

Article

Research on Fuzzy Temporal Event Association Mining Model and Algorithm

Aihua Zhu ^{1,2}, Zhiqing Meng ^{1,*}  and Rui Shen ³ ¹ School of Management, Zhejiang University of Technology, Hangzhou 310023, China² Changzhou Technical Institute of Tourism & Commerce, Jiangsu Union Technical Institute, Changzhou 213032, China³ School of Economics, Zhejiang University of Technology, Hangzhou 310023, China

* Correspondence: mengzhiqing@zjut.edu.cn

Abstract: As traditional models and algorithms are less effective in dealing with complex and irregular temporal data streams, this work proposed a fuzzy temporal association model as well as an algorithm. The core idea is to granulate and fuzzify information from both the attribute state dimension and the temporal dimension. After restructuring temporal data and extracting fuzzy features out of information, a fuzzy temporal event association rule mining model as well as an algorithm was constructed. The proposed algorithm can fully extract the data features at each granularity level while preserving the original information and reducing the amount of computation. Furthermore, it is capable of efficiently mining the possible rules underlying different temporal data streams. In experiments, by comparing and analyzing stock trading data in different temporal granularities, the model and algorithm identify association events in disorder trading. This not only is valuable in identifying stock anomalies, but also provides a new theoretical tool for dealing with complex irregular temporal data.

Keywords: fuzzy temporal data mining; fuzzy temporal association rules; fuzzy temporal event; temporal type

MSC: 68Q32; 62G08; 03B52



Citation: Zhu, A.; Meng, Z.; Shen, R. Research on Fuzzy Temporal Event Association Mining Model and Algorithm. *Axioms* **2023**, *12*, 117. <https://doi.org/10.3390/axioms12020117>

Academic Editors: Feng Feng, Harish Garg and G. Muhiuddin

Received: 26 November 2022

Revised: 15 January 2023

Accepted: 19 January 2023

Published: 23 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Temporal data stream (also known as time-varying data streams) mining is currently of great importance in the field of data mining, where the discovery of implicit knowledge (associative features, categorical features, etc.) in irregular temporal data streams is critical. First, each record in a temporal data stream varying irregularly over time displays multiple granularities, magnitudes and dimensions, as shown by data in finance. Second, in the same data bank there are different types of data, i.e., qualitative vs. quantitative, and structured vs. unstructured, e.g., trading data vs. public opinion data in the stock market, consultation data vs. medical records in hospitals, and weather detection data vs. body sensory data in describing temperature, etc. Finally, features and patterns of the data flowing in temporal data streams are obscured by the overwhelming amount and complexity of the data. All of these make it extremely difficult to build effective knowledge-discovery algorithms [1], such as those for urban traffic management, medical systems, stock market trading, and social networks. The current widely used mining algorithms, including Apriori, SVM, BP, and deep learning, focus on non-temporal association rules, classification, and feature recognition [2,3]. For temporal data streams, there is still a lack of effective models and algorithms. Therefore, the study of temporal data stream mining models and algorithms is extremely valuable and challenging.

Since it is difficult to discover the information features and patterns in complex time data, deformation processing methods are needed. As early as 1979, for example,

LA Zadeh proposed a theory of information granulation [4] to deal with complex data. Subsequently, the idea was further developed into the theory of granular computing. With it, complex data (e.g., unstructured verbal text, time series data, etc.) are translated into digital information at different granularity, which reveals features and patterns at different scales [5]. Based on this, this paper reconstructed the information in the raw data under different time granularity combining fuzzy theory, which helps rebuild the problem from multi-granularity spaces and establish a fuzzy feature-discovery model for multi-granularity data. Wang Guoyin et al. [6] indicated that multi-grain hierarchical thinking is consistent with human cognitive mechanisms, which helps human–computer cognitive mechanism integration. In this way, it enhances the computability of information and the interpretability of results. Particularly for the information granularity processing of time series, Yu Fusheng et al. [7] summarized previous experiences and proposed that information granularity processing should preferably be based on both temporal granularity and value granularity. Therefore, processing temporal data streams should be a multi-granular parallel processing work. Thus, knowledge-discovery models and algorithms with multiple granularities or different scales play an important role in solving complex temporal data streams.

Research on mining the original timestamp data is divided into two main categories: one is to directly build models and algorithms without partitioning the data at any granularity, with the argument that data deformation will lose the features inherent in the raw information. It is mainly shown by econometric models such as AMAR and panel models, etc. However, it has proved to be extremely problematic for complex time series. Therefore, another category dealing with complex time series data has risen, i.e., the construction of temporal data models and algorithms using different time granularity information, which has become an important research direction for temporal data mining. The methods in the second category have been successfully applied in many fields where there are strong systematic associations with historical data, such as in transport, medical, insurance, finance, and meteorology. Temporal data-mining methods have already achieved many research results in terms of association rule mining, data trend analysis, pattern analysis, causality analysis, and periodic pattern analysis [8].

The following discussion focuses on five aspects concerning data mining of time or temporal divisions and fuzzy temporal mining.

- (1) Some researchers have adopted temporal-segmentation methods to perform temporal data mining, such as Lovrić who pointed out that temporal segmentation is considered as a pre-processing step and a core task for various temporal data mining methods [9], the core problem being solving the theory of temporal data-mining models. For example, some scholars proposed generating subsequences by using the sliding window method on the raw time series, and then mine association rules based on operations such as feature–symbol transformation or similarity analysis on the subsequences [10,11]. Some scholars have also proposed to segment the subsequence separately and extract features based on item lifespans, and then mine the association rules between them [12,13].
- (2) Other researchers focused on diversity treatment of time granularity: raw time is organized into granular pieces of fixed time length, and the reconstructed data are mined separately for different time granularity, such as the periodical association rule mining based on the calendar [14]. However, this kind of time granularity segmentation is relatively rough. Usually day-, month-, or year- scales are adopted for the segmentation standard and sometimes the length of the segmentation is even not consistent, which all bring problems to the accuracy of the following mining. There is research showing the importance of temporal data mining for the discovery of temporal correlations or approximate periodic phenomena, but in practice there is the problem that it is difficult to precisely distinguish feature changes. Therefore, in this paper, based on the existing temporal type definition [15], combined with

the fuzzy set theory [16], a new temporal fuzzy correlation model and algorithm were reconstructed.

- (3) In the area of fuzzy data mining with time series, fuzzy theory was introduced as early as 20 years ago to handle the relationship between fuzzy and temporal data [17]. Fuzzy theory mainly deals with imprecise, non-quantitative common information, and through abstracting and generalizing quantitative attributes, concise and easily-understood conclusions with a certain degree of precision are reached. Therefore, fuzzy data mining has shown its effectiveness in financial data prediction and traffic flow control, as well as in medical diagnosis [18–20].
- (4) The research on fuzzy temporal data mining is mainly divided into two classes, namely time fuzzy and attribute feature fuzzy. More research focused on the fuzziness while taking into account the time attribute of the dataset, i.e., the fuzzy treatment of time. For example, Mazarbhuiya et al. uncovered interesting time-dependent patterns using Tire algorithm to deal with a dataset with fuzzy trading times [21,22]. The fuzziness also works on time constraints, especially the backward and forward association constraints of items. Fuzzy temporal logic is used to deal with time-constrained incomplete information, and performs mining, for example, fuzzy periodic association rules for items [23–25]. Additionally, fuzziness works on the duration of rules, where the association rules may only hold for a fuzzy period of time instead of the entire data set [22,26].
- (5) Besides time fuzzy, another research point is that temporal data-mining models introduce fuzziness into attributes [17,27]. When mining temporal databases with quantitative attributes, fuzzy sets can be used to represent their quantitative values, thus defining smoother transitions between adjacent values [24]. A common operation mode is to convert non-temporal quantitative values from temporal data into symbols or linguistic terms to make it easier for further analysis and interpretation. Many other researchers have also contributed to the exploration of suitable fuzzy affiliation functions [28–30] and mining algorithms [31] for fuzzy temporal mining. Some examples are neuro-fuzzy systems [32,33] and fuzzy clustering [34] that combine intensive machine learning.

In summary, the time or temporal fuzzy data mining mainly involves two models of temporal fuzzy and attribute feature fuzzy. Therefore, based on information granularity reconstruction and combining fuzzy mathematics and temporal approaches, this paper reconstructed time-varying data sets in a multi-dimensional and multi-granularity fashion by simulating human perception. In this paper, a fuzzy temporal event was defined, associated properties of the fuzzy temporal approach as well as operations were given, and a fuzzy temporal event association rule mining model and algorithm were proposed to discover the knowledge and patterns hidden in the raw time-varying data. Numerical experiments indicated that the fuzzy temporal event mining model can detect the correlations hidden in daily trading events, which proves its effectiveness.

