

# Discovering Leitmotifs in Multidimensional Time Series

Patrick Schäfer  
Humboldt-Universität zu Berlin  
Berlin, Germany  
patrick.schaefer@hu-berlin.de

Ulf Leser  
Humboldt-Universität zu Berlin  
Berlin, Germany  
leser@informatik.hu-berlin.de

## ABSTRACT

A leitmotif is a recurring theme in literature, movies or music that carries symbolic significance for the piece it is contained in. When this piece can be represented as a multi-dimensional time series (MDTS), such as acoustic or visual observations, finding a leitmotif is equivalent to the pattern discovery problem, which is an unsupervised and complex problem in time series analytics. Compared to the univariate case, it carries additional complexity because patterns typically do not occur in all dimensions but only in a few - which are, however, unknown and must be detected by the method itself. In this paper, we present the novel, efficient and highly effective leitmotif discovery algorithm LAMA for MDTS. LAMA rests on two core principals: (a) a leitmotif manifests solely given a yet unknown number of sub-dimensions - neither too few, nor too many, and (b) the set of sub-dimensions are not independent from the best pattern found therein, necessitating both problems to be approached in a joint manner. In contrast to most previous methods, LAMA tackles both problems jointly - instead of independently selecting dimensions (or leitmotifs) and finding the best leitmotifs (or dimensions). Our experimental evaluation on a novel ground-truth annotated benchmark of 14 distinct real-life data sets shows that LAMA, when compared to four state-of-the-art baselines, shows superior performance in detecting meaningful patterns without increased computational complexity.

## KEYWORDS

Multidimensional, Sub-Dimensional, Multivariate, Time Series, Motif, Motif Set, Leitmotif

### PVLDB Reference Format:

Patrick Schäfer and Ulf Leser. Discovering Leitmotifs in Multidimensional Time Series. PVLDB, 18(2): 377 - 389, 2024.  
doi:10.14778/3705829.3705852

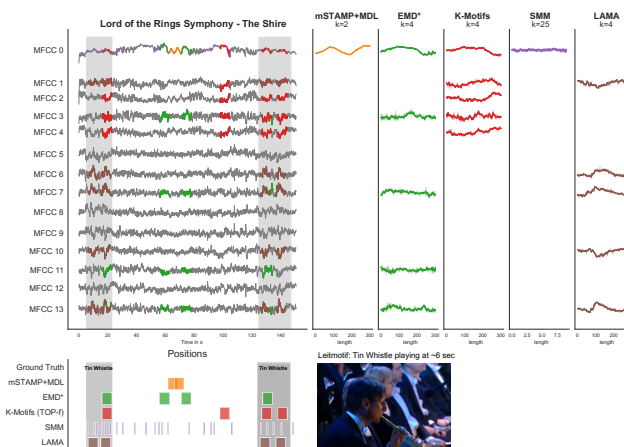
### PVLDB Artifact Availability:

The source code, data, and/or other artifacts have been made available at <https://github.com/patrickzib/leitmotifs>.

## 1 INTRODUCTION

A *leitmotif* (*leading motif*) is a recurring theme or motif that carries symbolic significance in various forms of art, particularly literature, movies, and music. The distinct feature of any leitmotif is that

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit <https://creativecommons.org/licenses/by-nc-nd/4.0/> to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing [info@vldb.org](mailto:info@vldb.org). Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.  
Proceedings of the VLDB Endowment, Vol. 18, No. 2 ISSN 2150-8097.  
doi:10.14778/3705829.3705852



**Figure 1:** *The Shire* theme played by the *Lord of the Rings* symphonic orchestra. The suite opens and ends with the *Hobbit* leitmotif, which is played on a solo tin whistle. Position of leitmotif found by LAMA (brown) compared to the ground truth (gray/bottom) and four competitor approaches (orange, green, red and purple). (Image best viewed in color)

humans associate them to meaning, which enhances narrative cohesion and establishes emotional connections with the audience. The use of (leit)motifs thus eases perception, interpretation, and identification with the underlying narrative [3]. A genre that often uses leitmotifs are soundtracks, for instance in the compositions of Hans Zimmer or Howard Shore. Figure 1 shows a suite from *The Shire* with 14 channels arranged by Howard Shore for *Lord of the Rings*. The suite opens and ends with the *Hobbits'* leitmotif, which is played by a solo tin whistle<sup>1</sup>, and manifests in a distinct pattern in several, but not all channels of the piece. However, the discovery of leitmotifs is not limited to soundtracks. We collected 14 real datasets from 6 domains and annotated leitmotifs.

Finding leitmotifs automatically is an unsupervised problem of high complexity, as patterns have an unknown length, occur an unknown number of times, and in an unknown number of dimensions. Furthermore, the different occurrences of a pattern are not identical but only similar to each other, where the degree of similarity may vary between the different dimensions. A specific problem is that in real-life data many dimensions often record only noise. Noise often is similar to itself and thus quickly leads to the detection of meaningless patterns. Finding the sub-dimensions that contain the meaningful pattern makes the problem distinct from the univariate

<sup>1</sup>"[Howard] Shore often arranges the Shire leitmotif for tin whistle or recorder cast in a simple homophonic texture, which can approximate the rustic countryside and carefree ways of the Hobbits." [17]

case [20]. It adds considerable complexity, as there exist  $d$  over  $f$  possibilities for a fixed number of dimensions  $1 \leq f \leq d$ , which is  $\binom{d}{f} \in \mathcal{O}(d^{\min(d-f, f)})$ .

This computational burden makes brute force enumeration of possible solutions infeasible and led to the development of heuristics. To date, three strategies have been proposed. Firstly, EMD [19] pre-process the MDTs with dimensionality reduction methods, such as principal component analysis (PCA), followed by a univariate pattern finding method. Thus, it first selects (or constructs) the subset of dimensions and then finds best patterns therein. Secondly, Minnen et al. [15] and Yeh et al. [22] describe methods that compute patterns in each dimension separately, then sort dimensions by pattern distance, and finally find a cut-off for determining the best  $f$  dimensions. Thirdly, Vahdatpour et al. [21] and Gao and Lin [6] also first seeks patterns separately in each dimension, and then determine the subset of dimensions in a bottom-up manner by picking dimensions with correlating pattern offsets. Our experiments (Section 4) with real data show that these approaches are highly noise-sensitive and often produce semantically irrelevant patterns.

In this work, we present the novel *LAtent leitMotif discovery Algorithm* (LAMA), which is designed to overcome the aforementioned challenges. LAMA is the first approach to perform dimension selection and motif mining in a joint fashion. Specifically, it seeks the set of  $k$  occurrences of a motif of length  $l$  that has minimal pairwise distances in a latent  $f$ -dimensional subspace, where  $l$  and  $k$  can either be user-specified or learned automatically from the data. It rests on two core principles. Firstly, leitmotifs emerge from a yet unknown subset of dimensions — which must be neither too few nor too many. Secondly, the set of sub-dimensions are not independent from the best pattern found therein, which implies that both problems should be approached in a joint manner. LAMA meticulously selects each subsequence as a potential leitmotif candidate and constructs a candidate set by considering the subsequence along with its  $k$ -nearest neighbors ( $k$ -NN). In a joint learning process, LAMA determines the most suitable dimensions for each candidate by assessing the  $k$ -NN distance within each dimension. This involves sorting the  $k$ -NN distances in ascending order and selecting the dimensions with the lowest distances. To find the best leitmotif, LAMA minimizes the pairwise distance among the constituent subsequences within all candidate sets of size  $k$ . This ensures the identification of the optimal Leitmotif across various sub-dimensional candidate sets.

Figure 1 shows the results of five core approaches LAMA, EMD\* [19]<sup>2</sup>, mSTAMP [22] using MDL for channel selection, SMM [16], and K-Motifs [12], with aggregated distances over the first 5 channels (same as for other competitors), details in Chapter 5. We use the same motif length  $l \approx 5s$  for all approaches to avoid bias. mSTAMP [22] identifies two occurrences of a seemingly straight line located in the loudness channel (MFCC 0), equal to a quiet section of the song. We reason this to be the fade out of one of the instruments within the Hobbits’ leitmotif. EMD\* finds a motif set with only 2 out of the 4 Hobbits’ leitmotif correctly identified, and 2 unrelated subsequences. Notably, even the similarity between subsequences of the same channels is low - see channels 7 or 11 resulting from the plot. K-Motifs use no channel explicit selection

<sup>2</sup>We re-implemented the relevant parts of EMD, as the source code is not available.

strategy. Using the first five channels it fails at identifying all occurrences with 3 out of 4 occurrences found, and one seemingly unrelated sub-sequence. SMM does not provide a way to define the motif length. Consequently, even with tuning, it detects 25 occurrences of what seems to be a short straight line in the loudness channel. Our LAMA is the only method to correctly identify 4 occurrences within the leitmotif. Other than EMD\*, LAMA’s occurrences show high pairwise similarity, too.

In summary, our paper makes the following contributions:

- (1) We introduce LAtent leitMotif discovery Algorithm (LAMA), an algorithm to uncover multi-dimensional leitmotifs. Notable, LAMA distinguishes itself as the pioneer by jointly tackling the dual challenge of selection sub-dimensions and discovering motif sets - unlike SotA, which prioritize one aspect over the other (Section 3.4).
- (2) A critical step in LAMA is the computation of a distance matrix of TS windows across dimensions. We show that a naive implementation would lead to excessive memory and present a new algorithm that reduces memory consumption by up to three orders of magnitude.
- (3) We describe innovative algorithms for automatically determining two of the three hyper-parameters of LAMA, which makes its application considerably easier compared to competitors, for which users have to define manually up to 7 user parameters.
- (4) We present the first annotated benchmark of 14 real-life MDTs, ranging from the distinctive punching motion in boxing, the fundamental steps of a dance routine, to the leitmotif of Darth Vader in Star Wars.

