



HAL
open science

A Relative Fuzzy Hamming-based Similarity for Temporal Semantic Sequences

Hiba Merakchi, Cyril de Runz, Thomas Devogele, Veronika Peralta

► **To cite this version:**

Hiba Merakchi, Cyril de Runz, Thomas Devogele, Veronika Peralta. A Relative Fuzzy Hamming-based Similarity for Temporal Semantic Sequences. 2025 IEEE International Conference on Fuzzy Systems (FUZZ), Jul 2025, Reims, France. pp.1-6, <10.1109/FUZZ62266.2025.11152160>. <hal-05265118>

HAL Id: hal-05265118

<https://hal.science/hal-05265118v1>

Submitted on 6 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A Relative Fuzzy Hamming-based Similarity for Temporal Semantic Sequences

Hiba Merakchi
LIFAT, University of Tours
Blois, France
hiba.merakchi@univ-tours.fr

Cyril de Runz
LIFAT, University of Tours
Blois, France
0000-0002-5951-6859

Thomas Devogele
LIFAT, University of Tours
Blois, France
thomas.devogele@univ-tours.fr

Veronika Peralta
LIFAT, University of Tours
Blois, France
0000-0002-9236-9088

Abstract—Temporal semantic sequences -temporally ordered elements with meaningful labels- are widely used in fields such as mobility, medicine, and e-learning. However, analyzing and comparing these sequences is challenging due to length variations, semantic richness, and temporal dynamics. Existing similarity measures are limited to sequences of identical lengths or fail to capture both semantic and temporal aspects simultaneously. We propose the Relative Fuzzy Temporal Hamming (*RFTH*) similarity measure, which leverages fuzzy logic to adapt to sequences of varying lengths. It not only captures semantic and temporal similarity but also guarantees robust performance with interesting properties. A running example illustrates its applicability and advantages. This measure promises interesting usages for analyzing human mobility, patient records, weather data, and many other applications that require robust sequence comparison.

Index Terms—Similarity measures; Uncertainty, “fuzzy,” and probabilistic; Time management; Data models.

I. INTRODUCTION

Semantic sequences are temporally ordered representations of elements, enriched with meaningful labels that highlight their contextual importance. They capture how these elements are arranged over time, making them a valuable tool in many fields such as mobility analysis [1], medicine [2], life trajectories [3], and more. This wide range of uses shows their adaptability and potential for understanding temporal dynamics across different applications.

Despite their widespread use, comparing semantic sequences remains a challenging but vital task. Motivation comes from the need to discern patterns, identify similarities, and find trends within these sequences. For example, in mobility analysis, comparing semantic sequences enables the detection of shared behavioral patterns among users, which is crucial for urban planning or personalized recommendations. In health-care, comparing patient treatment pathways can help identify best practices or deviations from standard protocols.

Many comparison methods (through distances and similarity measures) were proposed for specific cases (e.g. sequences of fixed length), but often fall short of capturing both temporal structure and semantic richness. Many challenges arise. First, semantic temporal sequences often vary in time length (t-length for short, which represents the sum of the durations of their elements), due to incomplete data or intrinsic differences, such as the complexity and length of medical

protocols. Traditional comparison methods, like the Hamming distance [4], which typically require sequences of equal length, struggle with these variations, leading to poor alignment or mismatches.

Semantic richness of sequences is also poorly considered. Many methods look for exact element matching, ignoring semantic similarity. For example, transport modes like “Bike” and “Walk” are not identical but share a common meaning as they both represent forms of *slow mobility*. These modes are semantically closer to each other than to “Bus,” which represents a distinct category of mobility.

Furthermore, temporal structure, such as different element durations or intervals, add complexity to the comparison. Indeed, while two sequences may share semantically similar elements, differences in timestamps or durations often obscure this equivalence. For example, the five sequences in Fig. 2 describe similar types of activities (work in blue, shopping in red, and transportation in yellowish tones), but they differ in temporal structure, i.e., in the timing (timestamps and durations) and the order of elements. Such a temporal structure is crucial for capturing the full context behind sequences, especially in domains like human mobility or activity analysis.

On top of this, most traditional methods emphasize either semantic or temporal structure, failing to integrate both effectively.