This paper presents three main innovations: (1) defining the fuzzy temporal type and fuzzy temporal granularity; (2) building a fuzzy temporal event model; and (3) proposing a fuzzy temporal event association rule model and algorithm.

Section 2 defines fuzzy temporal events and their operational rules and the associated properties. Section 3 constructs a fuzzy temporal event mining model based on fuzzy temporal events. In Section 4 stock trading data on the Lung Fu Bong (a Chinese stock daily winners list) is mined to show the effectiveness of the proposed fuzzy temporal even-association mining model.

2. Definition and Arithmetic of the Fuzzy Temporal Event

In this section, the definition and operations of the fuzzy temporal event are given to reveal features of temporal data streams at different temporal granularities. To better illustrate these, stocks were taken as an example.

Definition 1. Let $A = \{A_l\} (l = 1, 2, \dots, m)$ a finite set of m objects. Let $S = \{S_1, S_2, \dots, S_p\}$, $p \geq 1$ be the set of attributes of the objects containing a subset of p specific attributes. The set of state variables corresponding to attribute S_i is defined as $\bar{S}_i = \{s_{ij}\} (i = 1, 2, \dots, p; j = 1, 2, \dots, q_i)$, where s_{ij} are highly fuzzy due to decisions made by humans.

Definition 2. The fuzzy threshold interval corresponding to each attribute state variable s_{ij} is defined as $[d_{s_{ij}}, \bar{d}_{s_{ij}}]$ (s can be omitted), where $d_{s_{ij}}$ is the worst and $\bar{d}_{s_{ij}}$ is the best for s_{ij} . Let $D(S_i) = \left\{ [d_{s_{ij}}, \bar{d}_{s_{ij}}] \right\}$ be called the set of fuzzy intervals corresponding to s_{ij} of attribute S_i . When fuzzy state variable S_i takes an actual value d_i , fuzzy membership function $\theta_{s_{ij}} : R \rightarrow [0, 1]$ at state s_{ij} is defined by:

$$\theta_{s_{ij}}(d_i) = \begin{cases} \frac{d_i - d_{s_{ij}}}{\bar{d}_{s_{ij}} - d_{s_{ij}}}, & \text{if } d_i \in [d_{s_{ij}}, \bar{d}_{s_{ij}}] \\ 0, & \text{if } d_i \notin [d_{s_{ij}}, \bar{d}_{s_{ij}}] \end{cases} \quad (1)$$

Definitions 1 and 2 describe the state transformation degree of a transaction attribute. The following example illustrates the abovementioned concepts.

Example 1. The price of stock A_1 has three attributes $S = \{\text{Open price, Close price, Trading volume}\}$ in the stock market. Trading volume, an attribute of stock, is often described by vague conceptual terms such as $\{\text{Low, Medium, High}\}$ to depict its states. Corresponding the states Low, Medium, and High to $j = 1, 2, 3$, VOL is depicted as $\bar{S}_1 = \{s_{11}, s_{12}, s_{13}\}$ by Definition 1. In addition, stock prices, another attribute of stock, are often depicted as $\{\text{Rise, Fall, Unchanged}\}$ or $\{\text{Great Rise, Small Rise, Shock, Small Drop, Large Drop}\}$.

Furthermore by Definition 2, when s_{11}, s_{12}, s_{13} are defined as low, medium, and high VOL, respectively, and let its corresponding fuzzy interval $[d_{s_{1j}}, \bar{d}_{s_{1j}}]$ be abbreviated as $[d_{1j}, \bar{d}_{1j}]$ ($j = 1, 2, 3$), such as $[d_{11}, \bar{d}_{11}] = [0, 25]$; for s_{12} -medium volume, $[d_{12}, \bar{d}_{12}] = [20, 45]$; for s_{13} -high volume, $[d_{13}, \bar{d}_{13}] = [40, 60]$. By (1), for a VOL of A_1 , when $d_i = 45$, $\theta_{s_{11}}(d_i) = 0$, which means that a VOL of 45 is associated with 0% of the state Low. The value of $\theta_{s_{12}}(d_i) = 1$ indicates that a VOL of 45 is 100% in the medium state. A value of $\theta_{s_{13}}(d_i) = 0.25$ means a VOL of 45 is 25% in the high state. Then, a conclusion that stock A_1 is currently experiencing medium-high trading volume is drawn.

In the real world, everything changes over time. Let H be the time variable, i.e., a temporal dimension added to the description of object A and the attribute S . To further explore the interactions among (A, S, H) , it is necessary to define temporal type and some associated concepts as follows:

Definition 3. For any absolute moment $t (t \in R)$, when the following conditions are satisfied: (1) non-emptiness: $t \in v(t)$; (2) monotonicity: if $t \in v(t)$ and $v(t_1) \cap v(t_2) = \emptyset$, for any $t' \in v(t_1)$ and $t'' \in v(t_2)$, there is $t' < t''$, and hence $v(t_1) < v(t_2)$; (3) homogeneity: for any $t' \in v(t)$, there is $v(t') = v(t)$; (4) boundedness: for any $t' \in v(t)$, there is $|t'| < +\infty$, and then v is defined as a temporal type and $v(t)$ is called the temporal factor of the temporal type v [15]. A temporal type is called temporal granularity when it has a temporal factor of the same length.

In Definition 3, for example, second, minute, hour, day, month, and year are the most common temporal types. A specific day is a day temporal factor if day is the temporal type. Additionally, quarter of an hour, seven hours, five days, week, quarter of a year, and other time intervals can be temporal types. Temporal type is the concept of temporal factors, and temporal granularity is the concept of time factors with the same length, for example, day and hour, are temporal granularity, but month is temporal type, not temporal granularity. Definition 1 are all temporal types. Additionally, when raw datasets are reconstructed from

different temporal type perspectives, new patterns may be revealed, which may broaden and deepen temporal data mining.

Property 1. Given an absolute time period $[T, T']$, and the temporal type v which divides the time period $[T, T']$ into n segments, where T is the starting point and T' is the ending point, then there exists $t_1, t_2, \dots, t_n$, such that $[T, T'] = \bigcup_{k=1}^n v(t_k)$ holds, where $v(t_i) \cap v(t_j) = \emptyset$ $i \neq j, i, j = 1, 2, \dots, n, t_1 < t_2 < \dots < t_n$.

Property 2. For $t \in R$ and any temporal factor $v(t)$, there exists an upper bound $\sup(v(t))$ and a lower bound $\inf(v(t))$. From Definition 1, the length of the temporal factor $v(t)$ is $\text{length}(v(t)) = \sup(v(t)) - \inf(v(t))$. When $[T, T'] = \bigcup_{k=1}^n v(t_k)$, the value of n determines the value of $\text{length}(v(t))$, so a larger value indicates a smaller temporal granularity and vice versa.

Therefore, different granularity of the temporal type means a huge difference when describing the features of even the same period of time, by Property 2.

For example, if a factory operates on an hourly basis, there is a temporal length of one hour, while for a two-shift factory there is a temporal length of twelve hours and for a three-shift manufacturer there is a temporal length of eight hours.

Definition 4. For temporal type v at given time interval $[T, T']$ and the fuzzy features of objects A and attribute set S , $(A, S, [T, T']v)$ is defined as the space of fuzzy temporal events in $[T, T']$. A specific fuzzy temporal event can be denoted as $(A_l, s_{ij}, d_i(t_k), v(t_k))$, which means a fuzzy state of the event at $d_i(t_k)$ of s_{ij} for temporal factor $v(t_k)$ of object A_l . The event function is defined by:

$$E(A_l, s_{ij}, d_i(t_k), v(t_k)) = \theta_{s_{ij}}(d_i(t_k)) \quad (2)$$

and various states for the same temporal factor at the same attribute state are defined by:

$$E(A_l, \bar{S}_i, v(t_k)) = [\theta_{s_{ij}}(d_i(t_k))]_{1 \times j} \quad (3)$$

Then E is called the event function. Its value is determined by $\theta_{ij}(d_i(t_k))$, and obtained by computing the attribute state value $d_i(t_k)$ with the set $[\underline{d}_{ij}, \bar{d}_{ij}]$ corresponding to the state s_{ij} .

Example 2. In Definition 4, when the VOL of stock A_l is 45 at 13:01 on 8 October 2020, with the temporal type of minute, E is calculated: $E(A_l, s_{11}, 45, 13:01) = \theta_{11}(45) = 0$, $E(A_l, s_{12}, 45, 13:01) = \theta_{12}(45) = 1$, $E(A_l, s_{13}, 45, 13:01) = \theta_{13}(45) = 0.25$. In terms of attribute VOL, this event can be noted as $E(A_l, \bar{S}_1, 13:01) = (0, 1, 0.25)$, which means the event (A_l , medium to high in trading volume, 25% of High, 13:01) occurs.