We assess the performance of LAMA through a novel benchmark comprising 14 real-life manually labelled MDTs datasets. Our evaluation contrasts LAMA against four state-of-the-art competitors: EMD\*, mSTAMP, SMM, and K-Motifs, examining both the accuracy of pattern discovery against annotated ground truth and the computational runtime. We analyze the impact of four distance measures and noise on accuracy. Additionally, we demonstrate LAMA’s superior leitmotif detection quality through three compelling case studies spanning diverse domains, including the Star Wars Imperial March (soundtracks), Rolling Stones (pop music), and Boxing Routines (motion data). We find that LAMA finds considerably more meaningful patterns (Leitmotifs) in each of these data sets.

The rest of the paper is organized as follows: In Section 2 we present the background and definitions. Section 3 presents our Leitmotif approach. Section 4 shows the utility of our method through-out various case studies. Finally, Section 5 gives an overview of related work for multidimensional motif set discovery

## 2 BACKGROUND

In this section, we first formally define time series (TS) and the  $z$ -normalized Euclidean distance ( $z$ -ED), which we (like all prior work) will use throughout this work. We then introduce the definition of Leitmotifs for sub-dimension motif set discovery.

**DEFINITION 1.** *Multivariate Time Series (MTS): A multivariate time series  $T = (t_1, \dots, t_n) \in \mathbb{R}^{(d \times n)}$  is a list of  $n$  vectors with each  $t_i \in \mathbb{R}^d$  being a vector of  $d$  dimensions (sometimes referred to as*

channels). We denote the  $i$ -th observation of the  $k$ -th dimension by the scalar  $t_i^{(k)} \in \mathbb{R}$ , and the  $k$ -th dimension by  $T^{(k)} \in \mathbb{R}^n$

If every point in  $t_i \in T$  in the TS represents a single value ( $t_i \in \mathbb{R}$ ), the series is *univariate*. If each point represents the observations of multiple variables (e.g., temperature, humidity, pressure, etc.) we call the TS *multivariate* or *multidimensional*.

**DEFINITION 2.** Multivariate Subseq.: A subsequence  $S_{i:l}$  of a MTS  $T = (t_1, \dots, t_n) \in \mathbb{R}^{(d \times n)}$ , with  $1 \leq i \leq n$  and  $1 \leq i+l \leq n$ , is a MTS of length  $l$  consisting of the  $l$  contiguous vectors, each with  $d$  dimensions, from  $T$  starting at offset  $i$ :  $S_{i:l} = (t_i, t_{i+1}, \dots, t_{i+l-1}) \in \mathbb{R}^{(d \times l)}$ .

Works in motif discovery (MD) typically exclude overlapping subsequences, as their distance is naturally low.

**DEFINITION 3.** Overlapping subsequences (Trivial Match): Two subsequences  $S_{i:l}$  and  $S_{j:l}$  of length  $l$  of the same TS  $T$  overlap iff  $[i - l \cdot \alpha] \leq j \leq [i + l \cdot \alpha]$  where  $\alpha \in [0, 1]$ .

By default, we set  $\alpha = 1.0$ , which excludes any overlapping subsequences sharing more than 1 value. However, this parameter can be adjusted through our API.

The z-normalized Euclidean distance (z-ED) is the prevalent distance measure for MD. It can be formulated as a dot-product. This facilitates rapid computation of the distance matrix among all pairwise subsequences sampled from a TS using MASS [25], with complexity  $\mathcal{O}(n^2)$ , regardless of subsequence length  $l$ . Internally, MASS calculates a rolling dot product in constant time. This enables the straightforward computation of z-ED, Euclidean distance (ED), cosine similarity (CS), or Complexity Invariant Distance (CID) [2]. Our implementation supports all four distance measures. For brevity, we will focus on z-ED. We first define z-ED within a single channel, then extend it to the multivariate case.

**DEFINITION 4.** z-normalized Euclidean distance (z-ED): Given two univariate subsequences  $A = (a_1, \dots, a_l)$  with mean  $\mu_A$  and standard-deviation  $\sigma_A$  and  $B = (b_1, \dots, b_l)$  with  $\mu_B$  and  $\sigma_B$ , both of length  $l$ , and dot-product  $A \cdot B$  their (squared) z-ED is defined as:

$$z\text{-ED}^2(A, B) = 2l \left( \frac{A \cdot B - l\mu_A\mu_B}{l\sigma_A\sigma_B} \right) \quad (1)$$

In the case of multivariate subsequences, we conceptually compute the distance between  $A$  and  $B$  in each dimension, separately, and sum up the result.

**DEFINITION 5.** Multivariate z-ED: Given two multivariate subsequences  $A = (a_1, \dots, a_l)$ , with  $k$ -th dimension  $A^{(k)} \in \mathbb{R}^{(n)}$ , and  $B = (b_1, \dots, b_l)$ , with  $k$ -th dimension  $B^{(k)} \in \mathbb{R}^{(n)}$  respectively, both of length  $l$  and dimension  $d$ , their multivariate distance is defined as:

$$\text{dist}(A, B) = \sum_{k=1}^d z\text{-ED}^2 \left( A^{(k)}, B^{(k)} \right) \quad (2)$$

Instead of using all  $d$  dimensions, we often want to compute the distance between two multidimensional subsequences using only a subset of dimensions  $X$ , with  $|X| = f \leq d$ .

**DEFINITION 6.** Sub-dimensional Distance: Given two multivariate subsequences  $A = (a_1, \dots, a_l)$  and  $B = (b_1, \dots, b_l)$ , both of length

$l$  and  $a_i, b_i \in \mathbb{R}^d$ , their sub-dimensional distance using only the subset of dimensions  $X$  is defined as:

$$\text{dist}^{(X)}(A, B) = \sum_{k \in X} \text{dist} \left( A^{(k)}, B^{(k)} \right) \quad (3)$$

where  $X$  is the subset of dimensions, and  $|X| = f$

To derive sub-dimensional patterns, the initial stage involves identifying the *pertinent sub-dimensions* that collectively *minimize the overall distance*. Given the number of dimensions  $f$  as input, the corresponding optimization problem is to find the closest subset  $X^*$  of dimensions, with  $|X^*| = f \leq d$ , for which the above distance between two subsequences  $A$  and  $B$  in Eq. 3 is minimal:

$$X^* = \min_X \text{dist}^{(X)}(A, B)$$

To avoid high similarity between overlapping subsequences extracted from a TS, we define  $e$ -matches as:

**DEFINITION 7.**  $e$ -match: Two subsequences  $A$  and  $B$  of  $T$  are called  $e$ -matching iff (a)  $\text{dist}^{(X)}(A, B) \leq e \in \mathbb{R}$  and (b) they are non-overlapping. A set  $S$  of subsequences of  $T$  is called  $e$ -matching, iff all subsequences in  $S$  are pairwise  $e$ -matching.

We next define the *extent* of a motif set.

**DEFINITION 8.** Extent: Consider a MTS  $T$ , a set  $S$  of subsequences of  $T$  of length  $l$ , and sub-dimensions  $X$ . The extent of  $S$  is the maximal pairwise distance of elements from  $S$ :

$$e = \text{extent}(S, X) = \max_{(A, B) \in S \times S} \text{dist}^{(X)}(A, B)$$

$k$ -Motiflets [18], are the set  $S^*$  of exactly  $k$  occurrences of a motif of length  $l$ , whose maximal pairwise distance is minimal.

$$S^* = \min_{S, |S|=k} \text{extent}(S) \quad (4)$$

A heuristic solution [18] for finding  $k$ -Motiflets involves a  $(k-1)$ -nearest neighbor search around each subsequence of a TS, and choosing the set  $S^*$  with minimal extent, where  $S^*$  includes the query and its  $k-1$ -NN.

Next, we introduce our novel concept, the *Leitmotif*, along with its distinctive minimization problem:

**DEFINITION 9.** Top Leitmotif: Given a multidimensional TS  $T$  of dimensionality  $d$ , and (i) parameters cardinality  $k \in \mathbb{N}$ , (ii) motif length  $l$ , and (iii) number of sub-dimensions  $f \leq d$ , the top Leitmotif of  $T$  is the set  $S$  with  $|S| = k$  subsequences of  $T$  of length  $l$  for which the following holds: All elements of  $S$  are pairwise  $e$ -matching, with  $e = \text{extent}(S)$ , and there exists no set  $S'$  with  $\text{extent}(S') < \text{extent}(S)$  also fulfilling these constraints.

**DEFINITION 10.** The Leitmotif optimization problem is then defined as finding the most similar (closest)  $|X^*| = f$  dimensions and  $|S^*| = k$  subsequences for which holds:

$$X^*, S^* = \min_{X, S} \text{extent}(S, X) \quad (5)$$

The Leitmotif optimization problem naturally extends the  $k$ -Motiflet optimization problem, which was originally defined for uni-dimensional motif set discovery [18]. When using a single dimensional TS as input, the two definitions are equivalent. As finding  $k$ -Motiflet exactly is likely NP-hard [18], heuristics are required.

### 3 LEITMOTIF DISCOVERY

We first address the challenges in finding sub-dimensional leitmotifs (Section 3.1). We then introduce our novel Leitmotif algorithm LAMA for discovering sub-dimensional motif sets in multivariate TS, providing an overview in Section 3.2 and our heuristic solution in Section 3.3. Section 3.4 discusses issues with sub-dimensional motif discovery and our solution. Algorithms to learn the input parameters  $k$  and  $l$  are presented in Section 3.5, followed by computational complexity analysis (Section 3.6). Finally, we discuss scalable Leitmotif discovery (Section 3.7).