This paper aims to introduce a new similarity measure that addresses the challenge of length variations, while capturing both semantic richness and temporal structure, those adapting to diverse applications.

The paper is organized as follows: Section 2 outlines the state of the art. Section 3 introduces our similarity measure (*RFTH*), explains its computation steps, and illustrates its usefulness through a running example. Section 4 discusses the properties of *RFTH* and proves that it is a similarity. Finally, Section 5 concludes and suggests future research directions.

II. STATE OF THE ART

Existing comparison measures for sequences are mostly based on Edit Distance [5] or Hamming Distance [4]. The former overlooks temporal structure, while the latter provides a more suitable framework. Hamming-based methods are particularly appealing because they consider the temporal structure of sequence elements, such as elements’ durations. In this

section, we review some of the existing methods for semantic temporal sequences, with a focus on Hamming-based distances and their variations.

Hamming Distance (*HD*): The Hamming Distance is a classical method that compares the elements appearing at the same position and counts the differences [4]. While *HD* is computationally efficient, its application is restricted to scenarios where sequences are of equal length. Note that, when applied to temporal semantic sequences, *HD* requires discretizing time by repeating symbols according to their duration. This transforms continuous time into discrete steps, enabling direct element-by-element comparison.

Dynamic Hamming Distance (*DHD*): *DHD* [6] incorporates the temporal aspect of sequences by associating specific elements with particular time intervals (e.g. [12:00–13:00] for lunch hours). Unlike *HD*, *DHD* measures distance based on the frequency of transitions, which refer to the changes between element types at consecutive time points in a sequence. For instance, a transition occurs when an individual switches from “Working” to “Break” at one time point. The cost of substitution measures how costly it is to replace one symbol with another at a given time, and is estimated using conditional probabilities between adjacent symbols. This makes the measure both time- and context-dependent, enabling a more nuanced comparison of sequences. In [7], the authors proposed a simplified version of this cost computation, which is now implemented in the TraMineR library. While the probabilistic approach of *DHD* is effective for identifying anomalies, its reliance on memoryless models makes it sensitive to temporal misalignments.

Fuzzy Temporal Hamming Distance (*FTH*): *FTH* [8], generalizes *HD* through a fuzzy approach. It leverages a fuzzy context function to quantify the temporal neighborhood of elements and uses Wu-Palmer similarity [9] to compare elements based on semantic closeness derived from a taxonomy. *FTH* assigns costs to edit operations (substitute) based on both temporal alignment and semantic similarity. The one-sided *FTH* calculates the cost of transforming one sequence into another, with the full *FTH* achieved by symmetrizing this measure. This approach tolerates small temporal shifts and permutations, integrates temporal and semantic proximities, and ensures computational efficiency through optimized algorithms. However, its applicability is limited when comparing sequences of varying t-lengths.

A. Limitations of current approaches

The previous methods demonstrate strong potential for comparing temporal semantic sequences but exhibit some limitations mainly in scenarios involving sequences with different t-lengths.

Table I compares state-of-the-art methods based on the criteria described below.

- **Permutation (Perm):** Whether the method is robust to the permutation of elements within the sequence. Two sequences with the same elements, except for the order,

Method	Perm	Rep	Dur	T-len	Sem. Sim
HD	-	-	-	-	-
DHD	-	-	✓	-	-
FTH	✓	✓	✓	-	✓
RFTH	✓	✓	✓	✓	✓

TABLE I

COMPARISON OF METHODS BASED ON DIFFERENT CRITERIA.

should have high similarity, especially if the permuted elements are temporally close.

- **Repetition (Rep):** Whether the method is robust to the repetition of elements in the sequence. Sequences with identical elements but with repetitions should have high similarity, especially if the repeated elements are temporally close.
- **Duration (Dur):** Whether the method incorporates the durations of elements in the sequence.
- **T-lengths (T-len):** Whether the method allows for sequences of different t-lengths to be compared.
- **Semantic similarity (Sem. Sim):** Whether the method takes into account the conceptual proximity between elements, reflecting how similar or related they are in terms of their meaning or role in a given context. Sequences with semantically close elements should have high similarity, especially if they are temporally close.

