series models is a positive direction.

According to (Petronio et al., 2016), interval forecasting is a form of forecasting that provides forecasting intervals, instead of point ones, that may contain future values of *c(t)*. There are not many studies that use fuzzy time series models to provide this kind of forecasting.

This study follows the mentioned studies to build a model by inheriting the terms generation method of the study (Tung & Thuan, 2019) and applying a new formula for computing forecasting values. The new model is tested for forecasting power on the Bitcoin time series, one of the highly volatile time series. The model also provides forecasting intervals.

The rest of the paper is organized as follows: Section 3 presents some concepts used to build the model. Section 4 presents the model-building steps. Section 5 presents the experimental results. Section six, the final section, presents some conclusions of the paper.

2. Preliminary

2.1. Fuzzy time series

In this section, we refer to (Song & Chissom, 1993) to briefly review some definitions of the Fuzzy time series.

Definition 1

Let $Y(t)$ (t = …, 0, 1, 2, …), a subset of R^1, be the universe of discourse on which $f_i(t)$ (i = 1, 2, …) are defined and $F(t)$ is the collection of $f_i(t)$ (i =1, 2, …). Then $F(t)$ is called FTS on $Y(t)$ (t=…, 0, 1, 2, …).

Definition 2. The relationship between $F(t)$ and $F(t - 1)$ can be presented as $F(t - 1) \rightarrow F(t)$. If let $A_i = F(t)$ and $A_j = F(t - 1)$; the relationship between $F(t)$ and $F(t - 1)$ is represented by $A_i \rightarrow A_j$, where A_i and A_j refer to the left-hand side and the right-hand side of the FLR.

Definition 3. Let $F(t)$ be a FTS. If $F(t)$ is caused by $F(t - 1)$ *or* $F(t - 2)$ *or* $\cdots$ *or* $F(t - m + 1)$ *or* $F(t - m)$ then this FR is represented by $F(t - m) \rightarrow F(t)$ *or* $\cdots$; $F(t - 2) \rightarrow F(t)$ *or* $F(t - 1) \rightarrow F(t)$ and is called a first-order FTS model.

2.2. Hedge Algebras

This section refers to (Ho & Long, 2007) to present some basic concepts in

Hedge Algebras. These are applied to the model presented in the next section.

Definition 4. The HA is defined by $AX = (X, G, C, H, \leq)$, X is a set of terms, $G= \{c^+, c^-\}$ is the set of primary generators, c^+ and c^- are, respectively, the negative and positive term belongs to X, $C =\{0, 1, W\}$ is a collection of constants in X, H is the set of hedges, $H = H^+ \cup H^-$, whereH^+, H^- is, respectively, the set of all positive and negative hedges of X; "≤" is a semantically ordering relation on X.

Each hedge is considered a unary operator. When applying $h \in H$ to x, we obtain $hx \in X$. The positive hedges increase semantic tendency and vice versa with negative hedges. It can be assumed that $H^-= \{h_{-1}<h_{-2}< ... <h_{-q}\}$ and $H^+= \{h_1<h_2< ... <h_p\}$.

$H(x)$ is the set of terms $u \in X$, $u = h_n\ldots h_1x$, with $h_n,\ldots, h_1 \in H$, generated from x by applying the hedges of H.

If X and H are linearly ordered sets, then $AX = (X, G, C, H, \leq)$ is called linear hedge algebras, furthermore, if AX is equipped with two additional operations $\sum$ and Φ that are, respectively, infimum and supremum of $H(x)$, then it is called complete linear hedge algebras (ClinHA).

Definition 5. Let $AX = (X, G, C, H, \leq)$ be a ClinHA. An $fm: X \to [0,1]$ is said to be a fuzziness measure of terms in X if:

(1). $fm(c^-)+fm(c^+) = 1$ and $\sum_{h\in H} fm(hu) = fm(u)$, for $\forall u \in X$; in this case fm is called complete;

(2). For the constants 0, W, and 1, $fm(0) = fm(W) = fm(1) = 0$;

(3). For $\forall x, y \in X$, $\forall h \in H$, $\frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)}$, this proportion does not depend on specific elements; therefore, it is called the fuzziness measure of the hedge h and denoted by $\mu(h)$.

