

ỨNG DỤNG LÝ THUYẾT MỀM TRUNG TÍNH TRONG PHÂN TÍCH HÀNH VI VÀ HỖ TRỢ RA QUYẾT ĐỊNH TỪ DỮ LIỆU CHUỖI THỜI GIAN

Lưu Thị Kim Ngọc, Nguyễn Thị Nhã Linh, Nguyễn Thanh Nhã, Bùi Quang Thịnh
Trường Đại học Tiền Giang

Tóm tắt: Nghiên cứu này đề xuất một mô hình dự báo chuỗi thời gian đa thuộc tính dựa trên lý thuyết mềm trung tính (Neutrosophic Soft Sets – NSS) nhằm nâng cao độ chính xác trong bối cảnh dữ liệu biến động và chứa nhiều yếu tố không chắc chắn. Mô hình MA–NSSFTS được xây dựng dựa trên ba thành phần chân trị, bất định và phản chân trị, kết hợp với entropy thông tin để xác định trọng số thuộc tính, cùng cơ chế láng giềng gần nhất có trọng số trong không gian NSS. Bộ dữ liệu TAIEX năm 2004 được sử dụng để kiểm chứng mô hình, với bốn thuộc tính gồm giá mở cửa, giá cao nhất, giá thấp nhất và giá đóng cửa. Kết quả thực nghiệm cho thấy MA–NSSFTS đạt độ chính xác cao (chỉ số MAPE = 1.58%). Nghiên cứu khẳng định tiềm năng của phương pháp trong phân tích xu hướng, dự báo hành vi và hỗ trợ ra quyết định trong các lĩnh vực có dữ liệu nhiễu và không xác định.

Từ khóa: Lý thuyết mềm trung tính, Dự báo chuỗi thời gian, Hỗ trợ ra quyết định, Trọng số entropy

APPLICATION OF NEUTROSOPHIC SOFT SET THEORY IN BEHAVIORAL ANALYSIS AND DECISION SUPPORT FOR TIME SERIES DATA

Abstract: This study proposes a multi-attribute time series forecasting model based on Neutrosophic Soft Sets (NSS) to improve prediction accuracy under fluctuating and highly uncertain data conditions. The proposed MA–NSSFTS model integrates the truth, indeterminacy, and falsity components of NSS with an information-entropy-based attribute weighting scheme and a weighted nearest-neighbor mechanism in the NSS feature space. The 2004 TAIEX dataset, consisting of four attributes—open, high, low, and close prices—was employed for empirical evaluation. Experimental results demonstrate that the proposed model achieves high forecasting accuracy (MAPE = 1.58%). These findings highlight the potential of the model for trend analysis, behavioral forecasting, and decision-support applications in domains characterized by noisy and uncertain data.

Keywords: Neutrosophic Soft Set Theory, Time Series Forecasting, Decision Support, Entropy Weighting

Nhận bài: 12/10/2025

Phản biện: 12/11/2025

Duyệt đăng: 15/11/2025

I. ĐẶT VẤN ĐỀ

Dữ liệu chuỗi thời gian đóng vai trò quan trọng trong nhiều lĩnh vực như kinh tế, tài chính, khoa học xã hội, giáo dục và tâm lý học, nơi các hiện tượng được quan sát liên tục theo thời gian và chịu ảnh hưởng của nhiều yếu tố bất định. Các chuỗi thời gian hành vi đặc biệt phức tạp do chịu tác động đồng thời của tâm trạng, cảm xúc, động cơ và môi trường, dẫn đến mức độ nhiễu và không chắc chắn cao hơn so với các dữ liệu đo lường truyền thống.

Các mô hình dự báo cổ điển như ARIMA, Holt–Winters hoặc các mô hình tuyến tính ít hiệu quả trong môi trường dữ liệu nhiễu và không chắc chắn. Để cải thiện khả năng xử lý mơ hồ, các mô hình Fuzzy Time Series (FTS) được đề xuất nhằm biểu diễn biến động bằng các mức ngôn ngữ thay vì giá trị số tuyệt đối. Tuy nhiên, FTS truyền thống vẫn chưa thể mô tả sự bất định – yếu tố cốt lõi trong các dữ liệu hành vi.