These challenges motivate the development of a new similarity measure, which addresses these limitations by incorporating mechanisms to handle sequences of varied t-lengths while preserving their semantic and temporal structures.

III. RELATIVE FUZZY TEMPORAL HAMMING (*RFTH*)

RFTH is an extension of *FTH* that allows the comparison of sequences of different t-lengths. *RFTH* combines two key criteria using a T-norm: t-length comparability and semantic + temporal similarity. Indeed, two sequences are considered similar, if they have comparable t-length and contain semantically similar elements positioned closely within the sequence.

To achieve this, t-length comparability is ensured through a fuzzy function, while semantic similarity and temporal structure are handled by normalizing the sequences’ t-lengths to a common scale, enhancing the original *FTH* method.

Fig. 1 illustrates the pipeline of *RFTH*, which steps are described in next subsections, after formally defining temporal semantic sequences.

Definition. Let \mathbb{E} be the set of possible semantic elements, and \mathbb{D} the set of possible durations. We define a temporal semantic sequence $S = \langle p_1, p_2, \dots, p_n \rangle$ as a collection of pairs $p_i = (e_i, d_i)$, where $e_i \in \mathbb{E}$ represents the semantic element occurring at position i , and $d_i \in \mathbb{D}$ denotes the duration of this element in units of time.

In the following, let \mathbb{S} be the set of all possible sequences.

Example 1: Consider as a running example, some sequences of human activities throughout the day. A main sequence S is compared against four sequences S_1, S_2, S_3 , and S_4 . They

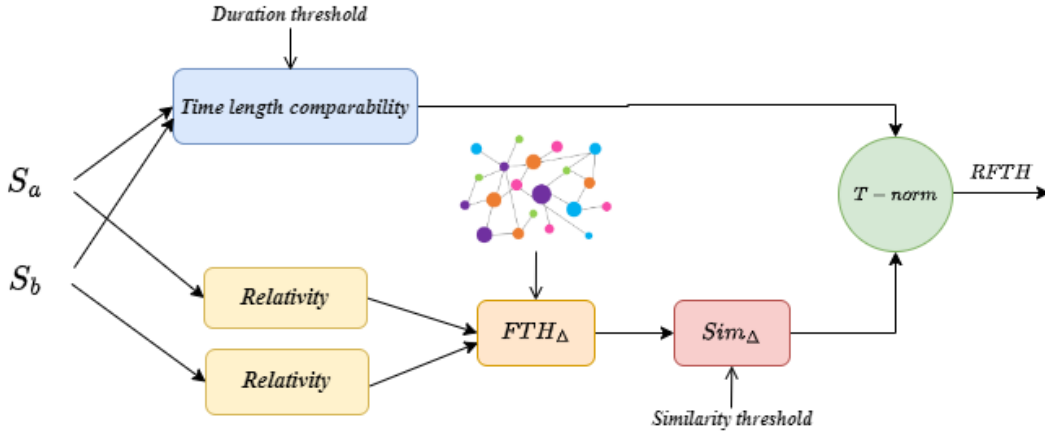


Fig. 1. Pipeline for Calculating Relative FTH Similarity.

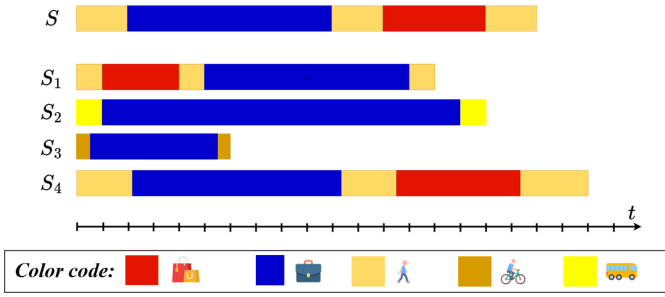


Fig. 2. Examples of temporal semantic sequences: activity types are represented by colors, and their durations are indicated by time slots, with each slot representing half an hour.

are sketched in Fig. 2 and an example of representation (for S_2) is provided below:

$$S_2 = \langle (\mathbf{Bus}, 30\text{min}), (\mathbf{Work}, 7\text{h}), (\mathbf{Bus}, 30\text{min}) \rangle$$

The activities in these sequences are associated with the taxonomy shown in Fig. 4.