Property 3. Given $[T, T'] = \bigcup_{k=1}^n v(t_k)$, then the sequence of fuzzy temporal events $(A_l, s_{ij}, d_i(t_k), v(t_k))$ forms of the following fuzzy temporal matrix:

$$F_{A_l S_\theta}(t) = \begin{bmatrix} E(A_l, \bar{S}_i, v(t_1)) \\ E(A_l, \bar{S}_i, v(t_2)) \\ \vdots \\ E(A_l, \bar{S}_i, v(t_n)) \end{bmatrix} = \begin{bmatrix} \theta_{i1}(d_i(t_1)) & \theta_{i2}(d_i(t_1)) & \cdots & \theta_{iq_i}(d_i(t_1)) \\ \theta_{i1}(d_i(t_2)) & \theta_{i2}(d_i(t_2)) & \cdots & \theta_{iq_i}(d_i(t_2)) \\ \vdots & \vdots & \cdots & \vdots \\ \theta_{i1}(d_i(t_n)) & \theta_{i2}(d_i(t_n)) & \cdots & \theta_{iq_i}(d_i(t_n)) \end{bmatrix} \quad (4)$$

Example 3. When n is given a different value, a different set of fuzzy temporal events is generated for the same time period. Again take stocks. For attribute VOL of stock, three fuzzy attribute states i.e., Low, Medium and High, are used, with its fuzzy attribute states as $D = \{\text{Low}, \text{Medium}, \text{High}\}$ and the corresponding set of fuzzy state intervals: $D = \{[\underline{d}_{11}, \bar{d}_{11}], [\underline{d}_{12}, \bar{d}_{12}], [\underline{d}_{13}, \bar{d}_{13}]\} = \{[0, 25], [20, 45], [40, 60]\}$. If VOL of stock A_l for the past three days is 20, 30, and 45, respectively, a fuzzy temporal event sequence $(A_l, s_{ij}, 20, v(t_1)), (A_l, s_{ij}, 30, v(t_2)), (A_l, s_{ij}, 45, v(t_3))$ is created. Three fuzzy membership degrees combined in one temporal factor for each fuzzy temporal event are as follows: $E(A_l, \bar{S}_1, 20, v(t_1)) = (0.8, 0, 0)$, $E(A_l, \bar{S}_1, 30, v(t_2)) = (0, 0.4, 0)$, $E(A_l, \bar{S}_1, 45, v(t_3)) = (0, 1, 0.25)$. These events are combined to form a fuzzy state matrix:

$$F_{A_i \text{vol}}(t) = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 1 & 0.25 \end{bmatrix}$$

Definition 5. For a temporal event $E(A_i, s_{ij}, d_i(t_k), v(t_k))$, let a fuzzy threshold α_i ($\alpha_i \in [0, 1]$) exist for the corresponding attribute S_i ; when $\theta_{s_{ij}}(d_i(t_k)) > \alpha_i$ and $E(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k)) = 1$, it means s_{ij} of S_i at temporal factor $v(t_k)$ occurs, otherwise $E(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k)) = 0$ means it does not occur. So the occurrence of a fuzzy state event can be recorded by the occurring attribute state s_{ij} after code conversion (s can be omitted in expressions), while non-occurrence is recorded as 0:

$$\tilde{E}(A_i, \bar{S}_i(\alpha_i), v(t_k)) = [s_{ij} \text{ or } 0]_{1 \times j} = [ij \text{ or } 0]_{1 \times j} \quad (5)$$

Example 4. In Definition 5, assuming the fuzzy threshold α_i being 0.25 for attribute VOL, for the above three days, the fuzzy temporal events are transcribed as follows: $\tilde{E}(A_i, \bar{S}_i, 20, v(t_1)) = (11, 0, 0)$, $\tilde{E}(A_i, \bar{S}_i, 30, v(t_2)) = (0, 12, 0)$, $\tilde{E}(A_i, \bar{S}_i, 45, v(t_3)) = (0, 12, 13)$, with an updated fuzzy state code matrix:

$$\tilde{F}_{A_i \text{vol}}(t) = \begin{bmatrix} 11 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 12 & 13 \end{bmatrix}$$

Definition 6. Let $\text{Sup}(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k))$ be the support rate (or frequency) of s_{ij} of object A_i for the temporal state type v over the time period $[T, T']$:

$$\text{Sup}(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k)) = \frac{1}{n} \sum_{k=1}^n E(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k)) \quad (6)$$

Given a frequent threshold β ($\beta \in [0, 1]$), if $\text{Sup}(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k)) > \beta$, then $(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k))$ is said to be a frequent fuzzy temporal event with respect to threshold α_i over time period $[T, T']$, and the validity value of the event occurring during that time period is $\text{Sup}(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k))$.

Definition 6 depicts the correlation or connection of events in different temporal factors.

Example 5. In Definition 6, with $\tilde{F}_{A_i \text{vol}}(t)$, a further calculation of $\text{Sup}(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k))$ assumes the support rate of Low, Medium, and High as 1/3, 2/3, and 1/3 for a VOL of A_i , respectively. When $\beta = 0.5$, then $(A_i, (\text{medium trading volume}), \text{day})$ is a frequent fuzzy temporal event during the three days.

Property 4. Given $(A_i, s_{ij}(\alpha_i), d_i(t_k), v(t_k))$, a frequent fuzzy temporal event about α_i over time period $[T, T']$, when there exists $\alpha_{in} \in [0, 1]$ and $\forall \alpha_{in} \leq \alpha_i$, the event is also a frequent fuzzy temporal event about α_{in} over time period $[T, T']$.

Definition 7. With two fuzzy temporal events $(A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i}))$, $(A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))$ in $(A, S, [T, T']v)$, the operations of fuzzy events in the fuzzy temporal event space are as follows:

- (1) The operation "AND", notated as " $\wedge$ ": If both $(A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i}))$ and $(A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))$ occur, then $E((A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i})) \wedge (A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))) = 1$, which indicates that fuzzy temporal events $(A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i}))$ and $(A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))$ are logically AND correlated on α_i and α_p .
- (2) The operations "OR", notated as " $\vee$ ": If $(A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i}))$ or $(A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))$ occurs, then $E((A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i})) \vee (A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))) = 1$, which indicates fuzzy temporal events $(A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i}))$ and $(A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))$ are logically OR correlated on α_i and α_p .

Definition 7 depicts the correlation or connection between two different fuzzy temporal events. Through Definition 7 we obtain the following results.

Property 5. There exists $\alpha_{i\wedge} = \min(\alpha_i, \alpha_p)$, satisfying $\forall \alpha \leq \alpha_{i\wedge}$, which makes

$$E((A_m, s_{iq_i}(\alpha), d_i(t_{k_i}), v(t_{k_i})) \wedge (A_n, s_{pq_p}(\alpha), d_p(t_{k_p}), v(t_{k_p}))) = 1$$

Additionally, if $\forall \alpha \geq \alpha_{i\wedge}$,

$$E((A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i})) \wedge (A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))) = 0$$

The threshold for the logical AND relationship between events $(A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i}))$ and $(A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))$ is $\min(\alpha_i, \alpha_p)$.

Property 6. There exists $\alpha_{i\vee} = \max(\alpha_i, \alpha_p)$ satisfying $\forall \alpha \leq \alpha_{i\vee}$ and making

$$E((A_m, s_{iq_i}(\alpha), d_i(t_{k_i}), v(t_{k_i})) \wedge (A_n, s_{pq_p}(\alpha), d_p(t_{k_p}), v(t_{k_p}))) = 1$$

Additionally, if $\forall \alpha \geq \alpha_{i\vee}$, $E((A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i})) \vee (A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))) = 0$.

The threshold for the logical OR relationship between events $(A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i}))$ and $(A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))$ is $\max(\alpha_i, \alpha_p)$.

Axiom 1. In $(A, S, [T, T']v)$, for any three fuzzy temporal events $X = (A_m, s_{iq_i}(\alpha_i), d_i(t_{k_i}), v(t_{k_i}))$, $Y = (A_n, s_{pq_p}(\alpha_p), d_p(t_{k_p}), v(t_{k_p}))$, $Z = (A_l, s_{oq_o}(\alpha_o), d_o(t_{k_o}), v(t_{k_o}))$, the following hold:

Exchange law: $X \wedge Y = Y \wedge X$, $X \vee Y = Y \vee X$

Combination law: $X \wedge (Y \wedge Z) = (X \wedge Y) \wedge Z$, $X \vee (Y \vee Z) = (X \vee Y) \vee Z$

Distribution ratio: $X \wedge (Y \vee Z) = (X \wedge Y) \vee (X \wedge Z)$, $X \vee (Y \wedge Z) = (X \vee Y) \wedge (X \vee Z)$

If $E(X \wedge Y) = 1$, it is denoted as $X \xrightarrow{\alpha_i|\alpha_p} Y$ called the fuzzy temporal association rule. Moreover, there are following properties:

Property 7. If $E(X \wedge Y) = 1$, then $E(X) = 1$, $E(Y) = 1$.

