Extracting interpretable muscle activation patterns with Time Series Knowledge Mining

Fabian Mörchen* Alfred Ultsch† Olaf Hoos‡

May 16, 2005

Abstract

The understanding of complex muscle coordination is an important goal in human movement science. There are numerous applications in medicine, sports, and robotics. The coordination process can be studied by observing complex, often cyclic movements, which are dynamically repeated in an almost identical manner. The muscle activation is measured using kinesiological EMG. Mining the EMG data to identify patterns, which explain the interplay and coordination of muscles is a very difficult Knowledge Discovery task. We present the *Time Series Knowledge Mining* framework to discover knowledge in multivariate time series and show how it can be used to extract such temporal patterns.

*corresponding author, Data Bionics Research Group, University of Marburg, Hans-Meerwein-Str., D-35032 Marburg, Germany, phone +49-6421-2823922, fax +49-6421-2828902, fabian@informatik.uni-marburg.de

†Data Bionics Research Group, University of Marburg, D-35032 Marburg, Germany

‡Department of Sports Medicine, University of Marburg, D-35032 Marburg, Germany

1 Introduction

In the research field of biomechanics and human movement science the kinesiological EMG (Electromyography) has been used extensively to gain insight into function and coordination of muscles in different movements and postures [CC93], [CC97], [Cla00]. A few recent publications have shown promising results using the kinesiological EMG in addition to kinematical data to discover and classify movement patterns by means of artificial neural networks (e.g. [HK93], [VVVBB01], [Cha01]). The neural networks are usually used for classification and prediction after supervised training. They do not, however, explain the correspondence of certain EMG patterns to the movements. With this supervised learning process the ultimate goal of Knowledge Discovery is not reached.

The representation of complex muscle coordination in terms of interpretable activation patterns would open up new possibilities to optimize, manipulate, or imitate the movements. The discovered knowledge could be used in medicine (e.g. rehabilitation), sports (e.g. movement optimization), and robotics (e.g. movement simulation). In this paper we define and use our temporal rule mining framework, called *Time Series Knowledge Mining* (TSKM) [MU04a], to discover interpretable temporal movement patterns from raw EMG measurements. The data is collected from tests in professional In-Line Speed Skating.

Our temporal rule mining framework decomposes the problem of rule discovery into the mining of single temporal concepts. The resulting rules have a hierarchical structure corresponding to these concepts. This gives unique opportunities in relevance feedback during the Knowledge Discovery process and in the interpretation of the results. An expert can focus on particularly interesting rules and discard valid but known rules before the next level constructs are searched. After obtaining the final results, an expert can zoom into each rule to learn about how it is composed and what its meaning and consequences might be. The decomposition is also of advantage for the mining algorithms, because the hypothesis space for a single mining step is smaller.

The patterns are expressed with the *Unification-based Temporal Grammar* (UTG) [Ult96], [Ult04b], a hierarchical rule language developed especially for the description of patterns in multivariate time series. The whole process is data driven and the resulting rules can be used for classification. Most importantly, they also offer further insight into the underlying biomechanical and neuromuscular phenomena.

In Section 2 we mention some alternative methods. Section 3 explains the temporal concepts, defines the minig steps of the framework, and gives details on two algorithms. The application of TSKM in sports medicine is described in detail in Section 4. The results are discussed in Section 5. Section 6 summarizes the paper.

2 Related work

The following two approaches extract rules from multivariate time series. Both convert the time series to labeled intervals using segmentation and feature extraction. Höppner [Höp02] mines temporal rules expressed with Allen's [All83] interval logic and a sliding window to restrict the pattern length. The patterns are mined with an *a priori* algorithm using support and confidence and ranked by an interestingness measure afterwards. Last *et al.* [LKK01] mine association rules on adjacent intervals using the Info-Fuzzy Network (IFN). The rule set is reduced using Fuzzy theory.