A. Time-length Comparability

The goal of this step is to calculate a fuzzy index that reflects the comparability of two sequences in terms of their t-length. This index helps ensure that the comparison is meaningful and consistent. Indeed, when sequences have very different t-lengths, even if stretched to the same relative scale, their temporal structures may artificially appear more similar than they actually are. This introduces skew into the similarity evaluation, potentially leading to misleading results.

Let S_a and S_b be two temporal semantic sequences. Let $l_t(S_a)$ be the t-length of S_a and $l_t(S_b)$ be the t-length of S_b . Let l be the acceptable t-length difference beyond which we consider the comparability of two sequences meaningless. The t-length comparability index of S_a and S_b is computed using a fuzzy function that assigns a similarity value based on the difference in their t-lengths. Specifically, we utilize a

symmetric triangular function for its simplicity, interpretability, and ability to represent a balanced transition between acceptable and unacceptable sequence t-lengths. This choice is motivated by the requirement that the function be continuous, symmetric, and defined as a fuzzy number with a punctual core [10]. Furthermore, a triangular function, unlike a trapezoidal one for example, ensures the separability property, which guarantees that the measure returns a value of 1 only when the sequences S_a and S_b are identical, a fundamental requirement for effective similarity assessment. It is defined as follows:

$$I_c(S_a, S_b) = \begin{cases} 0, & \text{if } |l_t(S_a) - l_t(S_b)| > l, \\ 1 - \frac{|l_t(S_a) - l_t(S_b)|}{l}, & \text{otherwise.} \end{cases} \quad (1)$$

This index quantifies how close the two sequences are in terms of t-length. Thus, it returns 0 if a sequence is too short or too long compared to the other.

The choice of parameter l , which represents half of the fuzzy function support, is intrinsically tied to the specific domain of application and is, therefore, fixed by domain experts who possess a nuanced understanding of the data and the acceptable limits of the temporal variability. Alternatively, it can be set as a ratio (e.g., 40%) of the median sequence t-length in a domain (for instance, for career trajectories spanning around 40 years, l could be set to 16 years); the standard deviation of t-length in a dataset can also be used.

Example 2: For our example, we set the threshold l to 4 hours. The t-length comparability index scores, are shown in Table II. We remark that sequence S_3 being too short obtains a zero score. \square

B. Relativity: Duration Normalization

Duration normalization involves adjusting the element durations of each sequence to relative values, expressed as proportions of the total t-length of the sequence.

Definition. Let $S = \langle (e_1, d_1), (e_2, d_2), \dots, (e_n, d_n) \rangle$ be a temporal semantic sequence, with $e_i \in \mathbb{E}$ and $d_i \in \mathbb{D}$, for $i \in$



Fig. 3. Representation of \hat{S} , \hat{S}_1 , \hat{S}_2 , \hat{S}_3 , and \hat{S}_4 , the normalized versions of sequences S , S_1 , S_2 , S_3 , and S_4 . Each time slot represents 0.1

Sequence	S_1 / \hat{S}_1	S_2 / \hat{S}_2	S_3 / \hat{S}_3	S_4 / \hat{S}_4
I_c	0.5	0.75	0	0.75
FTH_{Δ}	0.42	0.51	0.44	0.04
sim_{Δ}	0.58	0.49	0.56	0.96
Zadeh T-norm $\min(I_c, sim_{\Delta})$	0.5	0.49	0	0.75
Probabilistic T-norm $I_c \times sim_{\Delta}$	0.29	0.36	0	0.72
Łukasiewicz T-norm $\max(0, I_c + sim_{\Delta} - 1)$	0.08	0.24	0	0.71

TABLE II

T-LENGTH COMPARABILITY INDEX, FTH_{Δ} DISTANCE, sim_{Δ} SIMILARITY AND 3 T-NORMS, BETWEEN SEQUENCE S AND EACH OF THE SEQUENCES S_1 , S_2 , S_3 AND S_4 .

