Published in final edited form as:

IEEE J Sel Top Signal Process. 2016 August ; 10(5): 832–841. doi:10.1109/JSTSP.2016.2543679.

Discovering Multidimensional Motifs in Physiological Signals for Personalized Healthcare

Arvind Balasubramanian,

Department of Computer Science, The University of Texas at Dallas, Richardson, TX, 75080 USA

Jun Wang, and

Department of Bioengineering, The University of Texas at Dallas, Richardson, TX, 75080 USA

Balakrishnan Prabhakaran

Department of Computer Science, The University of Texas at Dallas, Richardson, TX, 75080 USA

Abstract

Personalized diagnosis and therapy requires monitoring patient activity using various body sensors. Sensor data generated during personalized exercises or tasks may be too specific or inadequate to be evaluated using supervised methods such as classification. We propose multidimensional motif (MDM) discovery as a means for patient activity monitoring, since such motifs can capture repeating patterns across multiple dimensions of the data, and can serve as conformance indicators. Previous studies pertaining to mining MDMs have proposed approaches that lack the capability of concurrently processing multiple dimensions, thus limiting their utility in online scenarios. In this paper, we propose an efficient real-time approach to MDM discovery in body sensor generated time series data for monitoring performance of patients during therapy. We present two alternative models for MDMs based on motif co-occurrences and temporal ordering among motifs across multiple dimensions, with detailed formulation of the concepts proposed. The proposed method uses an efficient hashing based record to enable speedy update and retrieval of motif sets, and identification of MDMs. Performance evaluation using synthetic and real body sensor data in unsupervised motif discovery tasks shows that the approach is effective for (a) concurrent processing of multidimensional time series information suitable for real-time applications, (b) finding unknown naturally occurring patterns with minimal delay, and (c) tracking similarities among repetitions, possibly during therapy sessions.

Index Terms

Data Mining; Pattern Discovery; Time Series; Multidimensional Motifs; Personalized Healthcare; Wearable Technology

I. Introduction

Wearable body sensors along with data analytics have been widely used for providing continued monitoring of patient performance during rehabilitative therapies and adding value to personalized healthcare. Based on sensor data analytics, physicians and medical care providers can verify compliance, evaluate performance and provide personalized assistance and therapy to the patient for optimal efficacy. However, there are some

challenges associated with the monitoring of patient performance using body sensor data –

- (i) The captured sensor data might be too patient specific and inadequate for the differentiation of normal and abnormal behavior to be modeled as a classification problem.
- (ii) Over the treatment period, the physician may need to modify the diagnostic paradigm to suit the evolving characteristics in patient performance. Thus, a personalized and context aware approach is required that (a) analyzes patient activity in an unsupervised or semi-supervised manner for identifying patterns; and (b) is scalable across multiple heterogeneous sensors capturing different data (e.g., ECG, blood pressure, motion, etc.). (see Figure 1).

Motif Discovery can be used in sensor generated multidimensional time series data for monitoring and tracking structural similarities in patient data. Motifs are approximately repeated subsequences or patterns in any time series data. Multidimensional motifs (MDMs) extend the concept to co-occurrences of such motifs across multiple dimensions. MDMs can facilitate personalization and context awareness, since they can identify the patterns generated in a patient data, without the need for classification. For instance, a physician can simply mark anomalous patterns and request to be alerted when such a pattern (typically patient specific) recurs in real-time.

Previous studies on multidimensional patterns [11], [12], [15], [22] either involve dimension wise processing of the time series, or prior knowledge of possible patterns, thus requiring all the data to be available on disk. This would not be helpful in discovering unknown multidimensional patterns in an online or real-time manner. In this paper, we present a MDM discovery approach that is compatible with real-time processing of time series streams. The contributions of this approach are:

- a. unsupervised identification of variable length multidimensional motifs, with support for two alternative models for MDMs – unordered and temporally ordered;
- b. concurrent online processing of all dimensions in the multidimensional time series for mining MDMs in real time, using efficient hashing based records to enable speedy update and retrieval of motif sets;

Motivation for two alternative models for MDMs

In synchronous biomedical sensor data, such as motion activity, co-occurrence information is useful in associating patterns from different dimensions of the sensor stream based on temporal proximity, which is addressed by our unordered MDM model. However, some specific studies involving heterogeneous sensor information might require the exact temporal sequence in which the co-occurring motifs are ordered. For example, muscle activity data from Electromyogram (EMG) sensors, captured concurrently with 3D motion activity data, can exhibit useful temporal relations. For example, in a jump activity, the onset of EMG signal (contraction of *tibialis anterior* in this example) occurs prior to onset of the toe movement as the subject prepares to jump [20] (See Figure 2). Thus, a recurring temporal ordering of motifs can be an indicator of an MDM as well, in contrast with simple co-occurrence of motifs. Similarities between MDMs having different constituent motifs can be observed based on the temporal relations between the constituent motifs.

II. Related Work

In recent years, there have been many studies [2]–[4], [8], [13] (for more references, please refer to [5]) that address the problem of offline as well as online motif discovery in unidimensional data, despite factors such as noise, scaling, translation etc. A number of recent studies have proposed solutions for MDM discovery. Tanaka et al [21] reduced multidimensional time series data to a single dimension using Principal Component Analysis (PCA), followed by motif discovery based on minimum description length principle. McGovern et al [10] suggested finding the salient dimensions and then identifying the key temporal motifs in those dimensions, with applications to severe weather prediction. Mueen et al [13] presented an online approach for unidimensional motif discovery and maintenance, and discussed scaling up to multiple dimensions. Chiu et al [4] addressed the issue of motif distortion in the presence of noise, and proposed a probabilistic approach based on random projection, to find similar SAX words [9] even when they differed by one or more symbols. Minnen et al [11] and Vahadtpour et al [22] extended this approach for MDMs, combining collision matrices resulting from random projection and partial similarity matching for each dimension, by identifying relevant dimensions and by graph clustering, respectively. Ordóñez et al. [15] proposed an approach for physiological data classification based on vector-space model for words and documents, presenting a Multivariate Bag of Patterns and Stacked Bag of Patterns representations of multivariate time series. More recently, Moazeni et al. [12] proposed a query based multidimensional signal search, with provision for missing or switched dimensions, noise, scaling and asynchrony.