#### 3.1 Finding Sub-Dimensional Leitmotifs

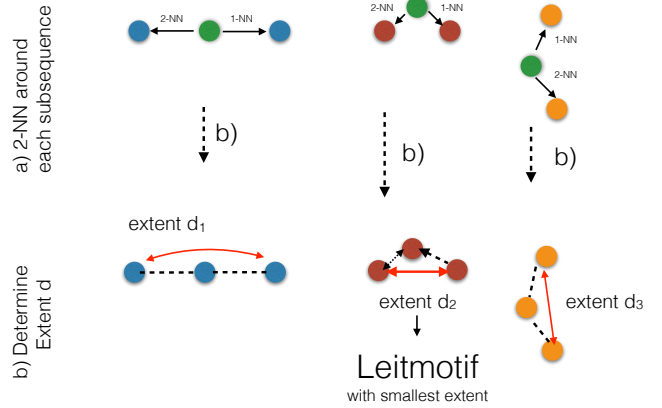
We’ve distilled two fundamental principles guiding leitmotif discovery. Firstly, leitmotifs emerge from an elusive, yet-to-be-defined subset of dimensions, striking a balance between not being too sparse nor too abundant. Secondly, the sub-dimensions set must be individually uncovered for each leitmotif candidate. For instance, consider the solo tin flute consistently playing within the same channel subset across all instances in our motivational example. LAMA addresses both challenges simultaneously, avoiding the sequential approach of dimension learning followed by motif discovery. Given the computational infeasibility of exhaustively enumerating all sub-dimensions, we use heuristics. These heuristics prioritize either channel selection or motif discovery first. We’ll explore the advantages and limitations of the state-of-the-art approach [22] for set discovery in Section 3.4.

#### 3.2 Overview of LAMA

Mathematically, we defined a *Leitmotif* as a collection of  $k$  approximate occurrences of a subsequence with a length of  $l$ , with minimal extent in sub-dimensional space. In a first pre-processing phase, the length  $l$  and the number of repetitions  $k$  of a motif are learned from sensible input ranges (see Section 3.5. Figure 2 illustrates the core idea behind LAMA for computing a leitmotif with 3 occurrences. LAMA iteratively perform two steps for each subsequence  $q$  (in green): (a) search for the  $(k - 1)$ -NN of  $q$ , and (b) determine their (sub-dimensional) pairwise extent of the candidate set. Finally, the set with minimal (sub-dimensional) extent is returned (in red at bottom).

The workflow for jointly detecting sub-dimensions with leitmotif discovery in LAMA is illustrated in Figure 3, and will be detailed in the following sections. It shows 4 dimensions of a penguin diving from [26]. For each multivariate query subsequence (Figure 3 TOP in orange), LAMA performs a 2-NN search (Figure 3 (a)) in each dimension. It next sorts the dimensions by the distance of their 2-th NN (Figure 3 (b)), which results in the order of Z-Acc, X-Acc, Y-Acc, Pressure. Finally, LAMA selects the best  $f$  dimensions (i.e.  $f=2$ ), based on lowest distance, from the TS. We may then take either of the  $k$ -NNs in dimension Z-Acc or X-Acc as Leitmotif candidate. We decided to take the first dimensions  $k$ -NNs as candidate (see Section 3.6.1 for the rationale).

To find the best Leitmotif, LAMA minimizes the extent over all sub-dimensional candidate sets of size  $k$ . For each candidate set, an optimal set of sub-dimensions is determined.



**Figure 2:** Depicted are three sets (blue, red, orange), centered around a query (green point) and its 2-NN. Leitmotif discovery involves two steps: (a) 2-NN search around each query subsequence, and (b) determine the extent of each set (red arrow), i.e.  $d_1, d_2, d_3$ . Finally, the top Leitmotif with smallest extent  $d_2$  is returned.

---

#### Algorithm 1 Compute Leitmotifs using LAMA

---

```

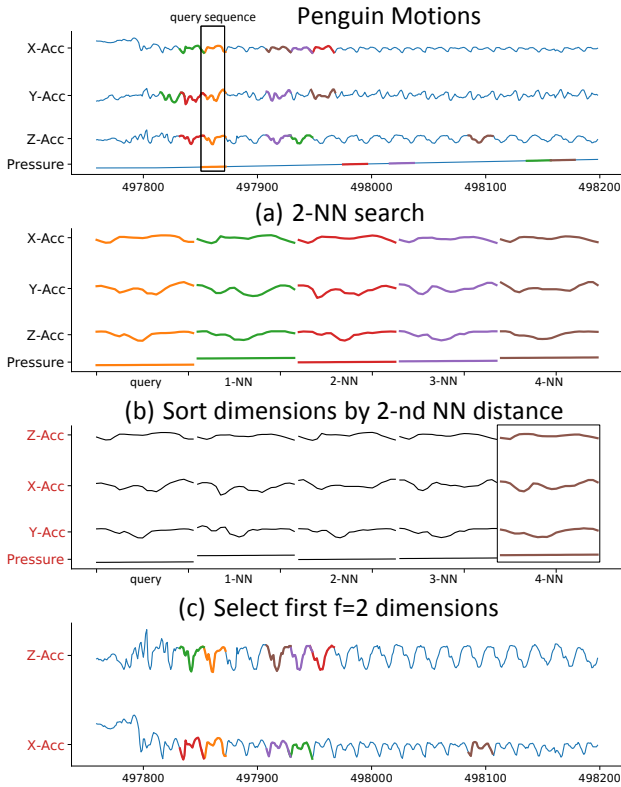
1: procedure RUN_LAMA( $T, D, f$ )
2:   ( $l, k$ )  $\leftarrow$  LEARN_PARAMETERS( $T, D$ )
3:   return LAMA( $T, D, f$ )
4: end procedure
5: procedure LAMA( $T, D, l, k, f$ )
6:    $d \leftarrow$  dimensions( $T$ )
7:    $n \leftarrow$  len( $T$ ) -  $l + 1$ 
8:   ( $leitmotif, e$ )  $\leftarrow$  ( $\{\}, inf$ ) ▷ bsf of extent
9:   for  $i \in [1, \dots, n]$  do
10:    ▷ Stage 1: Determine sub-dimensions
11:     $kNNs \leftarrow$  matrix of size  $d \times k$ 
12:    for  $dim \in [1, \dots, d]$  do
13:       $kNNs[dim] \leftarrow$  NON_TRIVIAL_ARGKNN( $D[dim, i], k$ )
14:    end for
15:    ▷ Stage 2: Check for leitmotif
16:     $f\_dims, candidate \leftarrow$  SELECT_F_DIMENSIONS( $D, kNNs, i, f$ )
17:     $dist_{knn} \leftarrow \sum_{dim \in f\_dims} D[dim, i, kNNs[dim, k]]$ 
18:    if  $dist_{knn} < e$  then
19:       $dist \leftarrow$  PAIRWISE_EXTENT( $D, f\_dims, candidate, e$ )
20:      if  $dist < e$  then
21:        ( $leitmotif, e$ )  $\leftarrow$  ( $candidate, dist$ )
22:      end if
23:    end if
24:  end for
25:  return ( $leitmotif, e$ )
26: end procedure

```

---

#### 3.3 Latent leitMotif discovery Algorithm (LAMA)

LAMA continuously undergoes the joint tasks for each leitmotif candidate: (a) it identifies the set of relevant sub-dimensions, and (b) the extent is minimized to find the best (TOP) leitmotif. Algorithm 1 illustrates our approach for discovering the Leitmotif. The LAMA procedure takes as input the TS  $T$ , the pairwise distance matrix  $D$ , of dimensionality  $d \times (n-m+1) \times (n-m+1)$ , computed separately along



**Figure 3: Dimensionality Reduction Workflow used in LAMA.** **Top:** a query sequence is selected. **(a)** For each of query sequence, a  $k$ -NN search is performed along each dimension. **(b)** The dimensions get sorted by the distance of the  $k$ -th NN. **(c)** The first  $f$  dimensions are kept for this sequence.

each dimension of the TS, and a hyper-parameter for the number of sub-dimensions  $f$  to use. The two parameters cardinality  $k \in \mathbb{N}$ , and motif length  $l \in \mathbb{N}$  are learned from the TS (Section 3.5). Using  $D$ , we can compute full or sub-dimensional pairwise distances between any pair of multivariate subsequences using Eq. 2. To make use of admissible pruning, we use the observation that the  $k$ -NN distance of a subsequence can be used to derive a lower and an upper bound on the extent of the candidate set, which effectively reduces the number of extent computations. Specifically, the extent must be at most 2 times the distance of the  $k$ -NN distance:

$$e \leq \text{dist}_{kNN} \leq 2 \cdot e$$

Intuitively speaking, the extent of a motif set must be between the radius and diameter, when two subsequences are at opposite bounds, of a hypersphere centered around the query subsequence (Figure 2).

LAMA applies admissible pruning to reduce the number of candidate sets using the lower bound on the extent, using  $k$ -NN distances  $\text{dist}_{knn}$ , and the current best-so-far extent  $e$ . The algorithm processes all subsequences (line 9) sequentially in a greedy fashion. It approximates the extent of the optimal set of subsequences using the  $(k-1)$ -NNs of a query to form a candidate set for the Leitmotif.

The NN have to be non-overlapping to each other to avoid any trivial matches in extent computation. In `NON_TRIVIAL_ARGKNN()` (line 13) we order the subsequences by their distance to the  $i$ -th subsequence and return the closest  $(k-1)$  non-trivial, i.e. non overlapping, neighbours in each dimension. Using these NNs we select the  $f$  best dimensions using `SELECT_F_DIMENSIONS()`, line 16. We will discuss this in the following Section 3.4. For the time being, assume we selected all dimensions to be used from  $D$ . The distance of the  $(k-1)$ -th NN is used for admissible pruning (line 18ff). The extend of a candidate set has to be checked, if and only if  $\text{dist}_{kNN}$  is lower than the current best-so-far extent  $e$ . If successful, we determine the pairwise extent of this set (line 19), again using the best  $f$  dimensions. To speed up computations, we apply admissible pruning in `PAIRWISE_EXTENT(_, _, e)`, too, by stopping the computation once any pairwise distance exceeds  $e$ . If the overall extent  $\text{dist}$  is smaller than the best-so-far  $e$ , we update the Leitmotif (lines 21). Once all  $n$  candidate offsets have been processed, we return the Leitmotif and its extent  $e$ .