$\{1 \dots n\}$. We define a normalized temporal semantic sequence $\hat{S} = \langle \hat{p}_1, \hat{p}_2, \dots, \hat{p}_n \rangle$ as a collection of pairs $\hat{p}_i = (e_i, \hat{d}_i)$, where \hat{d}_i represents the normalized duration of element e_i , given by:

$$\hat{d}_i = \frac{d_i}{\sum_{j=1}^n d_j} \quad (2)$$

Thus, sequences, regardless of their absolute t-lengths, are projected onto a common scale where the sum of normalized t-lengths equals 1.

Example 3: In this step, sequences S , S_1 , S_2 , S_3 and S_4 , with t-lengths of 9h, 7h, 8h, 5h and 11h respectively, are transformed into their normalized versions \hat{S} , \hat{S}_1 , \hat{S}_2 , \hat{S}_3 and \hat{S}_4 , respectively. They are shown in Fig.3.

For instance, \hat{S}_2 can be represented as follows:

$$\hat{S}_2 = \langle (\mathbf{Bus}, \frac{1}{16}), (\mathbf{Work}, \frac{14}{16}), (\mathbf{Bus}, \frac{1}{16}) \rangle$$

C. Application of FTH

Once the normalization is complete, the t-lengths of the normalized versions of the remaining sequences are all equal to 1, which means that FTH can now be applied onto them. Two variations of FTH are defined in [8]: FTH_{γ} and FTH_{Δ} . We opt for the use of FTH_{Δ} in our case. Its time-weighted cost function ensures that longer elements incur a higher editing cost, preserving their significance in the sequence, while shorter elements contribute proportionally less, preventing their overshadowing by longer ones. For

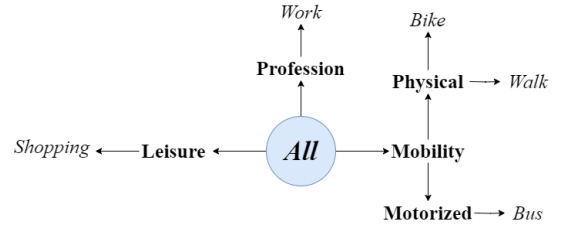


Fig. 4. Taxonomy of activities (only those used in the example are shown).

	Work	Shopping	Swim	Read	Bike	Walk	Bus
Work	1						
Shopping		1	$\frac{1}{2}$	$\frac{1}{2}$			
Swim		$\frac{1}{2}$	1	$\frac{1}{2}$			
Read		$\frac{1}{2}$	$\frac{1}{2}$	1			
Bike					1	$\frac{2}{3}$	$\frac{1}{3}$
Walk					$\frac{2}{3}$	1	$\frac{1}{3}$
Bus					$\frac{1}{3}$	$\frac{1}{3}$	1

Fig. 5. Wu-Palmer similarity between activities. Empty cells denote a similarity equal to 0.

\hat{S}_a and \hat{S}_b , normalized versions of S_a and S_b respectively, $FTH_{\Delta}(\hat{S}_a, \hat{S}_b)$ is given as:

$$FTH_{\Delta}(\hat{S}_a, \hat{S}_b) = \max\{FTH_{\Delta}(\hat{S}_a \rightarrow \hat{S}_b), FTH_{\Delta}(\hat{S}_b \rightarrow \hat{S}_a)\}$$

One-sided fuzzy temporal distance $FTH_{\Delta}(\hat{S}_a \rightarrow \hat{S}_b)$ is the sum of weighted costs required to transform \hat{S}_a into \hat{S}_b , considering both temporal and semantic similarity of elements. FTH_{Δ} is proven to be symmetric, and its robustness to time shifts is further enhanced by the introduction of a fuzzy time-window [11].