However, the approaches presented in [11], [15], [22] would not identify motifs of variable length beyond the SAX word segmentation range. Also, a common limitation among these approaches, including the one presented in [12], is that they either involve processing the entire information in one dimension before processing the next dimension, or require prior knowledge of the entire set of unidimensional patterns that can occur in the multidimensional data, thus implicitly requiring all the data to be available on disk. Our approach achieves real-time tagging of motif co-occurrences, along with discovering evolving pattern groups, by using a concurrent synchronous processing methodology. In a previous work [3], we extended the unidimensional variable length motif discovery technique proposed by Li et al [8] to enhance robustness and noise tolerance in motif discovery. In this paper, we present a parallelized motif co-occurrence record to mine MDMs synchronously across multiple dimensions.

With regard to mining the temporal relations between patterns in multivariate data, Allen's interval algebra [1] is particularly useful in modeling sequence information. This has been used in interval based event mining [17], frequent pattern mining from multiple data domains [18] for classification and similarity evaluation of event sequences [7]. A technique for mining temporal ordering of motifs in time series is presented in [16], but the approach is restricted to lag patterns and does not allow any lag variation within instances of the same lag pattern. Our approach uses the interval relations to formulate an alternative definition of an MDM based on the temporal ordering among unidimensional motifs.

The multidimensional motif discovery problem in time series data may bear some similarity to the popular biclustering problem in bioinformatics for identifying coherent activity in gene expression data [26]–[29]. However, their objectives are fundamentally different. Biclustering seeks to identify occurrences of the same expression occurring in different rows of the gene expression data. In contrast, the objective of our approach is to identify co-occurrences of different recurring unidimensional patterns across multiple dimensions. Thus, biclustering is a far more constrained problem than the problem addressed in this paper. Also, in terms of its operation, biclustering does not provide for partial differences in expressions when calculating similarity.

III. Preliminaries

A. Background

The following definitions of time series and time series motif pertain to a unidimensional/univariate time series.

Definition: A **time series** $T = (o_1, o_2, o_3, \dots, o_n)$ is a temporally ordered sequence of n real value observations.

For sake of simplicity of formulation, we assume the sampling rate to be uniform. Varying sampling rates can always be managed by processing the time series in terms of time durations rather than number of samples.

Definition: A **time series subsequence** $subT$ of a time series T is a temporally ordered subset of consecutive real value observations $(o_i, o_{i+1}, o_{i+2}, \dots, o_j)$ from $T(1 \leq i \leq j \leq n)$, where n is the length of T . The length of the subsequence is $j - i + 1 \leq n$.

Definition: The **begin timestamp** $beg_ts(subT)$ and **end timestamp** $end_ts(subT)$ of a time series subsequence $subT = (o_i, o_{i+1}, o_{i+2}, \dots, o_j)$ are the time instants t_i and t_j at which the first observation o_i and last observation o_j in $subT$ occur, respectively.

In our work, for sake of simplicity, the term timestamp ts when used by itself denotes the begin timestamp beg_ts .

Definition: A **unidimensional time series motif** M is a set $\{subT_i\}$ of two or more time series subsequences that are evaluated to be approximately similar to each other based on a given distance measure/metric.

Thus, $dist(subT_a, subT_b) \leq D, \forall \{subT_a, subT_b\} \in M$, where $dist$ is a distance measure/metric, and D is a threshold value set on $dist$. Also, subsequences belonging to a motif should be non overlapping, i.e. $\forall \{subT_a, subT_b\} \in M, (end_ts(subT_a) < beg_ts(subT_b)) \vee (end_ts(subT_b) < beg_ts(subT_a))$.

Definition: An **instance** m of a unidimensional motif M is any $subT \in M$. Thus, M can be considered to be a distinct motif type, and the motif instance m is one of the occurrences that belong to the motif type indicated by M .

In Figure 3, M_{11} , M_{21} , M_{22} , M_{31} , M_{41} are all time series motifs having two instances each, designated by m_{11} , m_{21} , m_{22} , m_{31} , m_{41} .

B. Formulation of Multidimensional Motifs

The following definitions pertain to multidimensional time series.

Definition: A **multidimensional time series** $MDT = \{T_1, T_2, T_3, \dots, T_d\}$ is a set of d synchronous time series. T_i is the i th time series in an MDT , and is therefore referred to as the i th dimension of the MDT .

Note 1: Implications of processing such synchronous time series information – (i) The information from all dimensions of the time series is available simultaneously; and (ii) the processing of MDT is done using a temporal window that is equal across all dimensions of MDT , thus normalizing any variation of sampling rate between dimensions.

Definition: An instance m_a of a motif M_a and an instance m_b of a motif M_b occurring in the same or different dimensions of an MDT are said to be **neighbors**, if $|ts(m_a) - ts(m_b)| \leq r$, where $ts(m_a)$, $ts(m_b)$ are the timestamps of m_a and m_b , respectively, and r is a prespecified integer value called the **neighbor distance**. Motif instances that are neighbors, are considered to be *co-occurring*.

Definition: A **motif neighborhood** is a consecutive set of time instants $(t_p, t_{p+1}, t_{p+2}, \dots, t_{p+r})$, where r is the prespecified neighbor distance, so that, any two motif instances m_a and m_b would be neighbors if $t_p \leq ts(m_a) < t_{p+r}$ and $t_p \leq ts(m_b) < t_{p+r}$.

Thus, any motif instances occurring within the same motif neighborhood are neighbors.

In Figure 3, if the neighborhood distance be 4, then the instance of M_{21} at 22 is a neighbor of the instance of M_{31} at 26, but not of the instance of M_{22} at 29.

Definition: A **MotifBag** $MBag = \{M_1, M_2, \dots, M_k\}$ ($k \geq 2$) is an unordered set of motifs such that, there exists at least one set *bag* of motif instances $\{m_1, m_2, \dots, m_k\}$ ($m_i \in M_i, i = 1, 2, \dots, k$) where, $\forall m_a, m_b \in bag(1 \leq a, b \leq k)$, m_a and m_b are neighbors.

The set of motif instances *bag* is referred to as an **instance** of the MotifBag $MBag$. The motifs in a MotifBag are referred to as its **members** and can span two or more dimensions.