Proposition 1. For each fuzziness measure fm on X, the following statements hold:

(1). $fm(hx) = \mu(h)fm(x)$, for every $x \in X$;

(2). $fm(c^-) + fm(c^+) = 1$;

(3). $\sum_{-q\leq i\leq p, i\neq 0} fm(h_i c) = fm(c)$, $c \in \{c^-, c^+\}$;

(4). $\sum_{-q\leq i\leq p, i\neq 0} fm(h_i x) = fm(x)$;

(5). $\sum_{-q\leq i\leq -1} \mu(h_i) = \alpha$ and $\sum_{1\leq i\leq p} \mu(h_i) = \beta$, where $\alpha, \beta > 0$ and $\alpha + \beta = 1$.

Definition 6. The fuzziness interval of the linguistic terms $x \in X$, denoted by $\Im(x)$, is a subinterval of [0,1], if $|\Im(x)| = fm(x)$ where $|\Im(x)|$ is the length of $fm(x)$, and recursively determined by the length of x as follows:

(1). If length of x is equal to 1 ($l(x)=1$), that mean $x \in \{c^-, c^+\}$, then $|\Im(c^-)| = fm(c^-)$, $|\Im(c+)|= fm(c^+)$ and $\Im(c^-) \leq \Im(c^+)$;

(2). Suppose that n is the length of x ($l(x)=n$) and fuzziness interval $\Im(x)$ has been defined with $|\Im(x)| = fm(x)$. The set $\{\Im(h_jx) | j \in [-q\wedge p]\}$, where $[-q\wedge p] = \{j | -q \leq j \leq -1 \text{ or } 1 \leq j \leq p\}$, is a partition of $\Im(x)$ and we have: for $h_px \leq x$, $\Im(h_px) \leq \Im(h_{p-1}x) \leq \ldots \leq \Im(h_1x) \leq \Im(h_{-1}x) \leq \ldots \leq \Im(h_{-q}x)$; for $h_px = \geq x$, $\Im(h_{-q}x) \leq \Im(h_{-q+1}x) \leq \ldots \leq \Im(h_{-1}x) \leq \Im(h_1x) \leq \ldots \leq \Im(h_px)$.

3. Proposed method

Input: *c(t)* is the time series to forecast;

Output: Forecasting values of *c(t)*

Model setting phrase (Tung & Thuan, 2019):

Step 1: Determine the number of terms to use for qualitative. The symbol for this number is *k*;

Step 2: Determine the universe of discourse *c(t)*, *U* = [*Dmin* – *D1*, *Dmax* + *D2*], where *Dmin*, *Dmax*, *D1*, and *D2* are respectively the lowest and highest historical values of *c(t)* and the values are chosen to ensure that future values of *c(t)* all belong to *U*.

Step 3: Use *AX* = (*X*, *G*, *C*, *H*, ≤) to generate terms, where *H* includes two hedges, h_{-1} and h_{+1}; $G = \{c^-, c^+\}$.

Let *p* be a FiFo list, *Lo* and *Hi* are the generators, respectively, and *t* is an integer variable.

Add *Lo* and *Hi* to *p*;

t=2;

Repeat until *t* >= *k*

{

Let *x* be a term variable;

If *p* is not empty, then *x* takes the value of the first element of *p*;

else break the loop;

If the fuzziness interval of *x* does not include any element of *c(t)* {

t = *t*-1;

continue;

}

Let *u* be an integer variable whose initial value is 0;

Use h_{-1} and h_{+1} operate *x* to produce two terms, $h_{-1}x$, and $h_{+1}x$;

Calculate the fuzzy interval of $h_{-1}x$ and $h_{+1}x$;

If the fuzzy interval of $h_{-1}x$ or $h_{+1}x$ contains any value of *c(t)*, increase *h* by one;

If *h* equals 2, then *t*=*t*+1;

If all elements of *p* include all historical values or include only one historical value, then break the loop;

}

Step 4:

Let *f(t)* be the set of fuzzy terms, initially *f(t)* = ∅;

For each *y* value of *c(t)*

If *y* belongs to the fuzzy interval of the term v generated in Step 3, then put *y* in *f(t)*.