Property 8. If $X \xrightarrow{\alpha_i|\alpha_p} Y$, then $X \xrightarrow{\min(\alpha_i, \alpha_p)} Y$, $Y \xrightarrow{\min(\alpha_i, \alpha_p)} X$.

Property 9. If $X \xrightarrow{\alpha_i|\alpha_p} Y$, $Y \xrightarrow{\alpha_p|\alpha_o} Z$, then $X \xrightarrow{\alpha_i|\alpha_p|\alpha_o} Z$, $X \xrightarrow{\min(\alpha_i, \alpha_p, \alpha_o)} Z$.

The definitions and properties of the abovementioned fuzzy temporal event model can be used to study the correlation between pricing, trading volume, etc. In Section 3, a fuzzy temporal event mining model and algorithm are proposed.

3. Fuzzy Temporal Event Mining Model and Algorithm

Based on the abovementioned, it is possible to describe the behavior of more complex fuzzy temporal events so as to find associations between them.

3.1. Fuzzy Temporal Event Association Rule Model

Definition 8. Let $r \geq 0, h \geq 0$ and be integers, and let $[T, T'] = \bigcup_{k=1}^n v(t_k)$, where $v(t_{k+r}), v(t_{k+r+1}), \dots, v(t_{k+r+h})$ are adjacent temporal factors and they are of an increasing relationship: $v(t_{k+r}) < v(t_{k+r+1}) < \dots < v(t_{k+r+h})$. $(A_n, s_{pq_p}(\alpha_p), d_p(t_{k+r}), v(t_{k+r})) \vee (A_n, s_{pq_p}(\alpha_p), d_p(t_{k+r+1}), v(t_{k+r+1})) \vee \dots \vee (A_n, s_{pq_p}(\alpha_p), d_p(t_{k+r+h}), v(t_{k+r+h}))$ is abbreviated as $(A_n, s_{pq_p}(\alpha_p), d_p(t_k), v(t_k))_r^h$ or $(A_n, s_{pq_p}(\alpha_p))_r^h$.

Define an operation:

$$\text{For } (A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \wedge \left[(A_n, s_{pq_p}(\alpha_p), d_p(t_{k+r+1}), v(t_{k+r+1})) \vee \dots \vee (A_n, s_{pq_p}(\alpha_p), d_p(t_{k+r+h}), v(t_{k+r+h})) \right],$$

i.e., $(A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \wedge (A_n, s_{pq_p}(\alpha_p), d_p(t_k), v(t_k))_r^h$,

$E((A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \wedge (A_n, s_{pq_p}(\alpha_p), d_p(t_k), v(t_k))_r^h) = 1$ indicates when event $(A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k))$ occurs, one of the subsequent events $(A_n, s_{pq_p}(\alpha_p), d_p(t_{k+r}), v(t_{k+r})), \dots, (A_n, s_{pq_p}(\alpha_p), d_p(t_{k+r+h}), v(t_{k+r+h}))$ after temporal factor r would occur, i.e., the fuzzy

temporal event association rule $(A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) (A_n, s_{pq_p}(\alpha_p), d_p(t_k), v(t_k)_r^h)$ holds. Otherwise, $E((A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \wedge (A_n, s_{pq_p}(\alpha_p), d_i(t_k), v(t_k)_r^h)) = 0$ and the rule does not hold.

Definition 9. Let $N_{[T, T']}^v(A_l, s_{ij}(\alpha_i))$ denote the number of repetitions of event $(A_l, s_{ij}(\alpha_i), d_i(t_k), v(t_k))$ occurring in the time period $[T, T']$:

$$N_{[T, T']}^v(A_l, s_{ij}(\alpha_i)) = \sum_{k=1}^n E(A_l, s_{ij}(\alpha_i), d_i(t_k), v(t_k)) \quad (7)$$

Similarly, the number of recurrences of rule $(A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \xrightarrow{\alpha_p | \alpha_o} (A_n, s_{pq_p}(\alpha_p), d_i(t_k), v(t_k)_r^h)$ in time period $[T, T']$ is expressed as:

$$N_{[T, T']}^v \left((A_m, s_{iq_i}(\alpha_i)) (A_n, s_{pq_p}(\alpha_p))_r^h \right) = \sum_{k=1}^n E \left((A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \wedge (A_n, s_{pq_p}(\alpha_p), d_p(t_k), v(t_k)_r^h) \right) \quad (8)$$

The occurrence of such a correlation is common in life. For example, prices of stocks rise sharply to form a high opening price, followed by a high probability of the price falling within a few days afterwards.

To mark the time interval during which event $(A_l, s_{ij}, d_i(t_k), v(t_k))$ frequently occurs in $[T, T']$, the concept of the lifespan of an event is defined.

Definition 10. Lifespan is defined as the period between the first temporal factor $v(t_{begin}^{\alpha_i})$ and the last temporal factor $v(t_{end}^{\alpha_i})$ when event $(A_l, s_{ij}, d_i(t_k), v(t_k))$ occurs in the database in $[T, T']$, and is denoted as $[v(t_{begin}^{\alpha_i}), v(t_{end}^{\alpha_i})]$ with its length $\phi(A_l, s_{ij}(\alpha_i), v)$ where $(v(t_{begin}^{\alpha_i}) \leq v(t_{end}^{\alpha_i}))$. The lifespan length of $(A_l, s_{ij}(\alpha_i), d_i(t_k), v(t_k))$ is calculated as follows: $len(\phi(A_l, s_{ij}(\alpha_i), v)) = v(t_{end}^{\alpha_i}) - v(t_{begin}^{\alpha_i})$, with the unit of measurement being temporal type v .

For example, when there is a great rise in the opening price (at an association of 0.6) for stock A_l on 20 October and another rise on 20 December during the observation period October to December, if day is chosen as the temporal type, then the lifespan length of the event $(A_l, \text{Opening Price}_{\text{great rise}}(0.6), \text{day})$ is 61 days.

Corollary 1. Starting temporal factor $v(t_{begin}^{\alpha_i})$ and ending temporal factor $v(t_{end}^{\alpha_i})$ of the lifespan of $(A_l, s_{ij}(\alpha_i), d_i(t_k), v(t_k))$ for the same temporal type in $[T, T']$ are determined by the threshold α_i . The length of lifespan is determined similarly.

Corollary 2. When $(A_l, s_{ij}(\alpha_i), d_i(t_k), v(t_k))$ occurs in $[T, T']$ at α_i , there exists $\alpha_{in} \in [0, 1]$, $\forall \alpha_{in} \geq \alpha_i$, making $len(\phi(A_l, s_{ij}(\alpha_i))) \geq len(\phi(A_l, s_{ij}(\alpha_{in})))$. Similarly there exists $\alpha_{in} \in [0, 1]$, $\forall \alpha_{in} \leq \alpha_i$, making $len(\phi(A_l, s_{ij}(\alpha_i))) \leq len(\phi(A_l, s_{ij}(\alpha_{in})))$.

For the above stock A_l , when the threshold is reduced from 0.6 to 0.4, the first date to observe great rise in opening price will probably be earlier, i.e., on or before 20 October; similarly, the last day to observe another great rise will most likely be later, i.e., on or after 20 December. Then lifespan of the event is conceivably extended. Similarly, when the affiliation threshold increases, in contrast, the lifespan may be shortened.

Definition 11. With the fuzzy temporal event association rule $(A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \xrightarrow{\alpha_i | \alpha_p} (A_n, s_{pq_p}(\alpha_p), d_p(t_k), v(t_k)_r^h)$, abbreviated as $(A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h$, $Sup \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h \right)$ is defined as the support of the fuzzy temporal event association rule $(A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h$ in $[T, T']$:

$$\begin{aligned} & \text{Sup} \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h \right) \\ &= \frac{N_{[T, T']}^v \left((A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \wedge (A_n, s_{pq_p}(\alpha_p), d_p(t_k), v(t_k))_r^h \right)}{\text{len}(\phi(A_m, s_{iq_i}(\alpha_i), v))} \end{aligned} \quad (9)$$

$\text{Conf} \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h \right)$ is defined as the confidence level of the fuzzy temporal event association rule $(A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h$ in $[T, T']$:

$$\begin{aligned} & \text{Conf} \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h \right) \\ &= \frac{N_{[T, T']}^v \left((A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \wedge (A_n, s_{pq_p}(\alpha_p), d_p(t_k), v(t_k))_r^h \right)}{N_{[T, T']}^v(A_m, s_{iq_i}(\alpha_i))} \end{aligned} \quad (10)$$

The distinctive feature of the fuzzy temporal event association rule is that it has both a fuzzy property and temporal dimension, so the confidence level and support of a fuzzy temporal event association rule can be used to measure the validity of the rule.