For example, if $m_x \in M_x$, $m_y \in M_y$ and $m_z \in M_z$ are motif instances that are neighbors, then the MotifBags formed are $\{M_x, M_y\}$, $\{M_y, M_z\}$, $\{M_x, M_z\}$ and $\{M_x, M_y, M_z\}$ with instances $\{m_x, m_y\}$, $\{m_y, m_z\}$, $\{m_x, m_z\}$ and $\{m_x, m_y, m_z\}$, respectively. In Figure 3, examples of MotifBags are $\{M_{11}, M_{21}, M_{31}\}$ and its subsets, $\{M_{22}, M_{41}\}$ etc.

Definition: If $m \in bag$, then $index(bag, m) = (dim, seq)$ where *dim* is the dimension of the MDT in which m occurs, and *seq* is the sequence number of m in its dimension within the range of *bag*. For example, in Figure 3, for the first instance of the MotifBag formed by $\{M_{22}, M_{41}\}$, $index(m_{22}) = (2, 1)$ and $index(m_{41}) = (4, 1)$.

Definition: The **begin timestamp** $beg_ts(mbag)$ and **end timestamp** $end_ts(bag)$ of an instance bag of a MotifBag $MBag$ are defined as $\{min(\{ts(m_x)\}), m_x \in bag\}$ and $\{max(\{ts(m_x)\}), m_x \in bag\}$, respectively.

Thus, the begin timestamp and end timestamp of bag are the timestamps of the motif instance belonging to bag that occur the earliest and latest, respectively.

Definition: An **Unordered Multidimensional Motif UMDM** is a MotifBag having two or more non-overlapping instances. Thus, for any two instances $umdm_a$ and $umdm_b$ of $UMDM$, $(end_ts(umdm_a) < beg_ts(umdm_b)) \vee (end_ts(umdm_b) < beg_ts(umdm_a))$.

Figure 3 uses our previous example to illustrate the concept of a UMDM. The motif instances of M_{11} , M_{21} and M_{31} together constitute a UMDM, since they form a MotifBag that has two instances. The same is true for M_{22} and M_{41} .

Definition: A **temporal relation** $R(m_a, m_b)$ between two motif instances $m_a \in M_a$ and $m_b \in M_b$ is the representation of the temporal occurrence of m_a with respect to that of m_b .

As dictated by Allen's Interval Algebra [1], the various definitions of $R(m_a, m_b)$ are listed in Table I.

Definition: A **MotifOrdering** is a set of pairwise temporal relations among instances of motifs belonging to a MotifBag. Thus, a MotifOrdering $MOrd$ is a set of temporal relations $\{R(m_a, m_b)\}$, $\forall m_a \in M_a, m_b \in M_b$ where $M_a, M_b \in bag$ and bag is a MotifBag instance, and $R_i \in \{R_1, R_2, \dots, R_7\}$ (see Table I).

An **instance** of a MotifOrdering is a single occurrence of a MotifOrdering in an MDT. Two instances of a MotifOrdering have the same number of participating motifs and constituent temporal relations with a dimension-wise one-to-one mapping among the temporal relations, but the need not have the same member motifs.

$ord_1 = \{R(m_a, m_b)\}$ and $ord_2 = \{R'(m_c, m_d)\}$ are instances of the same MotifOrdering $MOrd$ if and only if, $\forall R(m_a, m_b) \in ord_1, \exists R'(m_c, m_d)$ such that $R = R'$, $index(bag_1, m_a) = (bag_2, m_c)$ and $index(bag_1, m_b) = (bag_2, m_d)$, where bag_1 and bag_2 are MotifBag instances with $m_a, m_b \in bag_1$ and $m_c, m_d \in bag_2$.

A MotifOrdering can have multiple instances throughout an MDT.

In Figure 3, M_{11} (at 2), M_{21} (at 3) and M_{31} (at 2) are neighbors with temporal relations – $Overlaps(M_{11}, M_{21})$, $Equals(M_{11}, M_{31})$, $Overlaps(M_{31}, M_{21})$. Again, M_{11} (at 2), M_{21} (at 3) and M_{31} (at 2) are neighbors with temporal relations – $Overlaps(M_{11}, M_{22})$, $Equals(M_{11}, M_{31})$, $Overlaps(M_{31}, M_{22})$. These two form instances of the same MotifOrdering even though the member motifs are not the same, since both sets are governed by the same temporal relations.

Definition: An **Temporal Ordering based Multidimensional Motif TMDM** is a MotifOrdering having two or more non-overlapping instances. Thus, for any two instances

$tmdm_a$ and $tmdm_b$ of $TMDM$, $(end_ts(tmdm_a) < beg_ts(tmdm_b)) \vee (end_ts(tmdm_b) < beg_ts(tmdm_a))$.

In Figure 3, the MotifOrdering with instances M_{11} , M_{21} , M_{31} and M_{11} , M_{22} , M_{31} is a TMDM since it has two instances.

Note 2: Two instances of the same MotifOrdering can be instances of two different MotifBags. Similarly, two instances of the same MotifBag can be instances of two different MotifOrderings. **Note 3:** By definition, two instances of the same UMDM should have the exact same member motifs, irrespective of the temporal ordering among the motifs in the instance. In contrast, two instances of the same TMDM should have the exact same temporal ordering, while the actual member motifs can be different.

C. Problem Statement

Given a d -dimensional time series MDT , identify and retrieve every distinct Multidimensional Motif $UMDM$ or $TMDM$ and its instances.

IV. Multidimensional Motif Discovery

In this section, we provide a detailed discussion of the proposed Multidimensional Motif (MDM) discovery method.

A. Unidimensional Motif Discovery

For motif discovery in individual dimensions of the multidimensional time series, any state-of-the-art motif discovery technique could be used. However, one of the objectives of the proposed approach is to facilitate concurrent processing of all the dimensions, such that already processed sections need not be revisited. To enable this, we adapt the unidimensional motif discovery approach presented in [8] with the optimization suggested by [3] to enhance robustness and noise tolerance in motif discovery. The salient features of the approach are:

(i) Discretization of individual dimensions using SAX—For efficient step-wise processing, the sensor data streams are transformed into a string of discrete symbols using the widely popular Symbolic Aggregate approXimation (SAX) representation [9] (see Figure 4(a)) (The details of SAX are not elaborated here for sake of brevity; please refer to [9] and [3]). In an online streaming scenario, the pre-processing and discretization steps can be executed as a transformation filter that converts the streaming data to its symbolic representation in each dimension in real-time.