Step 5:

Establish relationships *Ai* → *Aj* corresponds to two consecutive terms of *f(t)*.

Group each relationship with the same opposite side into a relation group. For example, suppose we have relations: *Ai* → *Aj*, *Ai* → *Am*, *Ai* → *Aj*, then we have the relation group *Ai* → *Aj*(2)*Am*(1). Here (2), and (1) respectively, the number of occurrences of *Aj* and *Am* in the relations with the left side is *Ai*.

Forecasting phrase:

For the time series *c(t)* at time *tt*, the forecast value of *c(t)* at time *tt*+1 is calculated as follows:

- Determine the first difference values of *c(t)*, call these values the difference series *v(t)*. Let *h* be the total number of elements of *c(t)*.

- If *c(tt)* belongs to the fuzziness interval of *Ai*, then find the relation group whose left side is *Ai*, for example, $Ai \rightarrow Aj(p)Am(q)$. Then the forecast value is calculated according to the formula:

$$\frac{p*TB(Aj)+\cdots+q*TB(Am)}{p+\cdots+q}+\frac{Sum(v(t1))}{t1+h}$$

- If the relation group has the left side *Ai* but the right side is ∅, then the forecasted value is $TB(Ai)+\frac{Sum(v(t1))}{t1+h}$.

4. Experimental results

The Bitcoin time series, recording the values at the time point of closing the transaction, is used for testing the forecasting accuracy of the proposed model. Besides, the forecasting capacity of this model is also compared to the ARIMA model one, the most commonly used model in time series forecasting.

This paper uses different value ranges of the Bitcoin time series as the experimental data set. The first range is selected from May 3, 2019, to May 2, 2021, this range is named Bit1; the second range is named Bit2 which is taken from May 2, 2020, to May 2, 2021; The third range is named Bit3, which records data from May 2, 2020, to July 27, 2020.

R's auto.arima function is used to determine the forecast value of Bitcoin by interval and point. The interval forecasting values (Lo, Hi) when using ARIMA have 95% confidence.

To estimate the accuracy of the forecasts, this paper uses the index RMSE $=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x'_i-x_i)^2}$, where x'_i is the predicted value, x_i is the historical value, and *n* is the number of forecasted values.

In order to evaluate the accuracy of the forecast results. This paper proposes 2 evaluation criteria:

(1) The forecast interval must contain the value of the future time series.

(2) The forecast interval has a length, Hi (High) - Lo (Low), the shorter the better.

Forecast Result of Bit1

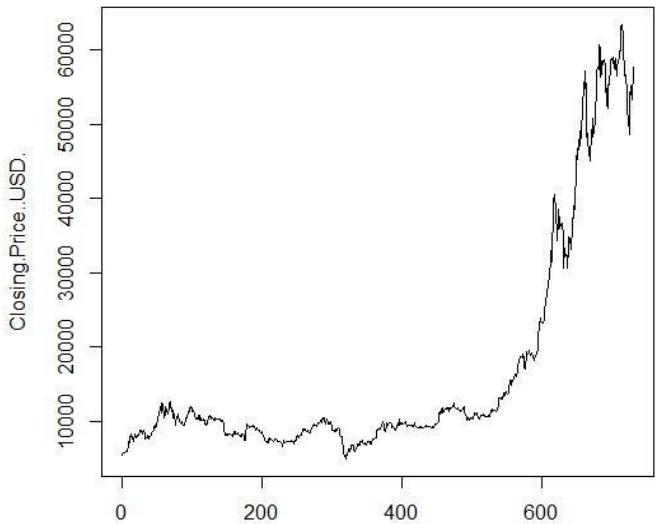

Fig 1: *Bitcoin from May 3, 2019, to May 2, 2021*

The graph of the Bitcoin time series over this period shows it is highly volatile and trending up. Bitcoin price is more and more volatile at a faster and stronger pace. From that, it can be seen that the law of change of historical values in this time series is very quickly outdated, that is, the law of motion in the previous time period may be very different from the law of motion in the previous time period.