Given n the number of v temporal factors in interval $[T, T']$, and let thresholds $\text{minsup}, \text{minconf} \in [0, 1]$ such that

$$\text{Sup} \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h \right) \geq \text{minsup} \quad (11)$$

$$\text{Conf} \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h \right) \geq \text{minconf} \quad (12)$$

The support and confidence level of all fuzzy temporal event association rules on the event space that satisfies the threshold requirement is obtained. Such a situation that satisfies fuzzy temporal event association rule $(A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h$ is common in the stock market. For example, when the price of stock A_j jumps high followed by a low in 2 days after an interval of 3 days and there are 7 such jumps out of all 20 opening high in a year for that stock, the annual validity is 7/20.

For the support and confidence level, clearly the following property holds:

Property 10. If $h_1 > h_2 \geq 0$ and all the others remain unchanged, then

$$\text{Sup} \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^{h_1} \right) \geq \text{Sup} \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^{h_2} \right) \quad (13)$$

$$\text{Conf} \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^{h_1} \right) \geq \text{Conf} \left((A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^{h_2} \right) \quad (14)$$

Definition 12. The effective lifespan of the fuzzy temporal event association rule $(A_m, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_n, s_{pq_p}(\alpha_p))_r^h$ is determined by the temporal interval $[v(t_{begin}^{\alpha_i}), v(t_{end}^{\alpha_i})]$, where $v(t_{begin}^{\alpha_i})$ means the first time when α_i and the fuzzy temporal event association rule are satisfied and $v(t_{end}^{\alpha_i})$ the last time. Similarly, the length of the lifespan is also determined by temporal factor, measured by temporal type v .

3.2. Fuzzy Temporal Event Association Rule Mining Algorithm

The fuzzy temporal event rule mining algorithm (**FTEAM algorithm**) proposed in this paper includes three parts: attribute fuzzification, temporal type transformation, and fuzzy association rule mining.

In real world, the same database depicted by different temporal types may display different amounts of information. Taking stock trading data as an example, for the same trading period, if smaller temporal granularity is taken, it will have more temporal factors, i.e., more total data, which will give a more detailed view of the changing pattern of stock

transaction with more dynamic information and, of course, more noise. In this case it may be possible to extract more features and retain more details of the transaction. In contrast, with larger temporal granularity, trading data are more noise tolerant and more likely to show features and trends. In particular, Ouyang Weiming et al. [35] demonstrated that the validity of association rules relates to temporal type. The choice of temporal type is critical to the validity of association rules. By selecting different temporal types, the database can be temporally transformed and reorganized into different temporal granularities, and correlation is mined at each granularity while considering commonality and individuality. The mining steps and algorithm are as follows.

The FTEAM algorithm:

Step 1 For one or more objects in $A = \{A_1, A_2, \dots, A_m\}$, select specific attributes $S_1, S_2, \dots, S_p$ to mine, based on the object's features.

Step 2 Based on practical experience and circumstances, select one or more suitable temporal types and transform and reorganize the raw data to form temporal events.

Step 3 Based on the actual situation of specific attributes and mining requirements, predetermine the number of k of the attribute state. Assume that domain $U = [D_{\min}, D_{\max}]$ covers the values of attribute S_i in all temporal events of the specific temporal type in $[T, T']$, where $D_{\min}$ and $D_{\max}$ are respectively the minimum and maximum values of $d_i(v(t))$. With the K-means clustering algorithm clustering values of the selected attribute with a pre-determined number k , k cluster centroids $o_i (i = 1, 2, \dots, k)$ are obtained, on which basis the domain of the attributes is delimited into k attribute state intervals $D(S_i) = \{[d_{s_{ij}}, \bar{d}_{s_{ij}}]\}$, ($j = 1, 2, \dots, k$).

Step 4 The midpoint $\frac{o_{k-1} + o_k}{2}$ of two adjacent clustering centers o_{k-1}, o_k is taken as the boundary point between two adjacent state intervals. Considering the fuzziness of the attribute state, set the proportion of interval cross-boundary part to 2c% (c-value is the extension ratio of the boundary point), then $\underline{d}_{ij} = (1 - c\%) \times \frac{o_{k-1} + o_k}{2} = \underline{D}_{k-1}$ and $\bar{d}_{ij} = (1 + c\%) \times \frac{o_{k-1} + o_k}{2} = \bar{D}_{k-1}$. Combining $D_{\min} = \underline{D}_0$ and $D_{\max} = \bar{D}_k$, the attribute domain is divided into k intersecting attribute state subintervals $(\underline{D}_0, \bar{D}_1), (\underline{D}_1, \bar{D}_2), \dots, (\underline{D}_{k-1}, \bar{D}_k)$, which correspond to the fuzzy states $s_{i1}, s_{i2}, \dots, s_{ik}$ of the attribute S_i .

Step 5 Select the fuzzy membership function, and calculate the fuzzy membership degree $\theta_{ij}(t_n)$ of the attribute value of each temporal factor on each fuzzy attribute state interval in $[T, T']$, to obtain the new fuzzy temporal state event $E(A_l, \bar{S}_i, v(t_k))$. Then set the fuzzy threshold α_i to transform events into a fuzzy temporal event set $\tilde{E}(A_l, \bar{S}_i(\alpha_i), v(t_k))$.

Step 6 Apriori algorithm [24,35] is applied to mine possible frequent sets and association rules among the fuzzy temporal event sets in Steps 6.1 to 6.3.

Step 6.1 Scan the fuzzy temporal event set $\tilde{E}(A_l, \bar{S}_i(\alpha_i), v(t_k))$ to obtain a set of non-duplicate fuzzy temporal events, noted as $C1 = [(A_l, s_{ij}(\alpha_i), d_i(t_k), v(t_k))]$ (called pre-1 sized frequent set). Scan the event set again and count the occurrences of all fuzzy events in $C1$ to calculate their $N_{[T, T']}^v(A_l, s_{ij}(\alpha_i), d_i(t_k))$. Remove the items from $C1$ that do not achieve the minimum support threshold (*minsup*) by (3.7), thus obtaining the fuzzy state 1-sized frequent set $L1$.

Support:

$$Sup(A_l, s_{ij}(\alpha_i), d_i(t_k), v(t_k)) = \frac{1}{n} N_{[T, T']}^v(A_l, s_{ij}(\alpha_i), d_i(t_k)) \geq minsup \quad (15)$$

Step 6.2 Set $C2$ (pre-2 frequent sets) containing two states is formed after combination and permutation of frequent set $L1$. The AND operation “ $\wedge$ ” is performed on all fuzzy temporal events two-by-two, scanning all combinations of events and counting the occurrences of items in $C2$. The items that do not achieve the *minsup* are dropped, resulting in the 2-sized frequent set $L2$ of fuzzy jointly state events. Repeat the above steps to generate a 3-sized pre-frequent set $C3$ based on the permutation of $L2$, and then obtain the 3-sized frequent set $L3$ by *minsup* value deletion, and so on until the p -sized frequent set Lp is obtained as a cut-off.

Step 6.3 Finally, based on the frequent sets of L_2 to L_p , confidences are calculated for fuzzy joint state events that occur frequently within the frequent set, and if the minimum support requirement (11) and minimum confidence threshold (12) are met, the association rule holds.

Since the number of fuzzy temporal events is limited, the FTEAM algorithm stops after a finite step operation.

4. Numerical Experiments and Analysis of Results

To demonstrate the effectiveness of fuzzy-temporal data association rule mining model and algorithm (FTEAM algorithm), in this part, intra-day trading data of stocks on Lung Fu Bong are mined.

4.1. Description of the Experimental Data

Stocks on Lung Fu Bong are identified as typical outliers in the Chinese stock market, which meet any of the following criteria and should to be one of the top 3: (1) daily price deviation: $\pm 7\%$; (2) daily turnover rate: 20%; (3) daily price amplitude: 15%; and (4) cumulative price deviation in three consecutive trading days: 20%. So when any stock is on the list, it indicates there is an obvious anomaly. The period from 1 October 2020 to 31 December 2020 was selected as the statistical time period. For the validity of the sample it is necessary to select the most representative stocks of *lung fu bong*. Therefore, Lanzi (002612.SZ) was selected, which was listed 36 times during 10 November 2020–31 December 2020, very frequently, indicating the complexity and irregularity of the selected temporal data.