### 3.4 Sub-dimensional Leitmotifs

Selecting the optimal sub-dimensions to minimize the extent of the Leitmotif is a challenging task. Theoretically, the process of choosing  $f$  out of  $d$  dimensions results in  $d$  over  $f$  possibilities, which is in  $\binom{d}{f} \in \mathcal{O}(d^{\min(d-f, f)})$  for fixed  $f$ . This computational burden makes direct computation practically infeasible.

**3.4.1 Independent Dimensionality-Selection using Pairs.** Fortunately, a log-linear solution, leveraging sorting techniques, has been proposed for addressing this challenge, particularly when minimizing the distance between pairs of subsequences in [15]. This approach has found success in *pair motif* discovery [22]. However, its efficacy diminishes when applied to the task of identifying *motif sets*, and we will delve into the reasons behind its limitations. Subsequently, we propose an extension, enabling the discovery of motif sets (Leitmotifs) in log-linear time in dimensions  $d$ .

The extent is defined as the maximum over all pairs of distances within the motif set  $S$ , and we need to find those sub-dimensions  $X$ , that minimize the overall extent to find the optimal set  $S^*$ . The objective is given by:

$$\min_{S, X} \text{extent}(S, X) \quad (6)$$

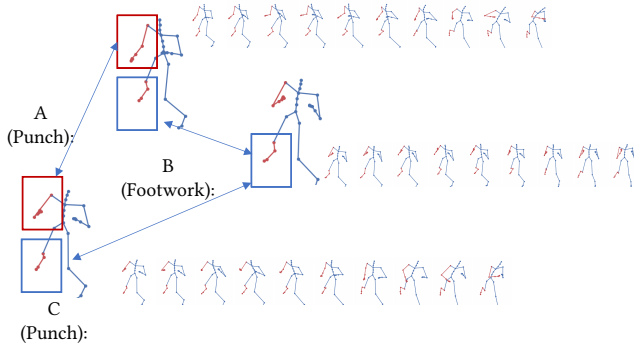
This objective becomes minimal, by expanding the definition of the extent, and find the best sub-dimensions for each pair of sub-sequences, such that:

$$\begin{aligned} S^* &= \min_{S, X} \text{extent}(S, X) \\ &= \min_S \left( \max_{(A, B) \in S \times S} \left( \min_X \text{dist}^{(X)}(A, B) \right) \right) \end{aligned} \quad (7)$$

This formulation entails first selecting a motif set  $S$ , followed by selecting sub-dimensions  $X$  with minimal distance for each pair of subsequences, to compute the maximum extent of the set. Minimizing yields the motif set  $S^*$  with overall lowest extent.

Furthermore, this formulation allows for a log-linear efficient implementation. For a multi-dimensional TS, with time and channel axis, finding the  $k$ -NN involves an ascend sort by distance along





**Figure 4: The figure shows three boxing routines A, B, C (each from left to right), which are unintuitively considered similar when independently optimizing/selecting dimensions between pairs (compare Eq. 7). While A and C show a punching motion, in B the actor simply turns his body around his own axis making small steps. (A, B) and (B, C) show similarity in steps, while (A, C) is similar in the punching motion. (Image best viewed in color)**

time within each channel, and choosing the first  $k$ -th entries. This is followed by sorting along the channels in ascending order, and always choosing the first  $f$  channels with smallest channel distance [15, 22]. This works well for pair motifs, i.e. the 1-NN [22]. Yet, it tends to find less intuitive motif sets, as it independently optimizes motif discovery and channel selection, as discussed next.

Consider the example illustrated in Figure 4. The example shows a boxing motion capture. Employing sub-dimensional leitmotif search, we find three subsequences denoted as A, B, C. Notably, subsequences A and B exhibit similarity in the punching motion (red). (B, C) and (A, B) share similarity in the foot-steps (blue). Despite minimal pairwise similarity in the selected dimensions, the overall motif set encompasses three distinct motions. This is a result of independent motif set discovery and channel selection.

**3.4.2 Joint Dimensionality Selection using Sets.** To address the aforementioned issue, we reformulate the optimization target. Our objective is to identify sub-dimensions  $X^*$ , with  $|X^*| = f$ , and subsequences  $S^*$ , with  $|S^*| = k$ , that exhibit highest similarity among all subsequences of the same set:

$$X^*, S^* = \min_X \left( \min_S \left( \max_{(A,B) \in S \times S} \text{dist}^{(X)}(A, B) \right) \right) \quad (8)$$

How are the two equations, Eq.7 and Eq.8, different? Let’s first consider a toy example with two sets of distances to illustrate the effect of swapping min and max: Set  $S_1 : \{1, 3, 5\}$  and Set  $S_2 : \{2, 4, 6\}$ . First, compute the max of each set:  $\max(S_1) = 5$  and  $\max(S_2) = 6$ . The min of these is 5. Next, compute the min of each set:  $\min(S_1) = 1$  and  $\min(S_2) = 2$ . The max of these is 2. By moving the max operation outside, we get a higher total distance, thus *extent* of the leitmotif.

Conceptually, when moving the minimization target sub-dimensions  $X$  outside the distance computations, as in Eq.8, we jointly learn a fixed set of sub-dimensions that lead to an optimal

---

### Algorithm 2 Sub-Dimension Selection ( $f$ out of $d$ )

---

```

1: procedure SELECT_F_DIMENSIONS( $D, kNNs, i, f$ )
2:    $d \leftarrow \text{dimensions}(kNNs)$ 
3:    $\triangleright$  1. Get  $k^{th}$ -NN distance by dimension
4:    $distkNN \leftarrow \text{Array of size } d$ 
5:   for  $dim \in [1, \dots, d]$  do  $\triangleright$  ( $k-1$ )-th NN by dimension
6:      $distkNN[dim] \leftarrow D[dim, i, kNNs[dim]]$ 
7:   end for
8:    $\triangleright$  2. Sort and take best  $f$  dimensions
9:    $f\_dims \leftarrow \text{FIRST}_f(\text{ARGSORT}(distkNN))$ 
10:   $\triangleright$  3. Use first (lowest distance) dimension as candidate
11:   $candidate \leftarrow kNNs[f\_dims[0]]$ 
12:  return ( $f\_dims, candidate$ )
13: end procedure

```

---

candidate leitmotif. This approach ensures that motif candidates share a common set of dimensions, such as the same feet or hand movements, to be considered similar. In contrast, Eq.7 identifies independent sub-dimensions for each pair of subsequences. This results in subsequences to be considered similar if any part of the body performs a similar movement between any two subsequences, even if these parts differ between other pairs of subsequences. Overall, this results in a less cohesive leitmotif.

**3.4.3 Solution.** Regrettably, there is currently no optimal and efficient solution available to address this joint minimization problem, necessitating the enumeration of all  $d$  over  $f$  solutions to identify the optimal one. An effective approximation is to select the best  $f$  dimensions based on minimizing the distance between the query and the  $(k - 1)$ -NN, which we will present next.

**3.4.4 Algorithm.** Algorithm 2 illustrates our approach. It takes as input the pairwise distance matrix  $D$ , of dimensionality  $d \times (n - m + 1) \times (n - m + 1)$ , computed separately along each dimension of the TS, the  $kNNs$  along each dimension, with dimensionality  $d \times (n - m + 1) \times k$ , the position  $i$  of the query sequence in the TS, and the number of dimensions  $f$  to choose from  $d$ .

It first extracts the distance for the  $(k - 1)$ -th NN in each dimension (lines 5–7) prior to sorting these distances (line 9) and choosing the first  $f$ . From these  $f$  dimensions, the set of  $(k - 1)$ -NN from the first dimension is used as Leitmotif candidate (lines 11-12).

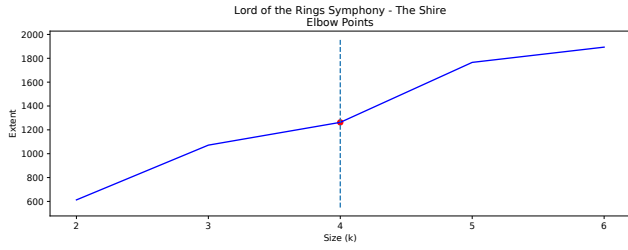
## 3.5 Learning Length and Set Size from the Data

We next present methods to learn suitable values for the Leitmotif length  $l$  and set size  $k$  to ease the discovery of motifs can be found without domain knowledge. These are based on an analysis of the extent function for different  $k$ , as introduced in [18]:

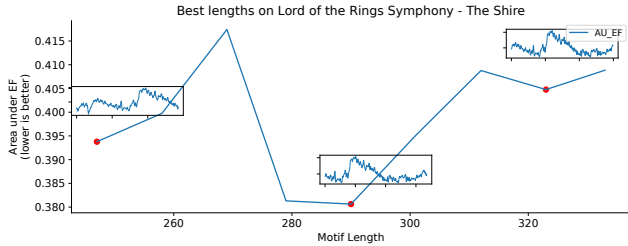
**DEFINITION 11.** Extent Function (EF): Assume a fixed length  $l$  and a TS  $T$ . Let  $S_k$  be the top  $k$ -Motiflet with length  $l$  of  $T$ . Then, the extent function  $EF$  for  $T$  is defined as  $EF(k) = \text{extent}(S_k)$ .

Using the EF lead to the following two ideas [18]:

1. Elbow points in the  $EF$  indicate changes of motifs. The last value of  $k$  before the elbow indicates a maximal motif, and is considered a meaningful value for  $k$ . Figure 5 shows the introductory *The Shire* use case. An elbow is detected at  $k = 4$  equal to the four repetitions of the leitmotif.



**Figure 5: The EF is a function of the cardinality of motif set to its extent. Elbow points represent large changes in similarity of the found motif, indicative of a concept change.**



**Figure 6: Area under the elbow function points to interesting lengths at local minima.**

2. Long flat stretches of the EFs indicate a high number of occurrences of a motif, but depend on the motif length  $l$ . Accordingly, window lengths  $l$  leading to long flat stretches are considered as particularly meaningful, which can be measured using the area under the EF (AU-EF).