Table 1: *Forecasting result on Bit1*

History Values	Point Forecast		Interval Forecasting			
	ARIMA	FTS	ARIMA		FTS	
57905.3	59981.25	56778	57853.60	62108.91	56523.0	58149.0
53633.3	61270.91	56778	57480.69	65061.13	56523.0	58149.0
57434.0	63093.48	56874	57029.23	69157.72	56523.0	58149.0
57186.3	64635.86	56778	56089.69	73182.04	56523.0	58149.0
57790.4	66325.56	56778	54973.36	77677.76	56523.0	58149.0
58272.2	67937.81	56778	53551.31	82324.30	56523.0	58149.0
57632.5	69590.77	59344	51927.84	87253.70	58149.0	60588.0
57951.4	71222.33	56778	50072.34	92372.31	56523.0	58149.0
55975.5	72865.14	56779	48022.30	97707.99	54898.0	56523.0
54697.2	74502.04	55898	45775.06	103229.01	53814.0	54898.0
RMSE	11458.8	991.0				

Experimental results show that, for a highly volatile series like Bitcoin, ARIMA gives a bad forecast while FTS gives a much better forecast. Especially for the interval forecast. In the Bit1 series prediction results, the forecast intervals by ARIMA are all longer than those predicted by FTS.

Forecast Result of Bit2

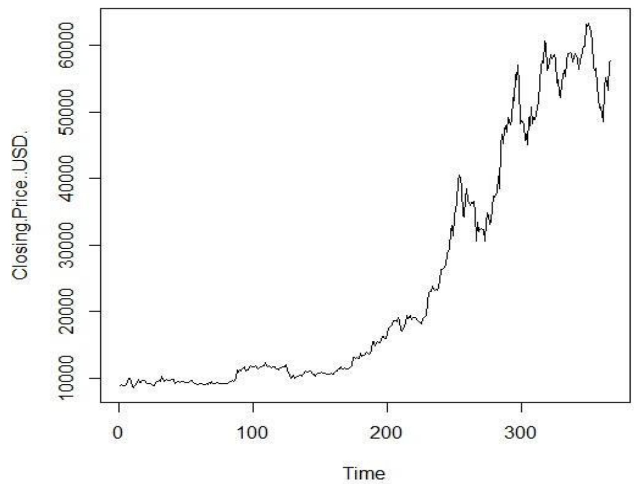

Fig 2: *Bitcoin from May 2, 2020, to May 2, 2021*

About the first 200 observations of Bit2 have fairly stable fluctuation, after that the time series increases quite suddenly and changes at a faster rate. This also shows that more than half of the observations at the beginning of

Bit2 have little effect on the variability of the second half of this time series.

Table 2: *Forecasting result on Bit2*

History Values	Point Forecast		Interval Forecasting			
	ARIMA	FTS	ARIMA		FTS	
57905.3	57677.98	56778	55287.84	60068.11	56523.0	58149.0
53633.3	57677.98	56778	54297.81	61058.14	56523.0	58149.0
57434.0	57677.98	56874	53538.13	61817.82	56523.0	58149.0
57186.3	57677.98	56778	52897.70	62458.25	56523.0	58149.0
57790.4	57677.98	56778	52333.46	63022.49	56523.0	58149.0
58272.2	57677.98	56778	51823.36	63532.59	56523.0	58149.0
57632.5	57677.98	59344	51354.26	64001.69	56523.0	58149.0
57951.4	57677.98	56778	50917.64	64438.31	56523.0	58149.0
55975.5	57677.98	56779	50507.56	64848.39	56523.0	58149.0
54697.2	57677.98	55898	50119.69	65236.26	54898.0,	56523.0
RMSE	1416.0	991.0				