(ii) Unidimensional Motif Discovery using Sequitur—The string encoded time series needs to be processed for finding naturally occurring repetitive patterns of variable lengths. For this, we employ a grammar induction approach based on the Sequitur string compression algorithm [14] to find variable length motifs, as suggested in previous studies [3], [8], as well as identify hierarchies among the motifs found (see Figure 4(b)).

(iii) Incorporating Flexibility for Spatial Variance—Valid instances of a motif might often be overlooked because of local variations due to noise and distortion. We have

proposed an enhancement to Sequitur in a previous work [3] to enable noise tolerant motif discovery, and have incorporated the same here.

B. Concurrent Processing of Multidimensional Time Series Data

The enhanced rule mining approach derived from Sequitur processes each of the strings in parallel, one symbol (SAX word) at a time. This can be visualized as a window that moves across all dimensions one symbol at a time, in what can be referred to as *symbol steps*. The timestamp of a symbol is the index of the symbol step in which it occurs. At each symbol step, the window encounters the next symbol in each dimension. Whenever motif instances are encountered and identified in any of the dimensions, they are indexed, as described in the following section.

C. Unordered Multidimensional Motif (UMDM) Discovery

The first motif discovery approach presented here provides the functionality of concurrently processing all dimensions of the multidimensional time series data, indexing unidimensional motifs and their co-occurrences as and when they are encountered, thus yielding UMDMs.

1) Identifying Co-occurrences of Unidimensional Motif Instance—Motif instances from different dimensions can be grouped together based on their timestamps as *co-occurring* motif instances. We execute concurrent motif discovery on all dimensions of the multidimensional time series (see Section V.A), and use an inverted timestamp index, which maintains each timestamp mapped to the list of unidimensional motifs that have an instance occurring at that particular timestamp, i.e. $\{ts \rightarrow \{M_{ij}\}\}$ (see Figure 5). The grouping of motif instances as neighbors is based on a predefined temporal proximity, referred to as the *neighborhood distance*, and can be achieved using a sliding window having a width equal to the neighborhood distance on the timestamp index keys. Figure 5 also shows the different sets of neighbors obtained from the example time series using a neighborhood distance of 4.

2) Indexing Co-occurring Motif Groups—MDMs are quite simply recurring groups of co-occurring unidimensional motif instances spanning multiple dimensions. Once the neighbor sets belonging to a particular duration of time have been identified, we need to record each combination of such neighboring motif instances, referred to as MotifBags, as a potential MDM. Each MotifBag would thus be a distinct combination of two or more motif instances from two or more dimensions. On the entry of each new motif instance in the timestamp index, every motif neighborhood that the newly identified motif instance belongs to is searched. From each such neighborhood, sets of all possible combinations of neighboring motif instances are computed and designated as MotifBags (see Algorithm 1).

Since recurring MotifBag instances are potential UMDMs, we need to track the occurrences of previously encountered MotifBags while simultaneously recording new MotifBags. For this purpose, we use a frequency index MBIndex to keep track of each distinct MotifBag and its frequency (number of instances) (see Figure 6(a)). Each distinctly identified MotifBag is recorded in the index, and mapped to a list of timestamps corresponding to the MotifBag's

instances, updated over time. Thus, the MBIndex is of the form $\{M Bag \rightarrow \{ts(bag_1), ts(bag_2), \dots\}\}$, where $MBag$ is a uniquely identified MotifBag with instances bag_1, bag_2 etc.

In the interest of efficiency, the frequency index MBIndex is implemented using a hashtable, where the MotifBags serves as keys, whereas the list of each MotifBag's instances along with timestamp information is stored as the value corresponding to the appropriate MotifBag key. The implementation of the MBIndex has the following features:

- i. Using a hashtable enables constant time addition and updation of MotifBag entries
- ii. MBIndex records co-occurrence associations in an incremental manner such that any subset of the cooccurrence is also recognized. Even after the creation of a motifBag $\{M_{ix}, M_{jy}, M_{kz}\}$, the motifBags $\{M_{ix}, M_{jy}\}, \{M_{jy}, M_{kz}\}$ and $\{M_{ix}, M_{kz}\}$ are not discarded and continue to be tracked. This provision makes the approach more robust in identifying the individual importance of each lower order MDM and their contributions to any higher order MDM they might constitute.
- iii. Since the MBIndex is updated on-demand, the number of updates is kept to a minimum.
- iv. MBIndex can be queried on in real time to retrieve the status of a specific MotifBag at any instant of time, such as its frequency, number of members, timestamps of instances etc.

Reducing Time and Space complexity: Enumeration of the exhaustive set of neighbor combinations involving a particular motif instance can take $O(k * 2^k)$ time in the worst case scenario, where k is the number of neighbors with respect to the motif instance. However, since the neighbors are common between motifs associated with the same timestamp, a simple implementational optimization of this step can reduce this time cost. Instead of executing on the entry of every new motif instance, the procedure can wait till unidimensional motif discovery processing is completed for that timestamp, and all currently discovered motif instances having this timestamp are indexed. The space cost of the algorithm is expensive, but required to be able to discover unknown combinations of unidimensional patterns. However, pattern combinations that are not tagged as MDMs beyond a certain time duration can be pruned, thus regulating the overall space cost.

3) Identifying UMDMs—Conceptually, an Unordered Multidimensional Motif is a recurring group of co-occurring motif instances. As explained in Section III, this concept translates to a UMDM being a MotifBag that has two or more instances. Thus, whenever the frequency of a MotifBag as recorded in the MBIndex exceeds the desired threshold, the algorithm flags the MotifBag as UMDM, with the instances of the MotifBag becoming instances of the UMDM.