Five temporal types were selected, with temporal granularity from small to large, including 1 min temporal, 5 min temporal, 10 min temporal, 15 min temporal, and 30 min temporal, based on which two comparison groups were formed, i.e., 5 min apart and 15 min apart. K-line data for 002612.SZ is exported from Comstar Financial Terminal for the period 2 November 2020–31 December 2020. By FTEAM algorithm, Python 3.8 and Excel 2016 were used to handle attribute states, dimensionless normalization, and code conversion of raw stock data. Table 1 illustrates fuzzy intervals and corresponding codes of the attribute states, where the first five attributes show the comparison of adjacent longitudinal temporal factors, and the last three reflect the inner trend within a transverse temporal factor.

Table 1. The fuzzy intervals and corresponding codes of the attribute states.

Attribute States and Codes	$[-0.1, -0.124]$ (s_{i1})	$[-0.68, -0.002]$ (s_{i2})	$[-0.124, 0.086]$ (s_{i3})	$[-0.002, 0.52]$ (s_{i4})	$[0.086, 0.1]$ (s_{i5})
Opening Price S_1	Great Drop (11)	Small Drop (12)	Shock (13)	Small Rise (14)	Great Rise (15)
Highest Price S_2	Great Drop (21)	Small Drop (22)	Shock (23)	Small Rise (24)	Great Rise (25)
Lowest Price S_3	Great Drop (31)	Small Drop (32)	Shock (33)	Small Rise (34)	Great Rise (35)
Closing Price S_4	Great Drop (41)	Small Drop (42)	Shock (43)	Small Rise (44)	Great Rise (45)
Volume Trade S_5	Low (51)	Moderate Drawdown (52)	Flat (53)	Moderate Release (54)	Rich (55)
Price Trends S_6	Great Drop (61)	Small Drop (62)	Shock (63)	Small Rise (64)	Great Rise (65)
Price Amplitude S_7	Smooth (71)	Slight Shock (72)	Wave Shock (73)	Visible Shock (74)	Wide Shock (75)
Turnover Rate S_8	Lower (81)	Low (82)	Medium (83)	High (84)	Higher (85)

Our first experiment was to mine the association rules in the raw daily data of Lanzi. After code-transforming the daily data using the same intervals, a total of 85 1-sized association rules were mined using Apriori algorithm, yet all of which implied a traversal correlation between the extreme states (Great Rise, Great Drop) of eight attributes, which has no real meaning. Due to the fact that the number of data mined was too small at only 43 items, the experiment was extended to the full year of 2020, where the association rules mined were still insubstantial. So, this is to say it is not feasible to find trade rules just based on raw daily data of stocks.

4.2. Results and Analysis

The FTEAM algorithm was applied to mine the adjacent association rules between different attributes of the same object $(A_l, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \xrightarrow{\alpha_i} (A_l, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)_1^0)$ (a special form as defined by Definition 8), which led to the following conclusions.

4.2.1. Discoverability of Fuzzy Temporal Association Features

Experiments show that these parameter thresholds directly affect the discovery of association rules at each temporal granularity, as conjectured from the nature of affiliation (Properties 4–6) and the definitions of support and confidence (Definitions 9 and 11). The threshold (as shown in Figure 1) flexibly adjusts the mined content. As it decreased from 0.7 to 0.5, the content and number of attribute state features mined under the same temporal type differed greatly, which demonstrates the necessity of mining data at different levels of granularity. Figure 1 and Table 2 indicate that the number of fuzzy temporal association rules decreased as the granularity increased, inferring that large temporal granularity makes it difficult to detect stock anomalies.

Table 2. Number of association rules for each temporal type at different support and confidence levels.

Minsup, Minconf	1 Min	5 Min	10 Min	15 Min	30 Min
0.1, 0.5	200,200	1017	705	597	398
0.1, 0.6	139,122	659	445	384	260
0.2, 0.5	18,450	77	49	39	44

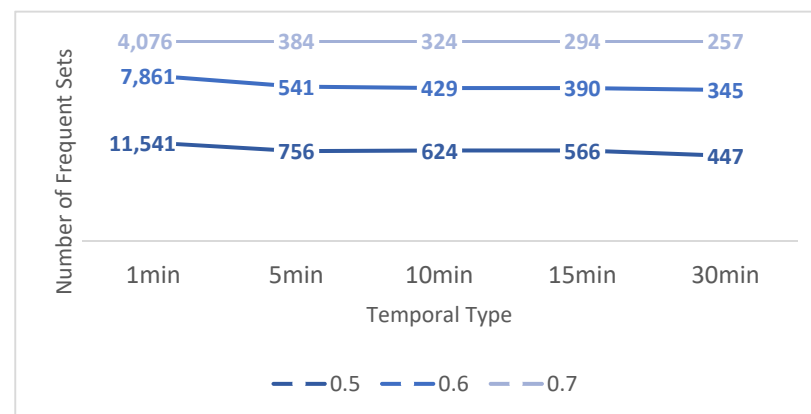


Figure 1. Number of frequent sets for different temporal types at different thresholds.

With all other parameters unchanged, *minsup* rose from 0.1 to 0.2, as shown in rows 1 and 3 of Table 2; additionally, *minconf* rose from 0.5 to 0.6 as shown in rows 1 and 2. It can be concluded that for the same stock trading data (at the same fuzzy affiliation threshold), the main factor affecting the mining of fuzzy temporal event association rules is the *minsup*, while *minconf* plays a role of fine-tuning the number of association rules.

Figure 2 gives the number of the frequent set mined under different temporal types. When small temporal granularity was taken, the attribute states concentrated more obviously in smooth or fluctuating states (right); yet, when large temporal granularity was taken, the attribute states diverged more greatly in great-rise and great-drop state (left). In terms of Lanzi, when temporal granularity changed from small to large, the extreme distribution of attribute states became more apparent. Furthermore, fuzzy temporal trading data at small granularity are more sensitive to changes in stock characteristics, making it easier to detect abnormal trading patterns.

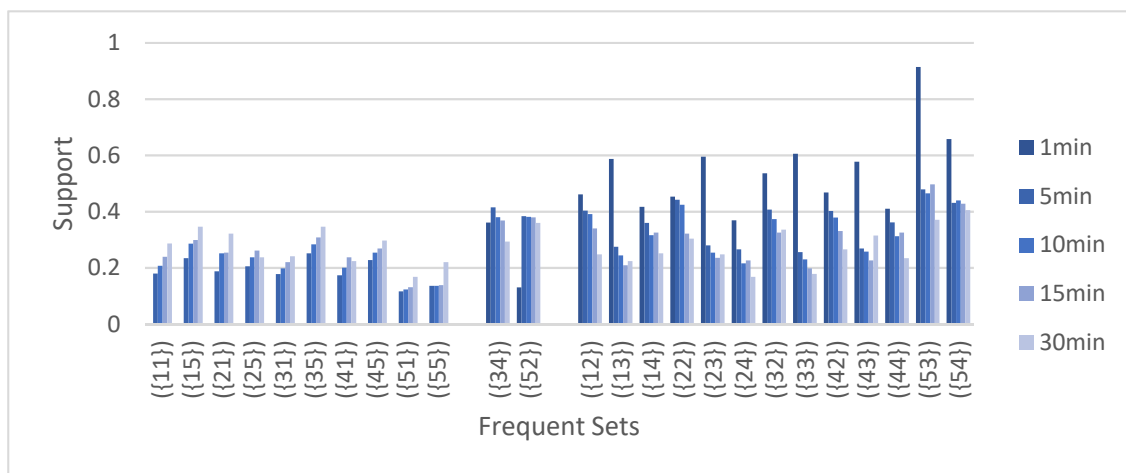


Figure 2. Frequent set distribution at different temporal granularity.

By mining the trading data of Lanzi, the association rules common to all temporal granularity are shown in Table 3, where Rule 1 and 2 (line 1-2) illustrate that the states of the closing price, i.e., rise or drop, can reflect the state of the highest price and VOL. Support for the temporal association rule is higher at smaller granularity, indicating that smaller granularity is more effective in portraying the association between features and that small granularity reveals a stock's anomaly. Rule 3 (line3) says lowest price shock (33) usually accompanies a slight drop in lowest price (32). This implies that when Lanzi's price fluctuates slightly, especially around lowest price, the longer the fight between the long and short sides, and the more likely there will be a downward trend of the price. This is more evident in large granularity temporal type, which indicates that stock trading trends are more specific and visible in large granularity data.

Table 3. Comparison of support for association rules at different temporal granularity of Lanzi.