Variable-length motif discovery involves two-fold: computing motifs along a range of lengths, and ranking the motifs found to reduce redundancy. For variable-length Leitmotifs, we compute the motif along a range of motif lengths  $lin[l_{min}, l_{max}]$ , and report those lengths, which show a local minimum in the AU-EF plot. These represent lengths, where the elbow-function is flat, and thus contain a motif set with the largest possible number of repeats. The concept is illustrated in Figure 6 for *The Shire* dataset.

### 3.6 Computational Complexity

We will next analyse the complexity of algorithm 1. It consists the building blocks: (a) extracting  $k$ -NN along each dimension, (b) computing the extent, and (c) selecting sub-dimensions.

*Regarding (a):* Extracting  $k$  nearest neighbors in  $d$  dimensions over all subsequences takes  $O(d \cdot n^2)$  using linear time selection (Quick-Select+Median of Medians) on each subsequence.

*Regarding (b):* Selecting  $f$  out of  $d$  sub-dimensions involves sorting the distances of the  $k-1$ -th NN along  $d$  dimensions, which is performed  $O(n)$  times. Using MergeSort, this totals  $O(n \cdot (d \log d))$ .

*Regarding (c):* The computation of the pairwise extent takes  $O(d \cdot k^2 \cdot n)$  in the worst case, if admissible pruning is not successful. The best case is in  $O(d \cdot k^2)$ , if the first candidate leads to admissible pruning of all other candidates.

Overall, the complexity of the algorithm using  $d$  dimensions is in  $O(d \cdot (n^2 + n \log d + k^2))$  in the best case, and  $O(d \cdot (k \cdot n^2 + n \log d + k^2))$  in the worst case.

**3.6.1 Theoretical Bounds.** The extent must be at most 2 times the distance of the  $k$ -NN distance. Figure 2 left illustrates the worst case, when two subsequences are at opposite sides of the hyper-sphere. Thus, the  $k$ -NN distance serves as an upper bound for the extent:  $e \leq \text{dist}_{kNN} \leq 2 \cdot e$

By sorting distances of the dimensions for each subsequences in ascending order, and choosing the sub-dimensions based on the  $k$ -NN (Algorithm 2 line 9), we minimize this upper bound (the  $2 \cdot e$  part). I.e. choosing a dimension with a larger  $k$ -NN distance yields a potentially larger upper bound. There is no guarantee though, as choosing a different subset of dimensions may still yield a smaller extent  $e$  of the Leitmotif. A same argument holds for choosing a different set of candidates other than from the first dimension (Algorithm 2 line 11). Overall, this approach is a 2-approximation to choosing the optimal subset of dimensions. Future work might include algorithms to provide better bounds.

### 3.7 Scalable Motif Discovery

One inherent limitation of the LAMA algorithm is scalability in terms of memory, which is quadratic in the size of the input. A (squared) distance matrix of a 30k (50k) TS requires roughly 3 GB (9 GB) of RAM. Thus, main memory quickly becomes a bottleneck for MD on large TS. A non-sufficient naive approach to reduce the memory, would be to only store the distance of each subsequence to its  $k-1$ -NN, aka,  $k$ -dimensional distance profile. However, to compute the extent of a candidate set, we need to compute the distance between all pairs of subsequences (compare Eq. 8), i.e. the distance between each NN.

The core idea is to compute the distance matrix twice in a row wise manner, always keeping only the distances in that row. When computing distances row-by-row, we only have access to the distance of the current subsequence  $i$  to its  $k$ -NN. Within these  $k$ -NNs there are pairs for which we have not seen any distance yet, i.e. when  $j > i$  with  $j \in k\text{-NNs}(i)$ . In the first pass, we compute the distance matrix row-by-row. For each row, we memorize the required pairwise distances from the candidate set. These offsets can point to preceding or succeeding rows of the distance matrix, which are not accessible. In the second pass, we iterate the matrix in a row-by-row manner again, and keep only those distances that were memorized in the first pass. This optimization reduces memory complexity by 2 to 3 orders of magnitude, at the cost of doubling the distance matrix computation complexity. To reduce runtime complexity, we case study finding leitmotifs in a TS segmentation in Section 4.5.

## 4 EXPERIMENTAL EVALUATION

Our experimental evaluation is two-fold. Firstly, in Section 4.2 we compare our LAMA against SotA in a **quantitative** analysis using precision and recall on a benchmark dataset with manually labelled ground truth. Secondly, in Section 4.3 we compare methods in terms of the **quality** of results on multiple case studies.

#### 4.1 First Annotated Leitmotif Benchmark Set

No public benchmark exists for sub-dimensional motif discovery. Table 1 provides an overview of the 14 datasets used in our evaluation. The Boxing, Ice Ice Baby, and Vanilla Ice datasets were used in previous studies [8, 22] without ground truth annotation. The other datasets are original use cases. We selected datasets where the leitmotif can be easily verified, such as audio or video, where refrains or common motions are recognizable by humans but challenging for algorithms. We manually inferred the ground truth positions for each dataset and provided gold standard parameters, including the size  $k$  of the motif set, motif length, and the number of relevant dimensions  $f$ .

**Four Motion Captures** Recordings originate from the CMU archive [9]. Each consists of 31 joints, subdivided into x, y, z coordinates, with a total of 93 dimensions, at a frame rate of 120 frames/second. *Charleston* contains a recording of a female subject dancing 20’s Charleston, a partner dance originated in the golden 20’s. The leitmotif is the basic footstep with swinging feed and arm movements, which is performed three times with a duration of 1.3s. *Boxing* captures of boxing routine. The leitmotif is a punch with the right arm with 10 repetitions of 1.6s each. *Swordplay* is a recording of a series of Japanese sword strokes. Each sword movement is distinct, but the common leitmotif is the raising of the sword above the head with 6 repetitions of 1s. *Basketball* shows the dribbling of a basketball, while an actor is moving forwards. The dribbling motion is repeated five times, which 0.42 each.

**Two soundtracks:** A genre that often uses leitmotifs are soundtracks. Each song has 20 MFCC channels. *Howard Shore* uses a large number of leitmotifs to support characters or the landscape of Middle-earth. The song *The Shire* contains the Hobbits’ leitmotif with four occurrences of 8 – 9s each. *John Williams* used leitmotifs to represent people and concepts in Star Wars, i.e., one leitmotif is played in the presence of Darth Vader. This leitmotif in the *Imperial March* is the trumpets playing, which is repeated 5 times with 8.3s.

**Five pop songs:** In pop songs a Leitmotifs can be a refrain, chorus, a rhythm section, or the bass line. Each song has 20 MFCC channels. *Rolling Stones’ Paint it Black* was released in 1966 on the *Aftermath* album. The sitar is featured prominently throughout the song, which adds a unique Middle Eastern flair. The verses of the song breakdown in a similar verse structure of B, C, B, C, B, C, D, C, B, C, E, F, where the C section is repeated most often. Its leitmotif is roughly 5s long and repeated twice per C section with a total of 10 times. *Linkin Park - Numb* was released in 2003 on the *Meteora* album. The refrain is repeated 6 times with 5.5s each. *Linkin Park - What I’ve done* was released in 2007 on the *Minutes to Midnight* album. The refrain is repeated 6 times with 10s each. *Queen - Under Pressure* was performed by Queen and David Bowie, and published in 1981 on the *Greatest Hits* album. It is most famous for its baseline, which was repeated by *Vanilla Ice - Ice Ice Baby* with a minor modification. This baseline is repeated 16 times for *Under Pressure* and 20 times for *Ice Ice Baby*, and 4.2s.

**Wildlife Recording:** We used a wildlife recording of birds in their natural habits, provided by the Royal Society for the Protection of Birds (RSPB) [5]. It has 20 MFCC channels. Bird vocalizations have been a challenging test case for TS MD [7]. The common starling is a noisy bird, with the male being the primary singer. Its

**Table 1: Datasets used in experiments. Ground leitmotifs were manually inferred. GT refers to the number of leitmotif occurrences.**

Use Case	Category	Length	Dim.	GT
Charleston [9]	Motion Capture	506	93	3
Boxing [9]	Motion Capture	4840	93	10
Swordplay [9]	Motion Capture	2251	93	6
Basketball [9]	Motion Capture	721	93	5
LOTR - The Shire	Soundtrack	6487	20	4
SW - The Imperial March	Soundtrack	8015	20	5
RS - Paint it black	Pop Music	9744	20	10
Linkin Park - Numb	Pop Music	8018	20	5
Linkin P. - What I’ve Done	Pop Music	8932	20	6
Queen - Under Pressure	Pop Music	9305	20	16
Vanilla Ice - Ice Ice Baby	Pop Music	11693	20	20
Starling [5]	Wildlife Rec.	2839	20	4
Physiodata [24]	Wearable Sensors	5526	5	20
Bitcoin Halving	Crypto/Stocks	3591	7	3

song comprises various melodic and mechanical sounds arranged in a ritualistic succession. The leitmotif contains repeated clicks and concludes with a burst of high-frequency sounds of 4 repetitions of 1.3s each.

**Crypto/Stocks: Bitcoin Halving Event:** This dataset includes daily closing values for seven cryptocurrencies over ten years, focusing on the price movements of Bitcoin following its halving events. Historically, Bitcoin prices rise significantly after each halving, typically over the course of a year. The leitmotif contains the 3 halving events that affect Bitcoin and Litecoin with 1 year duration.

**Wearable Sensor Data: Physiodata:** This dataset comprises inertial and magnetic sensor data from wearable devices during physical therapy exercises [24]. It includes an exercises performed in three ways (correct, fast, low-amplitude). The motif is repeated 20 times and 6s long.

**Competitors:** Achieving reproducibility proved to be a major challenge. Many of the approaches have roots dating back one or two decades [15, 19, 21], and unfortunately, none of the codes for Multidimensional Motif Set Discovery (MMSD) competitors have been publicly shared. We reached out to the two competitors within the last 5 years, namely TMDM [1] and SMM [16] asking for code, but only obtained the code for SMM. We use the following approaches: (a) *mSTAMP (MDL)*, a pair motif approach, which uses the Minimum Description Length (MDL) for sub-dimension selection, (b) *mSTAMP* using the  $f$ -dimensions from ground truth annotations, (c) *EMD\** [19], an approach using PCA for channel reduction followed by univariate motif set discovery, (d) *K-Motifs (TOP-f)*, a univariate method that aggregates the distances for the first  $f$  channels, and (e) *K-Motifs (all dims)* that aggregates distances over all channels, and (f) *SMM*.