Lý thuyết Neutrosophic Set do Smarandache phát triển mở rộng Fuzzy Set và Intuitionistic Fuzzy Set. Điểm mạnh là mô tả rõ ràng mức độ bất định (Indeterminacy), yếu tố rất quan trọng khi phân tích hành vi con người. Soft Set của Maji

cung cấp công cụ xử lý tham số linh hoạt; sự kết hợp giữa Soft Set và Neutrosophic Set tạo thành Neutrosophic Soft Set – NSS, đặc biệt mạnh trong các bài toán ra quyết định đa tham số. NSS được áp dụng trong nhiều lĩnh vực: Ra quyết định nhiều thuộc tính; Dự báo chuỗi thời gian; Tối ưu hóa hành vi và phân tích mơ hồ.

Tuy nhiên, phần lớn các nghiên cứu NSS dựa trên đơn thuộc tính hoặc coi trọng số các thuộc tính là như nhau, bỏ qua vai trò khác biệt giữa các khía cạnh của dữ liệu hành vi. Hơn nữa, các mô hình hiện có hiếm khi tích hợp phương pháp trọng số entropy để tăng tính thích ứng của mô hình trước sự biến động của từng thuộc tính.

Từ những hạn chế đó, nghiên cứu này đề xuất mô hình MA–NSSFTS (Multi-Attribute Neutrosophic Soft Set Fuzzy Time Series) kết hợp: Lý thuyết mềm trung tính; Phương pháp fuzzification; Trọng số entropy và Kỹ thuật láng giềng gần nhất dựa trên khoảng cách NSS. Mục tiêu nhằm cung cấp mô hình phân tích hành vi và hỗ trợ ra quyết định có độ chính xác cao, có thể áp dụng trong các lĩnh vực mang tính bất định như tâm lý – giáo dục, tài chính hành vi, phân tích cảm xúc hay đánh giá xu hướng học tập.

Phần còn lại của bài báo được cấu trúc như sau: Mục 2 trình bày kiến thức nền tảng; Mục 3 đề xuất mô hình MA-NSSFTS; Mục 4 giới thiệu bộ dữ liệu và kết quả thực nghiệm-so sánh; Mục 5 kết luận.

II. NỘI DUNG NGHIÊN CỨU

2.1. Các kiến thức nền tảng để xây dựng mô hình MA-NSSFTS

Định nghĩa 1. Chuỗi thời gian hành vi được sử dụng trong nhiều lĩnh vực như tâm lý học nhận thức, phân tích cảm xúc, hành vi tài chính và giáo dục. Các dữ liệu này thường chứa mâu thuẫn, bất định và thay đổi không tuyến tính (Kim, 2018). Phân tích hành vi qua chuỗi thời gian đòi hỏi mô hình có khả năng nhận diện xu hướng, phản ứng và nhiễu đồng thời.

Định nghĩa 2. Giả sử U là tập hợp lớn, $N(U)$ là tập hợp của tất cả các tập trung tính của U , E là một tập hợp của các tham số có thể mô tả các phần tử của U và $D \subseteq E$. Sau đó, một tập hợp mềm trung tính N trên U là một tập hợp được xác định bởi một hàm giá trị tập hợp f_N đại diện cho một ánh xạ: $f_N: E \rightarrow N(U)$ như vậy $f_N(e) = \emptyset$ nếu $e \in E - D$. Trong đó f_N là hàm xấp xỉ gọi là hàm tập hợp mềm trung tính N . Nó có thể được biểu diễn bởi một cặp $(f_N; E)$ như sau

$$N = (f_N; E) = \{(e, f_N(e)) : e \in E\} = \{(e, \{x; T_{f_N(e)}; I_{f_N(e)}; F_{f_N(e)} : x \in U\}) : e \in E\}$$

với: $T_{f_N(e)}, I_{f_N(e)}, F_{f_N(e)} \in [0, 1]$.

Định nghĩa 3. Giả sử $A = (T_A; I_A; F_A); B = (T_B; I_B; F_B)$ là các tập mềm trung tính, khoảng cách Euclidean

giữa A, B là $d(A, B) = \sqrt{\sum_{i=1}^n \frac{\Delta_{IT}^2(x) + \Delta_{IU}^2(x) + \Delta_{IF}^2(x)}{3}}$, với $\Delta_{IT}(x) = T_{A(e_i)} - T_{B(e_i)}; \Delta_{IU}(x) = I_{A(e_i)} - I_{B(e_i)}; \Delta_{IF}(x) = F_{A(e_i)} - F_{B(e_i)}$.