Pre-Event	Post-Event	1 min	5 min	10 min	15 min	30 min
Closing Price slight drop (42)	Highest Price slight drop (22)	0.649	0.619	0.644	0.542	0.500
Highest Price slight drop (22)	VOL Flat (53)	0.937	0.529	0.563	0.709	0.529
Lowest Price Shock (33)	Lowest Price slight drop (32)	0.572	0.579	0.646	0.660	0.725

4.2.2. Predictability of Fuzzy Temporal Event Association Rules

The inclusion of temporal granularity makes it possible to predict trading events that may be hidden when using standard approaches. Setting the following parameters, $threshold = 0.6$, $minsup = 0.1$, $minconf = 0.6$, experiment, yields 15 association rules between adjacent disparate attribute states $(A_l, s_{iq_i}(\alpha_i)) \xrightarrow{[T, T']^v} (A_l, s_{iq_i}(\alpha_i))_1^0$ (see Table 4). The rules are roughly divided into two categories, one with three rules concentrating on opening price and price trend: (1) small drop in opening price (12) in the previous temporal factor leads to a small drop in price trend (62) in the following temporal factor; (2) a shock in opening price (13) in the previous temporal leads to a stable fluctuation in price trend (63) in the following temporal; (3) a small rise in opening price (14) in previous temporal state corresponds to a small increase in price trend (64) in the following temporal state. These association rules have a relatively high degree of confidence in all temporal types, with validity generally reaching around 0.9, where associations very clearly occur. This suggests that for the five-day temporal type, the state of opening price of the previous temporal factor can be used to predict the state of Lanzi's Price Trend of the following temporal factor. The following 12 rules mean small fluctuations in volume and price in the previous

temporal factor easily lead to a low turnover state immediately following. Further, it demonstrates that volume states in the previous temporal factor are highly relevant to the trend of turnover rate in the immediate following temporal factor.

Table 4. Comparison of confidence for 1-sized fuzzy temporal adjacency association rules.

Association Rules	1 min	5 min	10 min	15 min	30 min
$(002612, s_{12}) \rightarrow (002612, s_{62})$	0.875	0.908	0.909	0.895	0.886
$(002612, s_{13}) \rightarrow (002612, s_{63})$	0.872	0.951	0.927	0.857	0.938
$(002612, s_{14}) \rightarrow (002612, s_{64})$	0.850	0.929	0.924	0.885	0.819
$(002612, s_{12}) \rightarrow (002612, s_{81})$	0.687	0.526	0.524	0.569	0.557
$(002612, s_{13}) \rightarrow (002612, s_{81})$	0.818	0.689	0.609	0.643	0.609
$(002612, s_{14}) \rightarrow (002612, s_{81})$	0.680	0.577	0.602	0.598	0.694
$(002612, s_{22}) \rightarrow (002612, s_{81})$	0.676	0.505	0.530	0.616	0.540
$(002612, s_{23}) \rightarrow (002612, s_{81})$	0.781	0.554	0.538	0.524	0.507
$(002612, s_{24}) \rightarrow (002612, s_{81})$	0.711	0.627	0.606	0.542	0.688
$(002612, s_{32}) \rightarrow (002612, s_{81})$	0.745	0.582	0.573	0.595	0.505
$(002612, s_{33}) \rightarrow (002612, s_{81})$	0.782	0.686	0.663	0.623	0.569
$(002612, s_{34}) \rightarrow (002612, s_{81})$	0.615	0.524	0.548	0.548	0.607
$(002612, s_{43}) \rightarrow (002612, s_{81})$	0.786	0.641	0.550	0.587	0.600
$(002612, s_{44}) \rightarrow (002612, s_{81})$	0.688	0.593	0.614	0.561	0.612
$(002612, s_{53}) \rightarrow (002612, s_{81})$	0.728	0.641	0.678	0.643	0.708
$(002612, s_{54}) \rightarrow (002612, s_{81})$	0.827	0.678	0.632	0.627	0.670

In experiments of raw daily data, these rules do not hold at all. Moreover, it illustrates that FTEAM algorithm is capable of identifying valid correlations between features at multiple temporal granularities, and thus provides appropriate predictions, which is crucial for the identification of some special stocks, such as the stocks on Lung Fu Bong.

4.2.3. Lifespan Validity of Fuzzy Temporal Event Association Rules

It is found that confidence of association rules for the same lifespan (see Definition 10) is different for different temporal granularities. It is not true that the smaller or larger the granularity, the more effective the fuzzy temporal event association rules (see Definition 12). For example, as shown by the fuzzy temporal adjacency association rule (12→62) in Table 5, even though the rule holds in the time period, 2 November 2020–31 December 2020, in all five temporal types, the confidence is different. The results showed that confidence was best at 5 min and 10 min, and association rules were significantly more effective at the 5 min temporal type. In real stock trading, this finding can help grid strategies to find the optimal grid size and achieve maximum grid arbitrage returns with minimum capital. It is also a helpful reference for other intraday quantitative trading strategies.

Table 5. Comparison of confidence for different temporal granularities of rule (12→62).

Temporal Granularity	Length of Active Lifespan: [start, end]	Number of Effective Granulations	Confidence
1 min	[2 November 2020 9:33, 31 December 2020 14:49]	10,306	0.875
5 min	[2 November 2020 10:20, 31 December 2020 14:50]	2052	0.908
10 min	[2 November 2020 10:40, 31 December 2020 10:30]	983	0.909
15 min	[2 November 2020 14:00, 31 December 2020 10:30]	648	0.895
30 min	[2 November 2020 11:00, 31 December 2020 10:30]	327	0.886

4.3. Conclusions of the Data Experiment

By effectively granularizing the raw data from multiple attribute dimensions (including the temporal dimension), the FTEAM algorithm can uncover hidden association rules and phenomena, including:

- (1) Small granularity of minute-level temporal type can mine many valid temporal association rules;
- (2) Large granularity of minute-level temporal type gives a better indication of the trend in the data;
- (3) When different temporal types are taken, it is possible to uncover both common and unique association rules, and common association rules have different lifespans and validity under different temporal types, as well as different meanings;
- (4) The choice of affiliation intervals and affiliation thresholds for fuzzy code transformation affects the discovery of temporal association rules;
- (5) The settings of thresholds for the fuzzy temporal association rule, consisting of those of *minsup* and *minconf*, should be based on the stringency of the realistic need for the validity of rules. Generally, when the settings are small, more rules will be acquired. Larger settings may result in rules not mined. Furthermore, the adjustment of support has a significant impact on the number of association rules mined;
- (6) There exists a very niche association event with low support for the previous event, but a high level of confidence in the overall association rules.

Experimental results show that the fuzzy temporal event mining model has significant academic value in theory and application, mainly insofar as (1) the model can discover the anomalous characteristics of special stocks under different granularity. For example, there is a clear tendency for Lanzi to have minute price swings in the same direction by small margins, which will manifest itself in the form of broad price oscillations over a longer temporal period; (2) the model has significant effectiveness in predicting stock trends at different levels of granularity. For example, it was found that in Lanzi's case previous temporal state of volume and price directly affects the trend of trading volume in the immediate following temporal state; (3) the model and algorithm are capable of discovering frequent features and association rules at various levels of temporal granularity, which, however, cannot be found in conventional data analysis.

5. Conclusions

To efficiently handle temporal data streams, this paper proposed to pre-process and re-structure raw data from a perspective of attribute state fuzziness and temporal type. Fuzzy temporal events were defined, their properties and operations deduced, and a rule mining mode with a fine algorithm was proposed to mine fuzzy temporal event associations. This work extends models for temporal data mining and enhances the performance and quality of temporal data association rule mining algorithms. The original information contained in the data was preserved, and flexible control of its volume and precision through temporal and state granularities is gained when multi-level and all-round data features were extracted and hidden knowledge was discovered from temporal data streams. Experimental results on intraday trading data of stocks on Lung Fu Bong indicate that the model and algorithm are highly sensitive to abnormal stock trading, as it uncovers trends hidden in daily trading data as well as related events. Statistically, these are useful in guiding the decision making of quantitative trading with high intraday frequency.

The fuzzy temporal event mining model mode studied in this paper is primarily concerned with raw data reconstruction. When it comes to temporal types, fuzzy temporal event association rules $(A_m, s_{iq_i}(\alpha_i), d_i(t_k), v(t_k)) \xrightarrow{\alpha_i|\alpha_p} (A_n, s_{pq_p}(\alpha_p), d_i(t_k), v(t_k)_r^h)$ (see Definition 8) can still be further studied in a number of ways:

- (1) In life, a large number of correlations among events do not occur immediately afterwards, as there is a buffer period. The adjustment of *r*-value in the fuzzy temporal association rules will uncover more undiscovered correlations.