**Parameters:** MD is an unsupervised learning task. LAMA has 1 hyper-parameter with the number of dimensions  $f$ , and two parameters ( $k$  and  $l$ ) that can be learned from the data given parameter ranges without access to any labelling. If learning was applied, the concrete ranges and learned parameters are defined within each experiment. For the experiment with ground truth available, we use



the associated parameters as input to all competitors. For SMM we tested the ranges:  $DEGAUSSIANTHRES$ : from 0 to 1,  $DELEVELTIME$  4, 6, 8,  $DELEVELDEPD$ : 4, 6, 8,  $DESIGMADEPD$ : 0.4, 0.5, 0.6, and  $R$ : 5 and 10. We provide source codes and results on our [website](#) [11].

## 4.2 Quantitative Comparison

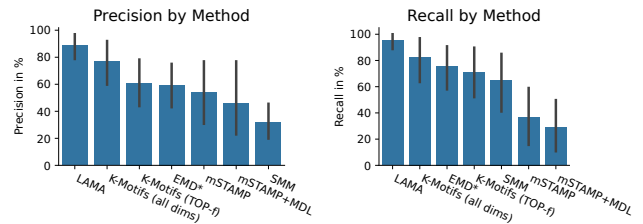
To measure the accuracy of the different approaches, we created a benchmark of ground-truth annotated datasets (Table 1) by manually inspecting the data for semantic motif sets. We use the gold standard parameters as input to the methods: the size  $k$  of the motif set, and the length of the motif set, or the number of relevant dimensions  $f$ . We assume that a method finds a motif set if the reported motif interval overlap with a ground truth interval by at least 50%. Figure 7 shows the results as (a) *precision*, measured as percentage of correct prediction out of all predictions, and (b) *recall*, measured as percentage of ground-truth annotations found.

LAMA has by far the highest precision and recall, see Table 2 with a margin of 12 to 42 percentage points for mean precision and 12.5 to 66 for mean recall. This underlines that jointly finding sub-dimensions and motifs is crucial for capturing leitmotifs. LAMA performs worst for the datasets *Vanilla Ice - Ice Ice Baby* and *Queen David Bowie - Under Pressure*, which both contain the famous bass line. The competitors struggle to find ground truth annotations. The best competitor is the univariate method *K-Motifs (all dims)*, which simply sums the distances computed over each dimension to derive a univariate signal. *K-Motifs (TOP-f dims)*, which selects the TOP- $f$  dimensions to sum up distances, has a significantly lower precision and recall. This underlines the need for choosing the ideal subset of dimensions rather than any subset. *mSTAMP* in both variants shows by far the worse mean and median recall, with median recall being 0 and 20, as it must only return pairs of motifs, whereas the motifs range from 4 to 20 occurrences (Table 1). Mean precision is slightly better, but still the worst among competitors. Channel selection using MDL is biased towards selecting noisy or near constant channels. *EMD\** globally selects dimensions using PCA, based on explaining the variance of the whole TS. This fails to capture the semantic motif sets, which is captured in small sections of a TS. SMM has 7 hyper-parameters that require tuning. Despite tuning efforts, SMM consistently returned between 1 and 50 motifs per dataset, each with up to 80 locations. We selected the best-fitting motif for each dataset, which inherently provides an advantage. Nevertheless, SMM’s precision remains suboptimal. Further analysis revealed that many motifs have a length of 1, making them single points rather than subsequences, and they are limited to a single dimension. Even after seeking guidance from the author, tuning the seven parameters of SMM remains the sole bottleneck.

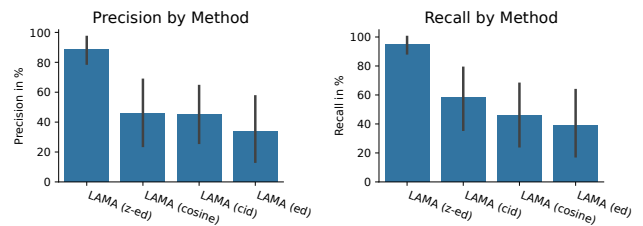
*Distance Measures:* We conducted experiments using the four implemented distance measures and LAMA. Figure 8 presents the results. Overall, the z-normalized ED (z-ED) achieved the highest precision and recall on our benchmark, followed by cosine distance (CD), complexity invariant distance (CID), and standard ED. The results indicate that ED and CD are highly sensitive to low-variance segments of the signal, such as the resting phase in the physio-data. CID performs slightly better as it compensates for low variance. One exception is *Vanilla Ice*, where ED outperforms z-ED by effectively capturing the baseline.

**Table 2: Precision and Recall by Method.**

Method	Precision		Recall	
	mean	median	mean	median
EMD*	59.3	65.0	75.9	80.0
K-Motifs (TOP-f)	61.1	70.0	70.8	<b>100.0</b>
K-Motifs (all dims)	76.8	83.3	82.6	<b>100.0</b>
mSTAMP	53.9	<b>100.0</b>	36.7	20.0
mSTAMP+MDL	46.2	0.0	29.0	0.0
SMM	31.8	26.5	65.4	95.0
LAMA	<b>88.7</b>	<b>100.0</b>	<b>95.1</b>	<b>100.0</b>



**Figure 7: Use cases with inferred semantic leitmotifs: Mean Precision (left) and Mean Recall (right) of the found motif sets by MD method. LAMA performs best.**



**Figure 8: Experiment with four different distance measures on benchmark dataset. z-ED performs best.**

*Influence of Noise:* Finally, we conducted experiments with increasing noise levels by adding Gaussian noise to the series, up to 50% of the dataset’s standard deviation (Figure 9). The addition of noise makes it progressively harder to distinguish between the leitmotif and low variance parts of the series. Consequently, the precision and recall of almost all methods decrease as the noise level increases, complicating the detection. Despite this, LAMA consistently achieves the highest scores across all noise levels. As mSTAMP initially identified low-noise sections of a signal irrelevant to the leitmotif, this makes it the least affected by noise.

## 4.3 Qualitative Comparison

In this section we discuss the quality of the discovered motif sets. The purpose is to compare methods not only by precision and recall of found motifs as in the previous section, but also to consider whether these motifs are actually meaningful, i.e., correspond to important concepts in the TS. For each experiment we first ran

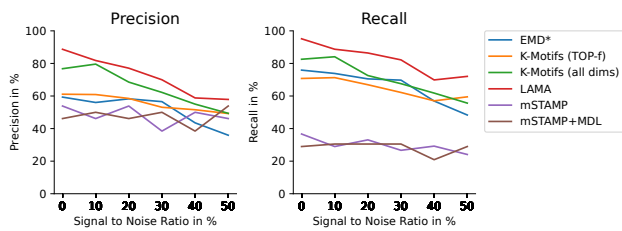


Figure 9: Experiment with increasing noise levels.

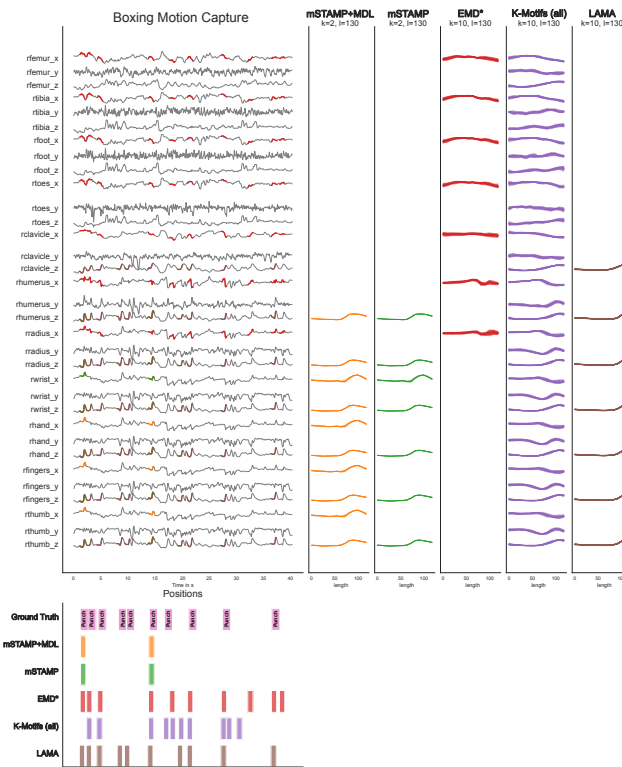


Figure 10: Punching motions found in a motion capture of a Boxing routine. (Image best viewed in color)

LAMA to learn optimal value for the length of the leitmotif  $l$ . As other methods offer no method for learning input parameters, we then use this length  $l$ , the number of occurrences  $k$ , and the number of channels  $f$  associated with the ground truth of each dataset as input to all methods for a fair comparison. SMM returns up to 50 motifs, with the best matching motifs consisting of sets of dozens of points that are barely visible to the eye (compare Figure 1). Thus, we chose to display SMM results only on our webpage.

**Motion Capture.** In this experiment we case study the analysis of motion capture data by identifying characteristic behaviour. Motion data is challenging to due its high number of dimensions, with 93 dimensions, for 31 joints each with  $x, y, z$  axis.