Định nghĩa 4. Giả sử Q là độ đo entropy thông tin, công thức được định nghĩa bởi biểu thức sau: $Q = -\sum_{i=1}^m P(i) \log_2(P(i))$ trong đó $P(i)$ biểu thị xác suất xuất hiện tại thời điểm tương ứng L_i ($i=1, 2, \dots, m$), trong đó $\sum_{i=1}^m P(i) = 1$ và $0 \leq P(i) \leq 1$.

2.2. Mô hình MA-NSSFTS được cải tiến từ các mô hình chuỗi thời gian trong

2.2.1. Mô tả bài toán

Bài toán đặt ra trong nghiên cứu này là dự báo giá đóng cửa kế tiếp của chuỗi thời gian tài chính trong điều kiện dữ liệu chứa nhiều yếu tố bất định, dao động mạnh và chịu tác động của tâm lý thị trường. Chuỗi thời gian được ký hiệu dưới dạng

$$D = \{("Open_t, High_t, Low_t, Close_t") | t=1, \dots, T\},$$

trong đó mỗi điểm dữ liệu biểu diễn thông tin của một phiên giao dịch. Mục tiêu chính là xây dựng mô hình có khả năng học được quy luật biến

thiên tiềm ẩn của chuỗi và từ đó dự báo giá đóng cửa $Close_t$ dựa trên các đặc trưng của những phiên trước đó.

Để tăng cường khả năng mô hình hóa biến động, bài toán được mở rộng bằng cách bổ sung hai thuộc tính dẫn xuất: Amplitude thể hiện mức dao động trong ngày $High_t - Low_t$ và Volume xấp xỉ (ước lượng biến động đóng cửa giữa hai phiên liên tiếp $|Close_t - Close_{t-1}|$). Sau đó, mỗi thuộc tính được chuyển thành chuỗi biến thiên $\Delta A_t = A_t - A_{t-1}$, giúp nhấn mạnh chiều hướng chuyển động thay vì giá trị tuyệt đối.

Thách thức trọng tâm của bài toán nằm ở việc: (i) Chuỗi giá tài chính có tính phi tuyến, khó dự đoán, và thay đổi liên tục; (ii) Các biến động nhỏ nhưng liên tiếp có thể dẫn đến sai lệch lớn khi dự báo; (iii) Yếu tố tâm lý, nhiễu ngẫu nhiên và sự không chắc chắn tích hợp trong dữ liệu khiến các phương pháp cổ điển dễ bị suy giảm hiệu suất.

Do đó, nghiên cứu đề xuất ứng dụng Mô hình MA-NSSFTS nhằm khai thác ưu thế của Fuzzy Time Series và Neutrosophic Soft Sets, cho phép diễn giải tốt hơn mức độ bất định, đồng thời tạo ra dự báo ổn định và chính xác hơn cho chuỗi thời gian tài chính.

2.2.2. Mô hình MA-NSSFTS

Các bước của thuật toán đề xuất được mô tả cụ thể như sau:

Bước 1: Chuẩn hóa dữ liệu đầu vào

$$D = \{(A_t, \Delta A_t) | A \in \{Close, Open, Volume, Amplitude\}\},$$

trong đó:

(1) Dữ liệu gốc: $D = \{(Date_t, Open_t, High_t, Low_t, Close_t) | t=1, \dots, T\}$.

(2) Thuộc tính Volume: $Volume_t = |Close_t - Close_{(t-1)}|$.

(3) Thuộc tính biên độ dao động: $Amplitude_t = High_t - Low_t$.

(4) Độ biến thiên: $\Delta A_t = A_t - A_{(t-1)}$.

Bước 2: Chuỗi biến thiên của mỗi thuộc tính được ánh xạ sang 5 mức ngôn ngữ:

$L_1 =$ strong down, $L_2 =$ slightdown, $L_3 =$ equal, $L_4 =$ slight up, $L_5 =$ strong up.