- (2) There may be more undiscovered correlations mined with adjusting of h , which does not restrict itself to the following temporal factor.
- (3) Models for mining association rules do not necessarily restrict themselves within the same object A . Take stock as an example. A great rise in the price of stock A_1 may lead to a small drop in the price of stock A_2 and stock A_3 .
- (4) Based on actual experience and transaction need, the number of s_{ij} can take different values. In Lanzi's case, five were taken. However, the rise and drop of prices can be further divided into more distinct groups, and then a richer set of correlation rule events may be gained.
- (5) Currently only event attributes are fuzzified, but time is also a key attribute dimension. Fuzzifying temporal states is also extremely useful, in, e.g., mining and discovering fuzzy periodic rules.

In addition, all other parameters in the model need further exploration and should be reasonably set based on actual trading data. A more effective and specific method of parameter moderation still needs to be found, as reasonable parameter setting helps improve the speed and accuracy of the algorithm, as well as the effectiveness and practicality of association rules.

Author Contributions: Conceptualization, A.Z. and Z.M.; methodology, software and validation, A.Z.; writing—original draft preparation, A.Z. and Z.M.; writing—review and editing, R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (NO.11871434).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: [<http://vip.stock.finance.sina.com.cn/q/go.php/vInvestConsult/kind/lhb/index.phtml> and <https://tushare.pro/document/2?docid=24>].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Laxman, S.; Sastry, P.S. A survey of temporal data mining. *Sadhana* **2006**, *31*, 173–198. [[CrossRef](#)]
2. Lin, W.; Orgun Mehmet, A.; Williams Graham, J. An Overview of Temporal Data Mining. *AusDM* **2022**, *8*, 83–90.
3. Lim, B.; Zohren, S. Time Series Forecasting with Deep Learning: A Survey. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2021**, *379*, 20200209. [[CrossRef](#)] [[PubMed](#)]
4. Zadeh, L.A. Fuzzy Sets and Information Granularity. *Adv. Fuzzy Set Theory Appl.* **1979**, *11*, 3–18.
5. Li, X.; Zhou, J.; Pedrycz, W. Linking granular computing, big data and decision making: A case study in urban path planning. *Soft Comput.* **2020**, *24*, 7435–7450. [[CrossRef](#)]
6. Wang, G.; Hong, Y. Multi-Granularity Cognitive Computing—A New Model for Big Data Intelligent Computing. *Front. Data Computing* **2020**, *1*, 75–85. [[CrossRef](#)]
7. Yu, F.; Cai, R. Optimized fuzzy information granulation of temporal data. In Proceedings of the 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, Shandong, China, 10–12 August 2010; Volume 1, pp. 419–423. [[CrossRef](#)]
8. Mitsa, T. Temporal data mining. In *Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2010.
9. Lovrić, M.; Milanović, M.; Stamenković, M. Algorithmic methods for segmentation of time series: An overview. *J. Contemp. Econ. Bus. Issues* **2014**, *1*, 31–53.
10. Rawassizadeh, R.; Momeni, E.; Dobbins, C.; Gharibshah, J.; Pazzani, M. Scalable Daily Human Behavioral Pattern Mining from Multivariate Temporal Data. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 3098–3112. [[CrossRef](#)]
11. Chen, C.-H.; Hong, T.-P.; Tseng, V.S. Fuzzy data mining for time-series data. *Appl. Soft Comput.* **2012**, *12*, 536–542. [[CrossRef](#)]
12. Chen, C.-H.; Hong, T.-P.; Lin, S.-B. Mining fuzzy temporal knowledge from quantitative transactions. In Proceedings of the 2011 International Conference on System Science and Engineering, Macau, China, 8–10 June 2011; pp. 405–409. [[CrossRef](#)]
13. Chen, C.-H.; Lan, G.-C.; Hong, T.-P.; Lin, S.-B. Mining fuzzy temporal association rules by item lifespans. *Appl. Soft Comput.* **2016**, *41*, 265–274. [[CrossRef](#)]
14. Hong, T.-P.; Lan, G.-C.; Su, J.-H.; Wu, P.-S.; Wang, S.-L. Discovery of temporal association rules with hierarchical granular framework. *Appl. Comput. Inform.* **2016**, *12*, 134–141. [[CrossRef](#)]

15. Meng, Z.; Jiang, H.; Jiang, M.; Liu, Y. An Algorithm of Discovering Approximate Periodicity Based on Self-Organizing Map for Temporal Data. In Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), Hainan, China, 24–27 August 2007; Volume 3, pp. 293–297. [\[CrossRef\]](#)
16. Mahmood, N.; Burney, A. Advances in Fuzzy Temporal Relational Databases: A Review. *WSEAS Trans. Inf. Sci. Appl.* **2011**, *8*, 171–184.
17. Wu, H.; Long, H.; Wang, Y.; Wang, Y. Stock index forecasting: A new fuzzy time series forecasting method. *J. Forecast.* **2021**, *40*, 653–666. [\[CrossRef\]](#)
18. Chen, W.; An, J.; Li, R.; Fu, L.; Xie, G.; Alam, B.M.Z.; Li, K. A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial-temporal data features. *Future Gener. Comput. Syst.* **2018**, *89*, 78–88. [\[CrossRef\]](#)
19. Xiao, F.; Ding, W. Divergence measure of Pythagorean fuzzy sets and its application in medical diagnosis. *Appl. Soft Comput.* **2019**, *79*, 254–267. [\[CrossRef\]](#)
20. Mazarbhuiya, F.A. An Efficient Algorithm for Mining Fuzzy Temporal Data. *Int. J. Math. Trends Technol.* **2016**, *37*, 1–5. [\[CrossRef\]](#)
21. Mazarbhuiya, F.A.; Shenify, M. Finding Local and Periodic Association Rules from Fuzzy Temporal Data. In Proceedings of the 2014 International Conference on Advances in Big Data Analytics, Beijing, China, 4–7 August 2014; Volume 9, pp. 21–24.
22. Lee, W.-J.; Jiang, J.-Y.; Lee, S.-J. Mining fuzzy periodic association rules. *Data Knowl. Eng.* **2008**, *65*, 442–462. [\[CrossRef\]](#)
23. Purificación, C. Fuzzy temporal association rules: Combining temporal and quantitative data to increase rule expressiveness: Fuzzy temporal association rules. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2014**, *4*, 64–70. [\[CrossRef\]](#)
24. Owuor, D.; Laurent, A.; Orero, J. Mining Fuzzy-Temporal Gradual Patterns. In Proceedings of the 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, 23–26 June 2019; IEEE: New Orleans, LA, USA, 2019; pp. 1–6. [\[CrossRef\]](#)
25. Alom, M.F. Mining Local Patterns from Fuzzy Temporal Data. *Int. J. Eng. Appl. Sci.* **2016**, *2*, 257768.
26. Dogan, O.; Kem, F.C.; Oztaysi, B. Fuzzy association rule mining approach to identify e-commerce product association considering sales amount. *Complex Intell. Syst.* **2022**, *8*, 1551–1560. [\[CrossRef\]](#)
27. Keyvanpour, M.R.; Mehrmolaei, S.; Etaati, A. PLI-X: Temporal Association Rules Mining in Customer Relationship Management Systems. *Comput. Knowl. Eng.* **2020**, *2*, 29–48. [\[CrossRef\]](#)
28. Sharmila, S.; Vijayarani, S. Association rule mining using fuzzy logic and whale optimization algorithm. *Soft Comput.* **2021**, *25*, 1431–1446. [\[CrossRef\]](#)
29. Chen, C.-H.; Chou, H.; Hong, T.-P.; Nojima, Y. Cluster-Based Membership Function Acquisition Approaches for Mining Fuzzy Temporal Association Rules. *IEEE Access* **2020**, *8*, 123996–124006. [\[CrossRef\]](#)
30. Poli, V.S.R. Fuzzy Temporal Data Mining Algorithms. *EasyChair*, No. 3560. 2020.
31. de Campos, S.P.V. Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature. *Appl. Soft Comput.* **2020**, *92*, 106275. [\[CrossRef\]](#)
32. El Meftahy, M.; El Kabbouri, M. Application of neuro-fuzzy methods for stock market forecasting: A systematic review. *Int. J. Account. Finance Audit. Manag. Econ.* **2022**, *3*, 437–454.
33. Li, H.; Wei, M. Fuzzy clustering based on feature weights for multivariate time series. *Knowl.-Based Syst.* **2020**, *197*, 105907. [\[CrossRef\]](#)
34. Ouyang, W.-M.; Cai, Q.-S. Discovering association rules with temporal constraints in databases. *J. Softw.* **1999**, *10*, 527–532.
35. Rahman, T.; Kabir, M.M.J.; Kabir, M. Performance Evaluation of Fuzzy Association Rule Mining Algorithms. In Proceedings of the 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 20–22 December 2019; IEEE: Khulna, Bangladesh, 2019; pp. 1–4. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.