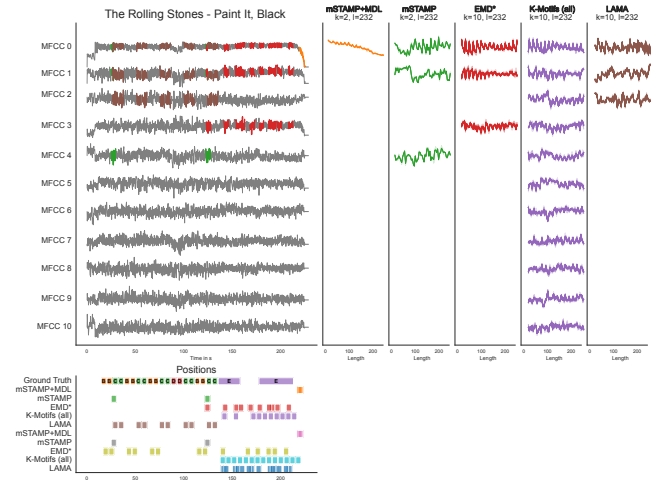
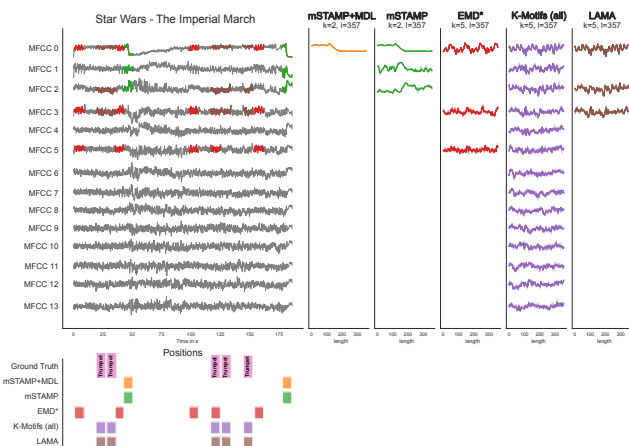


Figure 11: TOP-2 leitmotif discovery. The song *Paint it Black* by the Rolling Stones. The dataset has multiple characteristic sections. We focus on the two largest leitmotifs (in terms of occurrences). One leitmotif has 10 occurrences and is in the C verse. The second one is the humming of Mick Jagger and Keith Richards in the E section with 14 occurrences. For the sake of brevity, we display one leitmotifs by method to the right and the positions of both leitmotifs to the bottom. (Image best viewed in color)

Figure 10 shows a TS of a boxing routine. As part of the routine, the actor performs multiple other motions. The characteristic motion (leitmotif) is a punch/swing with the right hand. The punch is repeated 10 times with a duration of roughly 1.6s each. We only used the joints of the right body half. mSTAMP and LAMA mostly agree in important dimensions for punching, which are located at the right clavicle, humerus, radius, wrist, hand, fingers, and thumb. Other than that EMD\* identifies mostly joints at the right leg, which seems unusual for punching, and represents the triple steps performed prior to a punch. As such, LAMA is the only method to identify all punching motions. mSTAMP, in both variants, identifies two punches. EMD\* fails to identify all punches, correlating to identifying feed movement. K-Motifs similarly fails.

**Pop Music - Paint it Black.** In this experiment, we showcase leitmotifs being discovered in musical scores. Rolling Stones' *Paint it Black*, released in 1966 on the *Aftermath* album, exemplifies the band's rebellious image. The song is notable for its distinctive use of the sitar, an instrument popularized by the Beatles. The sitar, featured prominently throughout the song, adds a unique Middle Eastern flair. The verses of the song breakdown in the verse structure:

- B1 - Verse One: I see a red door ...
- C1 - Verse One: I see the girls walk by ...
- B2 - Verse Two
- C2 - Verse Two
- B3 - Verse Three
- C3 - Verse Three
- D1 - Verse Four: No more will ...



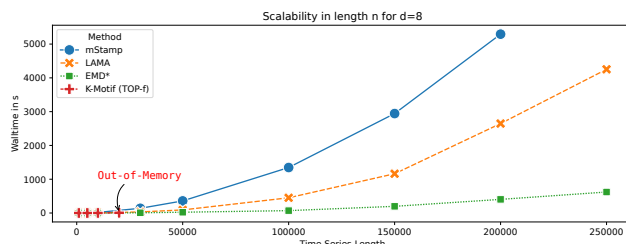
**Figure 12: The Star Wars soundtrack *Imperial March*. The dataset has a characteristic leitmotif played in the presence of Darth Vader. It repeats 5 times with the trumpets playing for roughly 8.3 each. (Image best viewed in color)**

- C4 - Verse Four
- B1 - Verse Five
- C1 - Verse Five
- E - Humming
- F - Outro

Instruments are all played in the same style for B and C verses. Overall, the B verse is repeated 4 times, and the C verse is repeated 5 times. The B verse contains the iconic "I see a red door and ...". B and C verses break down into two repeats with the same rhythmically. For preprocessing, we extracted the first 7 MFCCs from the song, equal to 7 channels with 9744 time stamps. The first leitmotif is equivalent to half of a C verse, which is split at "I see the girls walk by ..." and "I have to turn my head ...". It is 5s long and repeats twice in each, for a total of 10 times. The second one is the humming of Mick Jagger and Keith Richards in the E section with 14 occurrences.

Figure 11 shows the results. Using LAMA, both leitmotifs in sections C and E were identified. The motif set spans the first three MFCC channels. The second motif set found by LAMA overlaps with the humming section, and is located MFCC channels 0, 1 and 5. The B verse was not identified by LAMA. This use case showcases the core concept of LAMA to identify channels by leitmotif. The other competitors use the same channels independent of the leitmotif. mSTAMP, with the number of dimensions as input, identifies two repeats. The B section is found by EMD\*, but blurred by additional humming. The remaining motifs found by the other approaches seem to cluster around the humming section E. mSTAMP+MDL struggles to find a meaningful motif. It identified the fade out of the end in the loudness channel MFCC 0.

**Soundtrack - Imperial March.** Leitmotifs occur frequently in soundtracks to highlight the appearances of characters or landscapes. John Williams frequently uses leitmotifs in the Star Wars soundtrack. One motif is associated with Darth Vader's appearance, while another with the Death Star or the concept of the Force. The song *Imperial March* features a characteristic leitmotif played by



**Figure 13: Scalability in the length  $n$  of the TS.**

trumpets, and played in the presence of Darth Vader. It has a duration of 8.3 and repeats 5 times. Figure 12 shows the results. Using LAMA, we are able to identify the trumpets playing using 3 channels 0, 2 and 3. EMD\* covers only sub-sections of this leitmotif. It covers sections, where the music increases in volume, too. The best competitor is K-Motifs using all dimensions, finding the leitmotif, too. mSTAMP identifies sections of loudness decrease (fade-out).

#### 4.4 Scalability

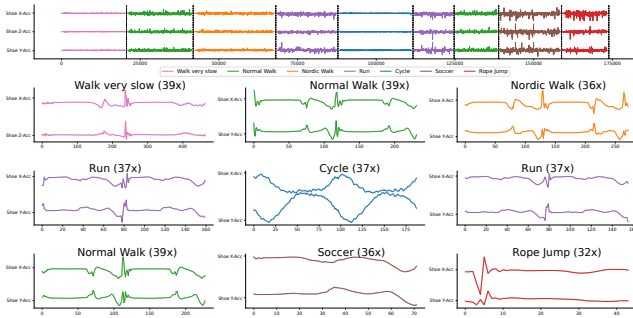
In these experiments, we systematically assessed the impact of increasing TS lengths on the penguin dataset [26], analyzing up to 250k time stamps. The benchmarking was conducted on a laptop equipped with an Apple M1 CPU and 16GB of RAM. Storing a distance matrix of 250k entries in main memory is impractical, requiring approximately 1.8 terabytes of space  $((250k)^2 \times 8 \times 32)$ , while our sparse matrix implementation for LAMA (Section 3.7) necessitates roughly 12 gigabytes of RAM, a significant reduction.

Note that mSTAMP, K-Motif, and LAMA have a worst-case time complexity of  $O(n^2 \times d \times L_{max})$ , pertaining to the computation of pairwise distances among candidates. EMD\* has complexity  $O(n^2 \times L_{max})$  by applying PCA to derive a one dimensional series first, i.e.  $d = 1$ . All implementations are Python-based. For mSTAMP we use the stumpy library, we re-implemented EMD\* and K-Motifs.

Figure 13 shows the results of our experiment. All algorithms scale quadratically in  $n$ . As expected from the complexity analysis, all method exhibit quadratic complexity the TS length, albeit with varying linear factors. LAMA performs faster than mSTAMP, yet is slower than EMD\*, which is due to EMD\*'s aggressive PCA dimensionality reduction technique which, however, also severely impacts its pattern extraction quality. K-Motif did not scale beyond 30k, due to its quadratic memory requirements in length  $n$ .

#### 4.5 Leitmotifs for Summarization

Leitmotif mining on large TS, such as those with over 150k points, can become prohibitive, taking up to 0.5 hours. However, many TS can first be segmented into semantically disjoint sequences, allowing leitmotif discovery to be performed on these segments. Figure 14 (top) shows a TS of a subject performing nine outdoor activities for 29 minutes, spanning 174k data points, sampled at 100Hz, and three channels [10]. If segments are not given, we may apply a segmentation algorithm, such as ClaSS [4], to derive disjoint segments, which takes 70 - 90s. As the data is quasi-periodic, we used the dominant Fourier frequency to determine the leitmotif length in each segment. Figure 14 (bottom) displays the nine leitmotifs



**Figure 14: The 9 Leitmotifs found (bottom) by LAMA in a physical outdoor activity by a subject performing 9 activities (top). First, we performed segmentation, then we ran LAMA on each segment.**

and corresponding two channels found for each segment. Notably, *Normal Walk* and *Run* activities are repeated twice, as indicated by the high similarity of the found leitmotifs. Running LAMA on each of the nine segments took a total of 92s (1900 ops/sec), compared to 739s (235 ops/sec) when running LAMA on the entire TS. This represents an 8-fold speedup. The approach is faster than real-time (29min, or 100Hz). Moreover, it demonstrates that we can summarize a physical activities by their leitmotifs.