Example 4: To calculate FTH_{Δ} between the sequences \hat{S} , \hat{S}_1 , \hat{S}_2 , \hat{S}_3 and \hat{S}_4 in our example, we use the taxonomy illustrated in Fig. 4. We compute the Wu-Palmer semantic similarity between the activities, which measures the closeness of two activities a_1 and a_2 based on their depth in the taxonomy and the depth of their least common ancestor (LCS) [9].

$$Sim_{\text{Wu-Palmer}}(a_1, a_2) = \frac{2 \cdot \text{depth}(\text{LCS})}{\text{depth}(a_1) + \text{depth}(a_2)}$$

Results are shown in Fig. 5. We set the fuzzy time-window to 0.5 since durations are normalized. Results are summarized in Table II. □

D. From FTH_{Δ} to Sim_{Δ}

By applying FTH_{Δ} , we obtain a distance measure that we transform into a similarity denoted as sim_{Δ} . FTH_{Δ} is proven to be bounded by the sequence t-length [8] (in this case 1). Hence, the similarity function is defined as follows:

$$sim_{\Delta}(\hat{S}_a, \hat{S}_b) = 1 - FTH_{\Delta}(\hat{S}_a, \hat{S}_b) \quad (3)$$

Example 5: The results of this step for our running example are shown in Table II. \square

E. Similarity Aggregation

This step involves aggregating the previous measures (sim_Δ and I_c) into a unified metric using a T-norm, which ensures a mathematically sound, flexible, and proportional fusion of both measures. The rationale behind this approach lies in the properties of T-norms, which are widely used in fuzzy logic and multi-criteria decision-making for modeling conjunctions (i.e., “and” operations) where all criteria must hold true to a certain degree.

- sim_Δ emphasizes the semantic and temporal alignment of sequences, preserving the order and meaning of elements.
- I_c ensures comparability in terms of the t-length of the sequences.

Hence, our similarity measure $RFTH$ reflects the simultaneous satisfaction of both measures.

$$RFTH(S_a, S_b) = T(I_c(S_a, S_b), \text{sim}_\Delta(\hat{S}_a, \hat{S}_b)) \quad (4)$$

The choice of a T-norm affects the aggregation outcome significantly. Several T-norms exist in the literature, from which we consider the usual Zadeh, Probabilistic and Łukasiewicz ones [12]. The Zadeh T-norm is considered the most “optimistic” because it devaluates the less the combination values, ensuring that the result reflects only the lowest degree of overlap between the two. The Probabilistic and Łukasiewicz T-norms are viewed as more “pessimistic,” as they tend to decrease the combined values more rapidly, meaning that both components significantly impact the final outcome. Among these, the Probabilistic T-norm strikes a balance: it is more optimistic than the Łukasiewicz T-norm but less so than the Zadeh T-norm. Therefore, we recommend using the Probabilistic T-norm for practical applications.

Example 6: The results of using different aggregation T-norms in this step are summarized in Table II. The code and the tests for the examples presented in this work are available anonymously on GitHub¹, along with a user guide. \square

IV. RFTH PROPERTIES

In this section we show that $RFTH$ satisfies the criteria outlined in Table I as well as similarity properties.

A. Satisfaction of expected criteria

Hereafter, we discuss the satisfaction of the expected criteria introduced in Section II and highlight the advantages of our approach:

- 1) **Duration:** By construction, the proposed method accounts for the durations of the elements in the sequences. Firstly, duration is a component of our representation of sequences. Duration is then considered for T-length

comparability and relativity, and is a core concept in FTH computation. Throughout the running example, we illustrated how durations affected each step of method.

- 2) **T-length:** The proposed method is able to robustly compare sequences of varied t-lengths, as illustrated through the comparison of sequence S with S_1 , S_2 , S_3 , and S_4 . This is achieved by combining the t-length comparability index and the sim_Δ measure into a single product.
- 3) **Permutation:** $RFTH$ is robust to permutation because it extends FTH, which was proved robust to permutation [8]. This was illustrated in the running example, where S_1 , permuting work and shopping activities, obtains a suitable similarity score.
- 4) **Repetition:** Similarly, $RFTH$ is robust to repetition because it extends FTH, which is robust to repetition [8].
- 5) **Semantic similarity** Also because of FTH, $RFTH$ is robust to semantic similarity. Indeed, it considers a taxonomy of element types (as the one illustrated in Figure 4) and uses Wu-Palmer similarity to compare elements. In the previous example, $\text{sim}_\Delta(\hat{S}, \hat{S}_3)$ is superior to $\text{sim}_\Delta(\hat{S}, \hat{S}_2)$ because the activity “Bike” is semantically closer to the activity “Walk” than the activity “Bus”.