Algorithm 1. GenerateMBMO()

Input: Timestamp ts , Neighborhood distance r

Output: set MB of MotifBags, set MO of MotifOrderings

```

 $MB \leftarrow \{\}, MO \leftarrow \{\}$ 
for valid timestamps  $p \leftarrow ts$  to  $ts + r$  do
     $currentNeighborhood \leftarrow p - r$  to  $p$ 
     $currentNeighborSet \leftarrow \{\}$ 
    for all timestamp  $s$  in  $currentNeighborhood$  do
         $motifSet \leftarrow tsIndex.getMotifsAt(s)$ 
         $currentNeighborSet \leftarrow currentNeighborSet \cup motifSet$ 
    end for
     $pwSet \leftarrow generatePowerSet(currentNeighborSet)$ 
     $pwSet \leftarrow pwSet - nullset\{\emptyset\}$ 
    for all set  $bag$  in  $pwSet$  do
        if  $bag$  is clean (No overlapping motifs) then
             $MB \leftarrow MB \cup bag$ 
             $mo \leftarrow \{\}$ 
            for all pair of motif instances  $m_a, m_b$  in  $bag$  do
                Form  $tempOrd$  tuple using relation  $R$  between  $m_a$  and  $m_b$ 
                 $mo \leftarrow mo \cup tempOrd$ 
            end for
             $MO \leftarrow MO \cup mo$ 
        end if
    end for
end for

```

D. Temporally ordered Multidimensional Motif (TMDM) Discovery

We propose an alternative model for representing MDMs based on tracking recurring groups of temporal relations among co-occurring motif instances, to yield TMDMs.

1) Representing temporal relations among motifs—We propose a *tempOrd* tuple format to store pairwise temporal relations between motif instances. Each *tempOrd* tuple contains three pieces of information $[a - p, b - q, R]$ – where $a - p$ and $b - q$ designate the participating motif instances and R is the temporal relation as given by Allen's interval algebra [1]. The motif indicated by $a - p$ is the p th motif in the a th dimension while the motif indicated by $b - q$ is the q th motif in the b th dimension, where p and q are orderings with respect to the particular combination under consideration (see Figure 6(b)). There are two main advantages of this representation format – (i) Only the local ordering and temporal relations are retained, and not the actual motif identifiers. This generic temporal ordering information can be compared with any other combination of motifs for similarity in temporal ordering; (ii) Use of Allen's interval algebra reduces redundancy by not requiring the inverse temporal relations to be recorded. For instance, if $[a - p, b - q, R]$ is a tuple formed, then Allen's formulation eliminates the need for the inverse tuple $[b - q, a - p, R']$ to be stored (if R is *Before*, then R' would be *After*).

2) Indexing Groups of Temporal relations—From each motif neighborhood, sets of all possible combinations of neighboring motif instances are computed. For each of these combinations the pairwise temporal relations are captured and encoded into tempOrd tuples. Thus, each combination of neighbor motifs yields a set of tempOrd tuples, which we refer to as a *MotifOrdering* (see Algorithm 1). **Note:** MotifOrderings can be formed by extracting temporal relations among motif instances that are members of a MotifBag (see Section IV.C), thus building upon existing functionality and making our method more efficient.

The process of indexing MotifOrderings is the exact same as that in case of MotifBags. We simply use a different frequency index, called the *MOIndex*, to track and record each distinct MotifOrdering and its frequency (number of instances) (see Figure 6(b)). Thus, the MOIndex is of the form $\{MOrd \rightarrow \{ts(ord_1), ts(ord_2), \dots\}\}$, where *MOrd* is a uniquely identified MotifOrdering with instances *ord₁*, *ord₂* etc. As in case of MBIndex, the MOIndex is also implemented using a hashtable to benefit from the advantages as described before.

3) Identification of TMDMs—By definition, a TMDM is a MotifOrdering that has two or more instances. Thus, whenever the frequency of a MotifOrdering as recorded in the MOIndex exceeds the desired threshold which is two, the algorithm flags the MotifOrdering as a TMDM, with the instances of the motifBag becoming instances of the TMDM.

Constrained TMDMs require constituent motifs as well as temporal relations to be identical. This can be achieved by cross-referencing entries in MBIndex and MOIndex and tracking frequencies of such joint entries.

V. Experiments and Results

In order to illustrate and evaluate the performance of the MDM discovery techniques presented in the previous sections, we present a set of experiments on both synthetic and real data.

A. Synthetic Data (Planted Motifs)

We evaluated the performance of the proposed method in identifying artificially introduced patterns. We generated a synthetic multidimensional time series with random walk data, and “planted” a number of motifs in every dimension. The resulting synthetic dataset provides examples for MDMs. Each planted motif has instances that are approximately similar to each other. We then executed unguided MDM discovery on this dataset using the proposed approach. A sliding window of 50 samples was used to scan the individual dimensions concurrently. Figure 7 illustrates a portion of the synthetic dataset used along with the planted motifs, and also provides some examples of the results obtained from motif discovery. As can be seen from the figure, the approach is able to identify every unique MDM resulting from the co-occurrences of the planted motifs, with all of their instances. The average recall of the method on this dataset is almost 100%. However, the unguided method also retrieves a number of MDMs using the interval data between planted motifs, which are not relevant and can be discarded. Precision therefore is not a suitable performance measure for unsupervised motif discovery, unlike classification. Some noteworthy points are: (i) discovered MDMs span different number of dimensions, and are

also hierarchically recorded. For instance, even though the MDM in Figure 7(c) is a formed by a subset of the motifs forming the MDM in Figure 7(a), maintaining records for hierarchical MDMs proves beneficial in finding instances that are not common to both MDMs; (ii) identified MDMs well exceed the length of the basis sliding window, while individual unidimensional motifs forming an MDM need not be of the same length either; (iii) given an intuitive idea of the expected patterns, the approach can use a suitable basis sliding window to track and “grow” patterns of lengths longer than the sliding window length.

Figure 8(a) presents an example of unconstrained TMDMs identified from the synthetic data. In the given example, the TMDM has three instances spanning two dimensions. We can see that two instances have common member motifs while the third has different member motifs. However, in each of these instances, the motif in the first dimension has a “During” relationship with the motif in the second dimension. In contrast, under the constrained TMDM, the instance with different member motifs is not indexed, as shown in Figure 8(b).

B. Comparative Evaluation

As mentioned previously, it is challenging to compare the performance of our approach with previous methods, since they are inherently offline and do not support sequential discovery of MDMs. Our approach follows a more greedy approach, identifying and tagging unidimensional as well as multi dimensional motifs working by a sequential first-come-first-serve best-effort principle. However, for comparative evaluation, we implemented the Graph Clustering method presented in [22]. This method could return almost all MDMs from the planted motifs dataset. However, it involves manual intervention after the unidimensional motif discovery to discard irrelevant motifs, and is not conducive to the end-to-end automated processing. The approach also does not process the dimensions concurrently, unlike our approach which does not suffer from this limitation.