Forecast Result of Bit3

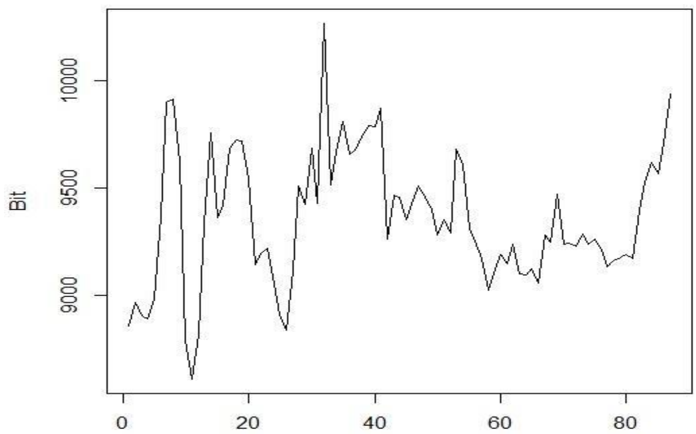

Fig 3: *Bitcoin from May 2, 2020, to July 27, 2020*

This is a series with quite stable fluctuations before a strong increase. Bit3 is chosen to test ARIMA in the case of a highly volatile series forecast.

Table 3: *Forecasting result on Bit3*

History Values	Point Forecast		Interval Forecasting			
	ARIMA	FTS	ARIMA		FTS	
11187.8	9761.346	11148	9313.936	10208.756	10842.0	11276.0
10939.7	9638.157	11221	9093.658	10182.657	10842.0	11276.0
11284.5	9552.712	11221	8967.212	10138.213	10842.0	11276.0
11119.0	9493.447	11458	8889.211	10097.682	11276.0	11565.0
11373.3	9452.340	11221	8839.294	10065.385	10842.0	11276.0
11766.7	9423.827	11458	8806.589	10041.066	11276.0	11565.0
11139.1	9404.051	11147	8784.805	10023.296	11276.0	11565.0
11261.8	9390.333	11220	8770.124	10010.543	10842.0	11276.0
11228.0	9380.819	11220	8760.147	10001.491	10842.0	11276.0
11653.4	9374.220	11221	8753.325	9995.114	10842.0	11276.0
RMSE	1834.9	207.0				

From the experimental results we see that, for highly volatile time series like Bitcoin, when using ARIMA to forecast, the large set of training values gives incorrect forecasting results compared to the training set of values. Meanwhile, FTS gives quite good forecast results on these series and is much better than ARIMA.

With the time series Bit1, ARIMA does not give accurate interval forecasting by FTS. ARIMA gives incorrect forecasting results in Bit2 when many historical values of this series are not in the forecast intervals, while the values of Bit2 are all within the forecasting intervals by FTS. The same thing happens with the forecast results in Bit3, even ARIMA in this forecast also gives forecast intervals that are quite far from the historical value of Bit3.

5. Conclusion

This paper presents a fuzzy time series model according to the Hedge Algebra approach with some improvements compared to the previous models. Specifically, a new formula for calculating forecasting value is proposed. The sum of the differences is used in this formula. This sum is used because it is considered a value that harmonizes the law of variation of the time series with respect to the forecast time. This sum is then divided by the number of historical values of the time series up to the time of the forecast and the total number of historical values of the time series.

The proposed model is applied to forecast Bitcoin time series with 3 value ranges with different lengths. The first two time series have fast, strong fluctuations, and the third time series is more stable. Forecast results show that the proposed model has higher accuracy than the ARIMA model, one of the most commonly used models in the field of time series forecasting.

REFERENCES

Arumugam, P. & Anithakumari, V. (2013). Fuzzy Time Series Method for Forecasting Taiwan Export Data. *International Journal of Engineering Trends and Technology (IJETT)*, *4*.