Our proposed method, $RFTH$, can be optimized by filtering out sequences that are too short or too long, as these sequences have a null t-length comparability index and give a $RFTH$ similarity score of 0 when combined with any other sim_Δ value using a T-norm. This is because 0 is the absorbing element of all T-norms, i.e.,

$$T(a, 0) = 0 \quad \text{for any } a \in [0, 1].$$

Thus, if the t-length comparability index of a sequence is equal to 0, it is excluded from the subsequent steps of computation to optimize the process.

Example 7: Sequences that are too short, such as S_3 in the previous example, can be excluded from the comparison because their t-length comparability index was null. This exclusion occurs despite S_3 having similar activities and order as S . \square

By removing such sequences, our method ensures optimized and meaningful comparisons by retaining only sequences with comparable temporal lengths.

B. Proof that $RFTH$ is a Similarity Measure

This section aims to verify that the $RFTH$ measure, defined using a general T-norm T , satisfies the essential properties of a similarity measure. Specifically, we will demonstrate that it is symmetric and separable.

Symmetry: The symmetry property requires that:

$$\forall S_a, S_b \in \mathbb{S}, RFTH(S_a, S_b) = RFTH(S_b, S_a).$$

According to Equation 4:

$$RFTH(S_a, S_b) = T(I_c(S_a, S_b), \text{sim}_\Delta(\hat{S}_a, \hat{S}_b))$$

¹Repository link: https://github.com/IamSherlockeddd/FUZZ2025/blob/main/Final_FUZZ_2025.ipynb

Analyzing the components:

1) **Symmetry of I_c :** According to Equation 1, there are two cases:

- a) **Case $|l_t(S_a) - l_t(S_b)| > l$**
- $I_c(S_a, S_b) = 0$
 - Similarly, $I_c(S_b, S_a) = 0$
 - Thus, $I_c(S_a, S_b) = I_c(S_b, S_a)$ in this case.
- b) **Case $|l_t(S_a) - l_t(S_b)| \leq l$**
- $I_c(S_a, S_b) = 1 - \frac{|l_t(S_a) - l_t(S_b)|}{l}$.
 - Similarly, $I_c(S_b, S_a) = 1 - \frac{|l_t(S_b) - l_t(S_a)|}{l}$.
 - Since $|l_t(S_a) - l_t(S_b)| = |l_t(S_b) - l_t(S_a)|$, we have $I_c(S_a, S_b) = I_c(S_b, S_a)$.

2) **Symmetry of sim_Δ :** Since FTH_Δ is proven to be symmetric [8], and according to Equation 3, we have:

$$\begin{aligned} \text{sim}_\Delta(\hat{S}_a, \hat{S}_b) &= 1 - FTH_\Delta(\hat{S}_a, \hat{S}_b) \\ &= 1 - FTH_\Delta(\hat{S}_b, \hat{S}_a) \\ &= \text{sim}_\Delta(\hat{S}_b, \hat{S}_a). \end{aligned}$$

Since $T(a, b) = T(b, a)$ for every T-norm [13], combining these results we obtain:

$$RFTH(S_a, S_b) = RFTH(S_b, S_a).$$

Thus, $RFTH$ is symmetric. \square

2. Separability: The separability property requires that:

$$\forall S_a, S_b \in \mathbb{S}, RFTH(S_a, S_b) = 1 \iff S_a = S_b.$$

When $S_a = S_b$:

- $I_c(S_a, S_b) = 1$ because $|l_t(S_a) - l_t(S_b)| = 0$.
- $\text{sim}_\Delta(\hat{S}_a, \hat{S}_b) = 1$ because $FTH(S_a, S_b) = 0$.

Substituting both terms in Equation 4, we have:

$$RFTH(S_a, S_b) = T(1, 1) = 1.$$

When $RFTH(S_a, S_b) = 1$: As a T-norm can only achieve 1 when both terms are 1-valued, and following Equation 4, we have:

$$\begin{aligned} \text{If } RFTH(S_a, S_b) = 1 &\implies T(I_c(S_a, S_b), \text{sim}_\Delta(\hat{S}_a, \hat{S}_b)) = 1 \\ &\implies I_c(S_a, S_b) = 1 \wedge \text{sim}_\Delta(\hat{S}_a, \hat{S}_b) = 1 \end{aligned}$$

Developing each term, we have:

- $I_c(S_a, S_b) = 1 \implies |l_t(S_a) - l_t(S_b)| = 0 \implies l_t(S_a) = l_t(S_b)$.
- $\text{sim}_\Delta(\hat{S}_a, \hat{S}_b) = 1 \implies FTH_\Delta(S_a, S_b) = 0$
And as FTH is separable [8]:
 $FTH_\Delta(S_a, S_b) = 0 \implies \hat{S}_a = \hat{S}_b$.