## 5 RELATED WORK

Research in multidimensional motif discovery includes algorithms that find motifs across all dimensions and within subsets (sub-dimensional). Reproducibility is challenging due to the lack of datasets with annotated ground truth, partially addressed by synthetic datasets [16]. Fine-tuning parameters often relies on trial and error. We present the first systematic benchmark with annotated ground truth from real use cases. K-Motifs [12] are defined as all subsequences within distance radius  $r \in \mathbb{R}$  to a centre subsequence of length  $l \in \mathbb{N}^+$  of the TS. We use K-Motifs as a baseline. *mSTAMP* [22] extends the matrix profile [23] for sub-dimensional pair motif discovery. To avoid enumerating all subsets of dimensions, the algorithm computes 1-NN distances in each dimension separately, then sorts dimensions in ascending order of 1-NN distances. From all  $k$  separately computed profiles, a  $k$ -dimensional cumulative profile is computed along the dimension axis. The optimal number of dimensions is identified using the minimal description length (MDL). In [6], the variable-length pair motif discovery algorithm *CHIME* is introduced. It transforms subsequences of minimal length  $l$  using SAX [13] along each dimension, converting the TS into words. These words are merged into larger patterns, and sub-dimensional co-occurring words are correlated across dimensions. Pairwise Euclidean distances between candidates sharing the same word are computed and ranked. CHIME produced runtime errors and only finds pairs, so it was not included in the experiments. In [19], the *EMD* method for motif set discovery is introduced. PCA is used to derive a univariate series, and motif sets are found by searching for subsequences within a distance radius  $r$  (similar to K-Motifs [12]). The TS is SAX transformed to prune the search space, and frequent words are used to find candidate subsequences.

Exact Euclidean distance computations within radius  $r$  are then performed. MDL identifies the most relevant patterns. No code is available, so we re-implemented relevant parts to use EMD\* in our experiments. In [15], a sub-dimensional method is introduced. The TS is transformed into SAX words, followed by random projections to identify frequent pair motifs. Euclidean distances of corresponding subsequences are computed along each dimension. Relevant dimensions are selected by sorting distances and applying a threshold. The top motif is extended to a set motif by searching for all subsequences within a threshold, similar to VALMOD [14]. For univariate series, VALMOD proved to be inferior to  $k$ -Motiflets [18] and K-Motifs [12], and we did not further consider it. In [21], an algorithm for discovering non-synchronous motifs in activity data is presented. It uses random projections of SAX words along each dimension and agglomerative clustering on a coincidence graph, connecting motifs with a lag below a hyper-parameter  $\alpha$ . Due to the different setting of asynchronous motifs and the lack of available code, we did not include it. In [1] an approach for sub-dimensional variable-length motif discovery is presented tailored for physiological signal data. The approach uses SAX to derive words, and a grammar induction approach to find frequent variable length motif sets. We have requested the code from the authors, but did not receive a response. As it is tailored to one use case, we did not further consider it. In [16], the salient multi-variate motif (SMM) algorithm is introduced for mining sub-dimensional motifs of variable length using metadata. A salient motif is a  $k$ -frequent subsequence that is distinct from its neighbors and has a pairwise distance of at most  $\epsilon$ . Patterns are identified through sub-sampling and clustering, using metadata to extract inter-dimensional relations. The method relies on metadata and has 7 user parameters.

## 6 CONCLUSION

A Leitmotif in a time series manifests as a dominant and recurrent subsequence, imbued with semantic meaning. In this work, we introduce a novel algorithm called LAMA for the sub-dimensional discovery of Leitmotifs in time series, and present a novel benchmark of 14 annotated datasets for finding Leitmotifs. Other than previous work, LAMA jointly optimizes motif discovery and channel selection thereby finds considerably more meaningful patterns (Leitmotifs) in each of the benchmark sets. Our findings demonstrate that summarizing a time series through its leitmotifs, as found by LAMA, offers a profound understanding of underlying patterns. This spans a spectrum of activities, from the distinctive punching motion in boxing to the fundamental steps of a dance routine, the wing flap pattern of a penguin, motions from human activity, or the unique vocalizations of a bird. In future work, we aim to further explore the scalability of LAMA by approximating the distance matrix through the application of locality-sensitive hashing.

## ACKNOWLEDGMENTS

We would like to express our gratitude to Silvestro Poccia, author of SMM, for providing code and dedicating time to emails and Zoom sessions, and for collaborative debugging. We extend our gratitude to the reviewers for their insightful comments, which significantly enhanced the quality of this paper.

## REFERENCES

- [1] Arvind Balasubramanian, Jun Wang, and Balakrishnan Prabhakaran. 2016. Discovering multidimensional motifs in physiological signals for personalized healthcare. *IEEE journal of selected topics in signal processing* 10, 5 (2016), 832–841.
- [2] Gustavo EAPA Batista, Xiaoyue Wang, and Eamonn J Keogh. 2011. A complexity-invariant distance measure for time series. In *Proceedings of the 2011 SIAM international conference on data mining*. SIAM, Mesa, Arizona, USA, 699–710.
- [3] Matthew Bribitzer-Stull. 2015. *Understanding the leitmotif*. Cambridge University Press, The Triangle Building Shaftesbury Road Cambridge CB2 8EA.
- [4] Arik Ermshaus, Patrick Schäfer, and Ulf Leser. 2024. Raising the ClaS of Stream- ing Time Series Segmentation. *Proceedings of the VLDB Endowment* 17, 8 (2024), 1953–1966.
- [5] The Royal Society for the Protection of Birds (RSPB). 2023. <https://www.rspb.org.uk/birds-and-wildlife>.
- [6] Yifeng Gao and Jessica Lin. 2019. Discovering subdimensional motifs of different lengths in large-scale multivariate time series. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, Beijing, China, 220–229.
- [7] Shima Imani and Eamonn Keogh. 2019. Matrix profile XIX: time series semantic motifs: a new primitive for finding higher-level structure in time series. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, Beijing, China, 329–338.
- [8] Kaveh Kamgar, Shaghayegh Gharghabi, and Eamonn Keogh. 2019. Matrix profile XV: Exploiting time series consensus motifs to find structure in time series sets. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, Beijing, China, 1156–1161.
- [9] Fernando De la Torre Frade, Jessica K. Hodgins, Adam W. Bargteil, Xavier Martin Artal, Justin C. Macey, Alexandre Collado I Castells, and Josep Beltran. 2008. *Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database*. Technical Report CMU-RI-TR-08-22. Carnegie Mellon University, Pittsburgh, PA.
- [10] Ilias Lamprinos, Daniel Steffen, and M Macedo. 2012. Physical activity monitoring for the aging population. In *IADIS International Conference e-Health, located at MCCSIS 2012 IADIS Multi Conference on Computer Science and Information Systems*. IADIS digital library, Lisbon, Portuga, 17–19.
- [11] Leitmotif - Source Code and Raw Results. 2023. <https://github.com/patrickzib/leitmotifs>.
- [12] Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Pranav Patel. 2002. Finding motifs in time series. In *Proc. of the 2nd Workshop on Temporal Data Mining*. ACM, Edmonton, Alberta, Canada, 53–68.
- [13] Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. 2007. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and knowledge discovery* 15, 2 (2007), 107–144.
- [14] Michele Linardi, Yan Zhu, Themis Palpanas, and Eamonn Keogh. 2018. VALMOD: A suite for easy and exact detection of variable length motifs in data series. In *Proceedings of the 2018 International Conference on Management of Data*. ACM, Houston TX USA, 1757–1760.
- [15] David Minnen, Charles Isbell, Irfan Essa, and Thad Starner. 2007. Detecting sub-dimensional motifs: An efficient algorithm for generalized multivariate pattern discovery. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*. IEEE, Omaha, NE, USA, 601–606.
- [16] Silvestro R Poccia, K Selçuk Candan, and Maria Luisa Sapino. 2021. SMM: Leveraging metadata for contextually salient multi-variate motif discovery. *Applied Sciences* 11, 22 (2021), 10873.
- [17] Vincent Rone. 2018. Scoring the Familiar and Unfamiliar in Howard Shore’s The Lord of the Rings. *Music and the Moving Image* 11, 2 (2018), 37–66.
- [18] Patrick Schäfer and Ulf Leser. 2022. Motiflets: Simple and Accurate Detection of Motifs in Time Series. *Proceedings of the VLDB Endowment* 16, 4 (2022), 725–737.
- [19] Yoshiki Tanaka, Kazuhisa Iwamoto, and Kuniaki Uehara. 2005. Discovery of time-series motif from multi-dimensional data based on MDL principle. *Machine Learning* 58 (2005), 269–300.
- [20] Sahar Torkamani and Volker Lohweg. 2017. Survey on time series motif discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 7, 2 (2017), e1199.
- [21] Alireza Vahdatpour, Navid Amini, and Majid Sarrafzadeh. 2009. Toward unsupervised activity discovery using multi dimensional motif detection in time series. In *Twenty-First International Joint Conference on Artificial Intelligence*. Citeseer, PASADENA, CALIFORNIA, USA, 272–277.
- [22] Chin-Chia Michael Yeh, Nickolas Kavantzias, and Eamonn Keogh. 2017. Matrix profile VI: Meaningful multidimensional motif discovery. In *2017 IEEE international conference on data mining (ICDM)*. IEEE, New Orleans, LA, USA, 565–574.
- [23] Chin-Chia Michael Yeh, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen, and Eamonn Keogh. 2016. Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th international conference on data mining (ICDM)*. IEEE, Barcelona, Spain, 1317–1322.
- [24] Aras Yurtman and Billur Barshan. 2014. Automated evaluation of physical therapy exercises using multi-template dynamic time warping on wearable sensor signals. *Computer Methods and Programs in Biomedicine* 117, 2 (2014), 189–207.
- [25] Sheng Zhong and Abdullah Mueen. 2024. MASS: distance profile of a query over a time series. *Data Mining and Knowledge Discovery* 38, 3 (2024), 1466–1492.
- [26] Yan Zhu, Makoto Imamura, Daniel Nikovski, and Eamonn Keogh. 2017. Matrix profile VII: Time series chains: A new primitive for time series data mining. In *2017 IEEE international conference on data mining (ICDM)*. IEEE, New Orleans, LA, USA, 695–704.