Chi-Chen, W. (2011). A comparison study between the fuzzy time series model and ARIMA model for forecasting Taiwan exports. *Expert Systems with Applications*, *38*, 9296–9304.

Ho, N.C., Long, N.V. (2007). Fuzziness measure on complete hedge algebras and quantifying semantics of terms in linear hedge algebras. *Fuzzy Sets and Systems*, *158*, 452 – 471.

Petronio, C. L. S. et al. (2016). Interval Forecasting with Fuzzy Time Series. *IEEE Symposium Series on Computational Intelligence*.

Song, Q., Chissom, B.S. (1993). Fuzzy time series and its models. *Fuzzy Sets and Systems 54* (3), 269–277.

Senthamarai, K. & Sakthivel, E. (2014). Fuzzy Time Series Model and ARIMA Model – A Comparative Study. *Indian journal of applied research*, *4*.

Tung, H., Thuan, N.D. & Loc, V.M. (2016). The partitioning method based on Hedge Algebras for Fuzzy Time Series Forecasting. *Journal of Science and Technology, Vietnam Academy of Science and Technology, 54*(5), 571-583. DOI: 10.15625/0866-708X/54/5/6518.

Tung, H., Thuan, N.D. & Loc, V.M. (2016). Time series forecasting method based on fuzzy time series by Hedge Algebra approach. *FAIR Conference 9*. DOI: 10.15625/vap.2016.00075.

Tung, H. & Thuan, N.D. (2019). Using model of fuzzy time series based on hedge algebras and variations to forecast time series. *Hội thảo khoa học Hệ thống thông tin trong Kinh doanh và Quản lý ISBM 2019, Trường Đại học Công nghệ thông tin* (p.22-34). Thành phố Hồ Chí Minh: Nxb Đại học Quốc gia Thành phố Hồ Chí Minh.

Thuan, N.D. & Tung, H. (2020). Forecasting with Improved Model of Fuzzy Time Series Based on Hedge Algebras. *International Journal of Advanced Trends in Computer Science and Engineering, 9*(5), 8069-8674.

Thuan, N.D. & Tung, H. (2020). Using Fuzzy Time Series Model Based on Hedge Algebras and Relationship Groups Following Time Points for Forecasting Time Series. *Proceedings of the 6th International Conference on Future Data and Security Engineering, 2020 (*tr. 401-410). Springer Verlag.

MỘT MÔ HÌNH CHUỖI THỜI GIAN MỜ MỚI DỰA TRÊN ĐẠI SỐ GIA TỬ DÙNG CHO DỰ BÁO GIÁ BITCOIN

Hoàng Tùng

Trường Đại học Đồng Nai

Email: tungaptechbd@gmail.com

(Ngày nhận bài: 1/10/2023, ngày nhận bài chỉnh sửa: 14/11/2023, ngày duyệt đăng: 18/12/2023)

TÓM TẮT

Mô hình chuỗi thời gian mờ đã trở thành một chủ đề nghiên cứu thu hút nhiều sự chú ý của các nhà khoa học bởi ứng dụng thực tế của nó trong dự báo chuỗi thời gian, đặc biệt là trong dự báo các chuỗi thời gian có số lượng quan sát nhỏ hoặc các chuỗi thời gian có sự biến động lớn về giá trị. Bài báo này giới thiệu một mô hình chuỗi thời gian mờ dựa trên Đại số gia tử với một đề xuất mới về công thức tính giá trị dự báo. Chuỗi Bitcoin được sử dụng để kiểm nghiệm năng lực dự báo của mô hình này. Kết quả thực nghiệm cho thấy, mô hình đề xuất cho kết quả dự báo chính xác hơn mô hình ARIMA, vốn là mô hình đã và đang được sử dụng phổ biến nhất trong dự báo chuỗi thời gian.

Từ khóa*: Chuỗi thời gian, Dự báo, Chuỗi thời gian mờ, Đại số gia tử, mô hình ARIMA, Bitcoin*