Giá trị chuẩn hoá trung bình tuyệt đối:

$$d_A = \frac{1}{T-1} \sum_{t=2}^T |\Delta A_t|.$$

Bước 3: Xây dựng MA-NSS với cửa sổ bậc $n = 3$: $W_t = \{L_{t-3}, L_{t-2}, L_{t-1}\}$

+ Tính xác suất từng nhãn: $P_i = \frac{\text{count}(L_i)}{3}$

+ Chuẩn hóa về NSS:

$$\Delta_{IT}(x) = T_{A(e_i)} - T_{B(e_i)}; \Delta_{II}(x) = I_{A(e_i)} - I_{B(e_i)}; \Delta_{IF}(x) = F_{A(e_i)} - F_{B(e_i)}.$$

Bước 4: Tính trọng số entropy của thuộc tính

$$A: w_A^* = \frac{1-E_A}{\sum_A (1-E_A)} \text{ với } E_A = \frac{-\sum P_i \log P_i}{\log 5}.$$

Bước 5: Tính khoảng cách NSS có trọng số

$$d(N(t), N(s)) = \sqrt{\sum_A w_A^* ((T_t - T_s)^2 + (I_t - I_s)^2 + (F_t - F_s)^2)}^{1/2}.$$

Bước 6: Dự báo FTS: Với tập huấn luyện $T = \{N(s) | s < t\}$,

+ Lân cận gần nhất: $s^* = \arg(\min)_{s < t} d(N(t), N(s))$.

+ Dự báo độ biến thiên: $\widehat{\text{Close}}_t = \Delta \text{Close}_{s^*}$.

+ Dự báo giá đóng cửa: $\widehat{\text{Close}}_t = \text{Close}_{t-1} + \widehat{\text{Close}}_t$.

2.3. Thí nghiệm kiểm chứng hiệu quả thuật toán trên bộ dữ liệu thực tế TAIEX 2004, nhằm đánh giá hiệu năng và so sánh kết quả

2.3.1. Dữ liệu thực nghiệm và chỉ số đánh giá

Bộ dữ liệu được sử dụng trong nghiên cứu là TAIEX 2004, bao gồm chuỗi giá giao dịch hàng ngày của Chỉ số Chứng khoán Đài Loan (Taiwan Capitalization Weighted Stock Index – TAIEX) trong năm 2004. Dữ liệu chứa các trường thông tin tiêu chuẩn của thị trường tài chính, bao gồm giá mở cửa (Open), giá cao nhất (High), giá thấp nhất (Low) và giá đóng cửa (Close) của từng phiên giao dịch. Tổng cộng có 249 phiên giao dịch, phản ánh đầy đủ các biến động theo chu kỳ ngắn hạn và các giai đoạn tăng–giảm mạnh của thị trường trong năm. Đây là bộ dữ liệu thường được sử dụng trong các nghiên cứu về dự báo chuỗi thời gian tài chính nhờ tính biến động cao và cấu trúc phi tuyến đặc trưng.

Để đánh giá hiệu quả dự báo của mô hình MA-NSSFTS, nghiên cứu sử dụng ba chỉ số sai số phổ biến trong phân tích chuỗi thời gian, bao gồm:

(1) Sai số tuyệt đối trung bình (Mean Absolute Error): $MAE = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t|$. Chỉ số MAE phản ánh mức độ sai lệch trung bình giữa giá trị thực tế và giá trị dự báo, với đơn vị đo trùng với đơn vị của dữ liệu đầu vào. Giá trị MAE càng nhỏ cho thấy mô hình càng ổn định và ít sai lệch tuyệt đối.

(2) Sai số căn bậc hai trung bình (Root Mean

Square Error): $RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}$

Chỉ số RMSE đánh giá độ lệch bình phương, nhấn mạnh các sai số lớn. Đây là chỉ số quan trọng trong các bài toán dự báo tài chính, nơi những sai số lớn thường gây ảnh hưởng nghiêm trọng đến độ tin cậy của mô hình.

(3) Sai số phần trăm tuyệt đối trung bình (Mean Absolute Percentage Error):

$$MAPE = 100 \times \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right|.$$

Chỉ số MAPE đo lường sai số dự báo tương đối theo phần trăm, cho phép đánh giá độ chính xác theo tỷ lệ thay vì giá trị tuyệt đối. MAPE dưới “2%” được xem là mức hiệu quả rất cao đối với dữ liệu tài chính có độ biến động mạnh như TAIEX 2004.

Ba chỉ số này được sử dụng đồng thời nhằm đảm bảo đánh giá toàn diện mức độ ổn định, độ chính xác và khả năng kiểm soát sai số lớn của mô hình đề xuất.