C. Real Data (Biomedical and Physiological Body Sensor Data)

To establish the utility of the proposed method in identifying patterns in real-world dataset, we evaluated its performance on two different physiological signal datasets.

Electromagnetic Articulography—We tested the approach on an Electromagnetic Articulography (EMA) dataset collected for research in silent speech interfaces [24]. The dataset used in this experiment is the articulatory data (tongue motion) for sequences of 25 words repeated by a single speaker, recorded using the Electromagnetic Articulograph (Carstens Inc., Germany) [25]. Four sensors – T1 (Tongue Tip), T2 (Tongue Blade), T3 (Tongue Body Front) and T4 (Tongue Body Back) - were attached approximately 10 mm from each other at the midline of the tongue [23], and were tracked within a calibrated electromagnetic field to record the tongue movement. The spatial accuracy of AG500 is 0.5 mm [25]. Only the y and z dimensions of the tongue sensors were used for analysis because the lateral movement (along the x dimension) of the tongue is not significant in normal speech production [23]. Figure 9 illustrates a few examples of the MDMs identified by the proposed approach in the motion trajectory of the tongue for the same sequence of 25 words

repeated twice. The approach was able to identify MDMs that which captured repetition of the same sequence of words. Any spurious motif instances found not corresponding to the correct sequence enable identification of similarities in trajectories for different word sequences. In case of silent speech interfaces, a high level of accuracy is required in classifying articulatory data, in order to differentiate between words pronounced. While MDMs may not be relied upon for classification purposes, they can serve as indicators of sets of words that have similar motion trajectories. The identification of such sets can assist in fine tuning a classifier, or making the differentiation stricter for those mix of words.

Motion Capture and Muscle Activity—In order to demonstrate the mining of TMDMs in real datasets, we use Mocap-EMG data collected for a “Raise Arm” activity [19]. The example uses the sensor data generated from motion capture sensors on the wrist joint (3 dimensions) captured synchronously with corresponding muscle activity data for the *tibialis anterior* muscle (1 dimension), captured at a sampling rate of 120 Hz. Examples of TMDMs from a section of the Mocap-EMG data for 10 repetitions of the activity using a sliding window and neighborhood distance of 500 samples each are shown in Figure 10. Figure 10(a) shows a TMDM, where the onset times of muscle activity shown by EMG data and the wrist joint displacement are the same. On the other hand, Figure 10(b) shows a TMDM, where the muscle activity shown by EMG data has an onset time before the onset time of the wrist joint displacement. Please note that in both cases, the constituent motifs are not identical across instances, while the temporal relations among the motifs are. This shows the utility of the proposed approach in capturing temporal relations among unidimensional motifs and repetitions of such relationship groups.

D. Personalization and Co-adaptation

From the provided examples, it can be observed that MDMs can be used to identify and tag recurring patterns having some contextual value. While these MDMs could be used as generic templates, their utility is better served in a personalized context. This is due to the approach being designed to identify those patterns that are consistent for each individual independently. For instance, the word groups identified by the MDMs in Figure 9 are personalized patterns identified for one individual, and therefore are representative of that individuals characteristics as determined by physiology and pronunciations (accent). This would help identify any anomalies or impediments unique to the individual for personalized treatment planning. In large datasets, MDMs can also be generalized across individuals to establish the varying manifestations of patterns in pronunciations and motion characteristics.

Moreover, it should also be noted that the recurring patterns are not identical and have ample variation within their individual instances, thus validating the robustness of the approach. This aspect of the approach is very suited for learning and refining the pattern class over time, with changes in the individuals characteristics as exhibited in the sensor data, eg. progress over time in speech therapy, or improvement in limb movement due to exercise. The approach is suited for adapting and refining MDMs over time with newer instances subsuming the older ones, and thus updating the index with the most current and relevant forms of recurring patterns. This can be used for conformance as well as tracking progress during treatment.

E. Runtime Performance and Scalability

In order to evaluate the scalability of the proposed method, we studied the processing time per window (the sliding window used for processing data) and the Processing delay per MDM, and their variations with increasing processing window size and increasing number of dimensions. All computations were done using a system with 3.4 GHz Intel dual core processor with 16 GB RAM. Table II and III present the results of our analysis using two datasets. The first dataset is taken from the PhysioNet/Computers in Cardiology Challenge 2009 Test Set B dataset [6] (see Figure 1) that provides synchronous body sensor data, including heart rate, systolic blood pressure, diastolic blood pressure, mean arterial pressure, and respiratory rate. The section of data used in our analysis has 5 dimensions of synchronous time series data having 5000 samples each, with a sampling rate of 125 Hz. The results (see Table II) indicate that the non-uniform variation the processing time per timestamp is not predictable, due to dependence on the frequency and concentration of MDMs discovered, which is very dataset-specific. The second dataset used is the Electromagnetic Articulation (EMA) dataset (described in Section V.B) having 10000 samples (sampling rate 200 Hz). It can be seen from Table III that the average time taken to process information generated over 2.5 seconds (500 samples), which is the time taken to utter a sequence of 4-5 words, is of the order of a milliseconds for data having upto 30 dimensions. This average time is extremely low in cases where most windows are processed extremely fast, with the exception of a few that take a large amount of time. Also, the average delay between the arrival and capture of an MDM in such data is only about 100 milliseconds.

This illustrates the scalability of the method for processing data having large number of dimensions in real-time, since the concurrent mining of the unidimensional and multidimensional motifs can be achieved within a very small fraction of the actual duration of recording. This can also be helpful in cases where the physician marks interesting or anomalous patterns and can be alerted when the patterns recur in realtime. It should be noted that the processing time would vary with the concentration of unidimensional motifs mined and frequency of the MDMs discovered, since a larger number of motifs results in a “pattern explosion”, thus requiring much more time to generate and index pattern combinations. With increasing dimensions, execution times may be handled better using distributed processing or multi-threading.