Combining both results we have $S_a = S_b$.

Thus, $RFTH$ is separable. \square

It is also important to note that $RFTH$ is not transitive. For example, it is possible to observe that both $RFTH(S, S_1) > 0$ and $RFTH(S_1, S_3) > 0$, while $RFTH(S, S_3) = 0$. This situation occurs due to the t -length compatibility index.

V. CONCLUSION

In this article, we introduced $RFTH$, a new Hamming-based similarity measure, tailored for the comparison of temporal semantic sequences with different t -lengths. By combining relative t -length normalization and t -length comparability, this method enables the comparison of diverse sequences while preserving the integrity of their temporal and semantic relationships.

We conducted several experiments on data from diverse fields, including human mobility, weather data, and log files, to validate the applicability of our proposed method. A running example based on mobility activities showcased it. The prospects for this method are promising.

The current approach assumes that sequences are continuous and do not contain gaps. Future work aims to extend this method to accommodate such discontinuities, enabling the comparison of sequences with missing data or irregular event patterns. This enhancement would make the measure more robust and applicable to a broader range of real-world scenarios. It also paves the way for the development of a query language capable of formulating complex queries on temporal semantic sequences. Such a language could provide the necessary flexibility to extract similar subsequences while considering temporal and semantic contexts, which would be particularly useful in fields such as mobility trajectory analysis or life-course tracking.

REFERENCES

- [1] P. R. Stopher and S. P. Greaves, "Household travel surveys: Where are we going?" *Transportation Research Part A: Policy and Practice*, vol. 41, no. 5, pp. 367–381, 2007.
- [2] S. A. Murray, M. Kendall, K. Boyd, and A. Sheikh, "Illness trajectories and palliative care," *British Medical Journal*, vol. 330, no. 7498, pp. 1007–1011, 2005.
- [3] A. Abbott, *Time matters: On theory and method*. University of Chicago Press, 2001.
- [4] R. W. Hamming, "Error detecting and error correcting codes," *The Bell System Technical Journal*, vol. 29, no. 2, pp. 147–160, 1950.
- [5] V. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Proceedings of the Soviet physics doklady*, 1966.
- [6] L. Lesnard, "Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns," *Sociological Methods & Research*, 2010.
- [7] A. Gabadinho, G. Ritschard, N. S. Müller, and M. Studer, "Analyzing and visualizing state sequences in r with traminer," *Journal of statistical software*, vol. 40, pp. 1–37, 2011.
- [8] C. Moreau, T. Devogele, C. De Runz, V. Peralta, E. Moreau, and L. Etienne, "A fuzzy generalisation of the hamming distance for temporal sequences," in *FUZZ-IEEE*, 2021.
- [9] Z. Wu and M. Palmer, "Verb semantics and lexical selection," in *Proceedings of the Association for Computational Linguistics*, 1994, pp. 133–138.
- [10] J. Dijkman, H. van Haeringen, and S. de Lange, "Fuzzy numbers," *Journal of Mathematical Analysis and Applications*, vol. 92, no. 2, pp. 301–341, 1983.
- [11] S. Jiang, J. F. Jr., and M. C. González, "Clustering daily patterns of human activities in the city," *Data Min. Knowl. Discov.*, vol. 25, no. 3, pp. 478–510, 2012.
- [12] D. Boixader and J. Recasens, "Vague and fuzzy t -norms and t -conorms," *Fuzzy Sets Syst.*, vol. 433, pp. 156–175, 2022.
- [13] E. P. Klement, R. Mesiar, and E. Pap, *Triangular norms*. Springer Science & Business Media, 2013, vol. 8.