2.3.2. Kết quả thực nghiệm

Sau khi áp dụng đầy đủ các giai đoạn của mô hình MA-NSSFTS lên bộ dữ liệu TAIEX 2004, kết quả thực nghiệm thu được như sau:

$MAE = 96.26$, $RMSE = 128.38$, $MAPE = 1.58\%$

Các kết quả trên cho thấy mô hình đạt được độ chính xác cao. MAE và RMSE đều nằm trong phạm vi thấp so với biên độ dao động trung bình của chỉ số TAIEX trong năm 2004 (khoảng 200–300 điểm), chứng tỏ mô hình kiểm soát tốt sai số tuyệt đối cũng như sai số bình phương. Đặc biệt, MAPE chỉ ở mức 1.58%, phản ánh khả năng dự báo theo tỷ lệ rất chính xác và vượt trội so với nhiều mô hình chuỗi thời gian truyền thống.

2.3.3. Thảo luận và đánh giá

Để đánh giá hiệu quả của thuật toán đề xuất, một so sánh kết quả thực nghiệm được tiến hành với nhiều mô hình sẵn có trước đó, sử dụng cùng một bộ dữ liệu trong khuôn khổ ra quyết định đa tiêu chí. Các mô hình được so sánh bao gồm NSS-Entropy và NSS-FTS. Kết quả được trình bày trong Bảng 1 bên dưới với các chỉ số tương ứng được tính theo công thức đề xuất phía trên và kết quả công bố trong.

Bảng 1. Bảng so sánh kết quả thực nghiệm giữa các mô hình

Mô hình	MAE	RMSE	MAPE
NSS–Entropy	N/A	53.60 -128.32	N/A
NSS–FTS	78.25	101.26	1.62%
MA-NSSFTS (đề xuất)	96.26	128.38	1.58%

Nhìn chung, kết quả thực nghiệm khẳng định rằng MA–NSSFTS là mô hình dự báo hiệu quả cho dữ liệu tài chính có tính biến động cao, nhờ khả năng kết hợp đồng thời thông tin từ nhiều thuộc tính và cơ chế mô hình hóa bất định thông qua lý thuyết mềm trung tính.

III. KẾT LUẬN

Nghiên cứu đề xuất mô hình MA–NSSFTS tích hợp đa thuộc tính, lý thuyết mềm trung tính và trọng số entropy, giúp xử lý hiệu quả bất định và

những trong chuỗi thời gian. Kết quả thực nghiệm cho thấy mô hình đạt hiệu suất dự báo cao, MAPE chỉ 1.58%. Mô hình có tiềm năng ứng dụng rộng rãi trong các lĩnh vực như tài chính hành vi, giáo dục, phân tích tâm lý, đánh giá mức độ cảm xúc theo thời gian, và hệ hỗ trợ ra quyết định. Hướng phát triển tiếp theo gồm: tối ưu hóa tham số của sở, tích hợp mô hình học sâu, mở rộng sang dữ liệu cảm xúc và hành vi người dùng theo thời gian thực.

TÀI LIỆU THAM KHẢO

Wu, X., Shi, B., Dong, Y., Huang, C., Faust, L., & Chawla, N. V. (2018, October). Restful: Resolution-aware forecasting of behavioral time series data. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (pp. 1073-1082).

Hwang, J. R., Chen, S. M., & Lee, C. H. (1998). Handling forecasting problems using fuzzy time series. *Fuzzy sets and systems*, 100(1-3), 217-228.

Song, Q., & Chissom, B. S. (1993). Fuzzy time series and its models. *Fuzzy sets and systems*, 54(3), 269-277.

Smarandache, F., Broumi, S., Singh, P. K., Liu, C. F., Rao, V. V., Yang, H. L., & Elhassouny, A. (2019). Introduction to neutrosophy and neutrosophic environment. In *Neutrosophic Set in Medical Image Analysis* (pp. 3-29). Academic Press.

Maji, P. K., Biswas, R., & Roy, A. R. (2003). Soft set theory. *Computers & mathematics with applications*, 45(4-5), 555-562.

Broumi, S. (2013). Generalized neutrosophic soft set. *Infinite Study*.

Karaaslan, F. (2014). Neutrosophic soft sets with applications in decision making. *Infinite Study*.

Zhao, A., Jie, H., Guan, H., & Guan, S. (2020). A multi-attribute fuzzy fluctuation time series model based on neutrosophic soft sets and information entropy. *International Journal of Fuzzy Systems*, 22(2), 636-652.

Guan, H., He, J., Guan, S., & Zhao, A. (2019). Neutrosophic soft sets forecasting model for multi-attribute time series. *IEEE Access*, 7, 25575-25588.

Singh, S. R. (2007). A simple method of forecasting based on fuzzy time series. *Applied mathematics and computation*, 186(1), 330-339.