VI. Conclusion

Monitoring patient conformance during personalized rehabilitative activities requires tracking repeating patterns in multidimensional time series data from body sensor networks with heterogeneous sensors. Multidimensional Motifs (MDMs) provide an intuitive and effective solution for this purpose without requiring prior evidence unlike classification techniques. This paper presents an enhanced and efficient method for MDM discovery, with ability to concurrently process multidimensional time series data, and identify and index MDMs and their instances using an efficient hashing based indexing approach. The unordered and temporally ordered models proposed for MDMs are beneficial in tracking variable length motif cooccurrences and temporal orderings among motifs across multiple

dimensions. The scalability and efficiency of the approach holds promise for online analysis and monitoring in healthcare applications. The utility of the proposed approach in concurrent processing of multidimensional time series, finding unknown naturally occurring patterns with minimal delay, and tracking repetitions and flagging patterns in real-time during therapy sessions makes it a valuable assistive tool in analysis and interpretation of body sensor data characteristics. Our future endeavors in this direction would involve extending the MDM mining approach to associated applications such as anomaly detection, association rule mining between MDMs, and patient grouping based on similarities based on characteristics extracted from their data in the interest of patient profiling and personalized treatment and rehabilitation.

Acknowledgments

This material is based upon work supported by the National Science Foundation Grant No.1012975, the National Institutes of Health Grant R03 DC013990 and the 2015 ASHFoundation New Century Scholar Research Grant.

References

1. Allen JF. Maintaining knowledge about temporal intervals. *Commun ACM*. Nov; 1983 26(11):832–843.
2. Armstrong, T., Drewniak, E. Proceedings of the 7th international conference on Machine learning and data mining in pattern recognition, mldm'11. Berlin, Heidelberg: Springer-Verlag; 2011. Unsupervised discovery of motifs under amplitude scaling and shifting in time series databases; p. 539-552.
3. Balasubramanian A, Prabhakaran B. Flexible exploration and visualization of motifs in biomedical sensor data. Proceedings of Workshop on Data Mining for Healthcare (DMH), in conjunction with ACM KDD. 2013
4. Chiu, B., Keogh, E., Lonardi, S. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, kdd '03. New York, NY, USA: ACM; 2003. Probabilistic discovery of time series motifs; p. 493-498.
5. Fu TC. A review on time series data mining. *Engineering Applications of Artificial Intelligence*. 2011; 24(1):164–181.
6. Goldberger AL, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*. 2000 Jun 13; 101(23):e215–e220. [PubMed: 10851218]
7. Kostakis, O., Papapetrou, P., Hollmén, J. Artemis: Assessing the similarity of event-interval sequences. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M., editors. ECML/PKDD (2), volume 6912 of Lecture Notes in Computer Science. Springer; 2011. p. 229-244.
8. Li Y, Lin J, Oates T. Visualizing variable-length time series motifs. Proceedings of SIAM Data Mining (SDM). 2012:895–906.
9. Lin, J., Keogh, E., Lonardi, S., Chiu, B. Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, dmkd '03. New York, NY, USA: ACM; 2003. A symbolic representation of time series, with implications for streaming algorithms; p. 2-11.
10. McGovern A, Rosendahl DH, Brown RA, Droegemeier KK. Identifying predictive multi-dimensional time series motifs: an application to severe weather prediction. *Data Min Knowl Discov*. Jan; 2011 22(1-2):232–258.
11. Minnen D, Isbell C, Essa I, Starner T. Detecting subdimensional motifs: An efficient algorithm for generalized multivariate pattern discovery. *Data Mining, 2007 ICDM 2007 Seventh IEEE International Conference on*. 2007:601–606.
12. Moazeni M, Mortazavi B, Sarrafzadeh M. Multi-dimensional signal search with applications in remote medical monitoring. *Body Sensor Networks (BSN), 2013 IEEE International Conference on*. 2013:1–6.

13. Mueen, A., Keogh, E. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '10. New York, NY, USA: ACM; 2010. Online discovery and maintenance of time series motifs; p. 1089-1098.
14. Nevill-Manning CG, Witten IH. Identifying hierarchical structure in sequences: a linear-time algorithm. *J Artif Int Res. Sep*; 1997 7(1):67–82.
15. Ordóñez P, Armstrong T, Oates T, Fackler J, Lehman CU. Multivariate methods for classifying physiological data. Workshop on Data Mining Medicine and HealthCare (DMMH 2013) at 2013 SIAM International Conference on Data Mining. 2013:37–45.
16. Patel, D., Hsu, W., Lee, M., Parthasarathy, S. Lag patterns in time series databases. In: Bringas, P.Hameurlain, A., Quirchmayr, G., editors. Database and Expert Systems Applications, volume 6262 of Lecture Notes in Computer Science. Springer; Berlin Heidelberg; 2010. p. 209-224.
17. Patel, D., Hsu, W., Lee, ML. Proceedings of the 2008 ACM SIGMOD international conference on Management of data, SIGMOD '08. New York, NY, USA: ACM; 2008. Mining relationships among interval-based events for classification; p. 393-404.
18. Patel D, Hsu W, Lee ML. Integrating frequent pattern mining from multiple data domains for classification. *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*. 2012:1001–1012.
19. Pradhan G, Engineer N, Nadin M, Prabhakaran B. Analyzing motoric and physiological data in describing upper extremity movement in the aged. *Universal Access in the Information Society*. 2011; 10(2):139–150.
20. Pradhan GN, Prabhakaran B. Analyzing and visualizing jump performance using wireless body sensors. *ACM Trans Embed Comput Syst. Aug*; 2012 11(S2):47:1–47:26.
21. Tanaka Y, Iwamoto K, Uehara K. Discovery of time-series motif from multi-dimensional data based on mdl principle. *Machine Learning*. 2005; 58(2-3):269–300.
22. Vahdatpour, A., Amini, N., Sarrafzadeh, M. Proceedings of the 21st International Joint Conference on Artificial intelligence, IJCAI'09. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; 2009. Toward unsupervised activity discovery using multi-dimensional motif detection in time series; p. 1261-1266.
23. Wang J, Green JR, Samal A, Yunusova Y. Articulatory distinctiveness of vowels and consonants: A data-driven approach. *Journal of Speech, Language, and Hearing Research*. 2013; 56(5):1539–1551.
24. Wang J, Samal A, Green J, Rudzicz F. Whole-word recognition from articulatory movements for silent speech interfaces. 13th Annual Conference of the International Speech Communication Association (InterSpeech). Sep.2012
25. Yunusova Y, Green J, Mefferd A. Accuracy assessment for ag-500 electromagnetic articulograph. *Journal of Speech, Language, and Hearing Research*. 2009; 52(2):547–555.
26. Madeira SC, Teixeira MC, Sa-Correia I, Oliveira AL. Identification of Regulatory Modules in Time Series Gene Expression Data Using a Linear Time Biclustering Algorithm. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. Jan-Mar;2010 7(1):153–165. [PubMed: 20150677]
27. Goncalves JP, Madeira SC. LateBiclustering: Efficient Heuristic Algorithm for Time-Lagged Bicluster Identification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. Sep-Oct;2014 11(5):801–813. [PubMed: 26356854]
28. Supper J, Strauch M, Wanke D, Harter K, Zell A. EDISA: extracting biclusters from multiple time-series of gene expression profiles. *BMC Bioinformatics*. 2007; 8(1):1–14. [PubMed: 17199892]
29. Amar D, Yekutieli D, Maron-Katz A, Hendler T, Shamir R. EDISA: extracting biclusters from multiple time-series of gene expression profiles. *Bioinformatics*. 2015; 31.12:i17–i26. [PubMed: 26072479]

Biographies

Arvind Balasubramanian is pursuing a PhD in Computer Science at the University of Texas at Dallas. His research interests are in Data Mining and Pattern Discovery, with special focus on Biomedical and Healthcare domain applications.

Jun Wang is an Assistant Professor of Biomedical Engineering and Communication Disorders at the University of Texas at Dallas. He earned his PhD in Computer Science from the University of Nebraska-Lincoln in 2011. His research areas include silent speech recognition, motor speech disorders, and computational neuroscience for speech communication.

Balakrishnan Prabhakaran is a Professor of Computer Science at the University of Texas at Dallas. He is a Member of the Executive Council of the ACM Special Interest Group on Multimedia (SIGMM) and is the Co-Chair of IEEE Technical Committee on Multimedia Computing (TCMC) Special Interest Group on Video Analytics (SIGVA). His research has been funded by Federal Agencies such as the National Science Foundation (NSF) and USA Army Research Office (ARO).

Fig. 1.

Various human body sensors recording unidimensional and multidimensional data. Examples of synchronously captured multidimensional data from heterogeneous sensors, and motifs identified (Data from [6]).

Fig. 2.

Example from Electromyogram (EMG) sensor data where muscle activity has different onset times than the associated motion activity.

Fig. 4.

(a) String encoding time series using SAX; (b) Motif mining from a SAX symbol string using Sequitur.

Fig. 5. The inverted timestamp index for the example time series shown in Figure 3, and the list of motif neighborhoods with starting timestamp.

Fig. 6.

(a) MBIndex: Hashtable recording the motifBags as keys and the corresponding time stamp lists; (b) MOIndex: Hashtable recording the MotifOrderings as keys and the corresponding timestamp lists.

Fig. 7.

Results of Unordered Multidimensional Motif Discovery on Synthetic Dataset with Planted Motifs of length of 60, 60, 90 and 50 samples, respectively. (a), (b) and (c) The three UMDMs in the dataset with their instances (highlighted in red, better viewable on electronic version). Please note that variable length motifs are captured, with motif lengths sometimes exceeding the base sliding window length, such as the one in (b).

Fig. 8. Examples of (a) Unconstrained TMDM and (b) Constrained TMDM (highlighted in red, better viewable on electronic version).

Fig. 9. Multidimensional motifs identified in Articulatory data (Sampling rate 200 Hz) for a sequence of 25 words repeated twice. In each of the above cases, the instances of the detected MDM correctly correspond to repetitions of the same word sequence – (a) ‘start, moose, trash, awe’; (b) ‘mute, them, fate, tang’; (c) ‘rap, vast, dab, ways’.

Fig. 10.

Examples of TMDMs from a section of the Mocap-EMG data for 10 repetitions of the “Raise Arm” activity (sampling rate 120 Hz) – (a) onset times for wrist joint displacement and tibialis anterior muscle activity are the same; (b) onset time of tibialis anterior muscle precedes that of wrist joint displacement (TMDMs are highlighted in red, better viewable on electronic version).

Table I
Temporal Relations set by Allen's Interval Algebra

<i>R</i>	Temporal Relation	Definition
R_1	<i>Precedes</i> (m_a, m_b)	$R_1 end_ts(m_a) < beg_ts(m_b)$
R_2	<i>Meets</i> (m_a, m_b)	$end_ts(m_a) = beg_ts(m_b)$
R_3	<i>Overlaps</i> (m_a, m_b)	$(beg_ts(m_a) < beg_ts(m_b)) \wedge (end_ts(m_a) < beg_ts(m_b)) \wedge (end_ts(m_a) > beg_ts(m_b))$
R_4	<i>Starts</i> (m_a, m_b)	$(beg_ts(m_a) = beg_ts(m_b)) \wedge (end_ts(m_a) < end_ts(m_b))$
R_5	<i>During</i> (m_a, m_b)	$(beg_ts(m_a) > beg_ts(m_b)) \wedge (end_ts(m_a) < end_ts(m_b))$
R_6	<i>Finishes</i> (m_a, m_b)	$(beg_ts(m_a) > beg_ts(m_b)) \wedge (end_ts(m_a) = end_ts(m_b))$
R_7	<i>Equals</i> (m_a, m_b)	$(beg_ts(m_a) = beg_ts(m_b)) \wedge (end_ts(m_a) = end_ts(m_b))$

Table II
Performance of Proposed Approach for varying length of processing window in PhysioNet data with 5 dimensions having 5000 samples each

Processing Window length	Average Processing Time per Window	Average Processing Delay Per MDM
Between 100 and 1000 samples	Varying non-uniformly between 2 ms and 7 ms	Varying non-uniformly between 2 ms and 64 ms

Table III
Performance of Proposed Approach for increasing number of dimensions in EMA data
with 10,000 samples, with a processing window of 500 samples (2.5 sec)

No. of Dimensions	Processing Time per Window (Avg, Min, Max) (ms)	Processing Delay Per MDM (Avg, Min, Max) (ms)
5	(0, 0, 3)	(3, 1, 5)
10	(0, 0, 4)	(4, 1, 7)
15	(0, 0, 9)	(7, 1, 14)
20	(0, 0, 36)	(18, 1, 39)
25	(1, 0, 91)	(49, 1, 94)
30	(1, 0, 144)	(95, 1, 159)