Kam and Fu [KF00] also use Allen's interval operators to formulate patterns in interval sequences from a temporal database. Such a sequence could also be obtained from time series in analogy to the approaches above. The rules are restricted to so called A1 patterns, that only allow concatenation of operators on the right hand side. The patterns are mined with an *a priori* algorithm.

We think, that the strict use of Allen's interval relations is problematic. With many real life time series no distinction should be made between intervals starting or ending exactly at the same time or within a small time window. The relations also mix very different temporal concepts. Two overlapping intervals represent the concepts of order and coincidence. We think, that the separation of these concepts in our rule language allows are more intuitive interpretation in many domains. Since each algorithm is mining only a single temporal concept, the hypothesis spaces are smaller than when mining the set of Allen's relations in one step. As disadvantages of the method by Last, we see the restriction to rules on neighboring intervals, the lack of gaps in the interval sequence obtained from one time series, and the need to specify a single target attribute.

3 Time Series Knowledge Mining

The temporal knowledge discovery framework Time Series Knowledge Mining (TSKM) aims at finding interpretable symbolic rules describing interesting patterns in multivariate time series. We define the data models, mining task, and algorithms for each level of the framework. Some tasks can be solved with well known data mining algorithms, while other require new algorithms. The rules are expressed in a language called *Unification-based Temporal Grammar* (UTG).

3.1 Unification-based Temporal Grammar

The UTG is a hierarchical temporal extension of unification-based grammars, developed especially for the description of patterns in multivariate time series (MVTS) [Ult96]. A hierarchy of semiotic levels is built from simple patterns, covering the temporal concepts duration, coincidence, and order. We briefly describe the elements of the UTG, for a more detailed description of the UTG see [Ult04b, MU04b].

At each hierarchical level the grammar consists of semiotic triples: a unique symbol (syntax), a grammatical rule (semantic), and a user defined label (pragmatic). The grammatical rule is produced by a mining algorithm for this hierarchy level (see below). An expert is needed to interpret the rule and complete the triple with a meaningful label. Optionally, the symbol can also be set to a meaningful abbreviation. An example of such a semiotic triple is shown in Figure 1. The generated rule describes a state assignment for a time point. The skater is in a certain movement phase if the angles measured at the hip and the ankle are within the respective intervals. An expert diagnosed this as the *preparation phase* for the foot contact, assigned the appropriate label, and choose *pre* as a symbol.

```
symbol: pre
rule: A PrimitivePattern is a 'pre' if
'hip' in [142.39,166.46] and 'ankl' in [63.19,79.44].
label: preparation phase
```

Figure 1: An example of a semiotic triple

In absence of a domain expert, the unique symbols and labels can be generated automatically during the mining process, but they should be adjusted later for better interpretation. One semiotic triple describes a class of constructs, each class usually has many instances occuring at certain time points or during certain intervals. On each semiotic level we allow a special blank symbol, called *Tacet* (the pausing of an instrument or voice in a musical piece), to express the fact, that there is no UTG construct for a time point or interval. Short, not plausible, interruptions of an otherwise persisting state are called *Transients*. The maximum length for Transients is application and level dependent.

Figure 2 shows how the UTG elements are formed in the space of time, features and alternatives. A *Primitive Pattern* describes a process state for a single point in time, shown as circles. It represents a temporal atom, because it has unit duration. A *Succession* expresses the temporal concepts of duration. It represents a time interval where nearly all time points in this interval have the same Primitive Pattern label (rectangles A and B). A projection in the direction of the feature axis forms intervals where the states coincide. This is called

Event (rectangle A+B). If the start points of all overlapping Succession are approximately equal and the same is true for the end points, the Event is called *synchronous*. Events thus represent the temporal concepts of coincidence, and synchronicity. A typical ordered occurence of Events forms the next hierarchy level, called *Sequences*. A Sequence expresses the temporal concept of order. Several similar ordered patterns, depicted by mostly the same gray shades, can be merged to form the final Temporal Pattern.

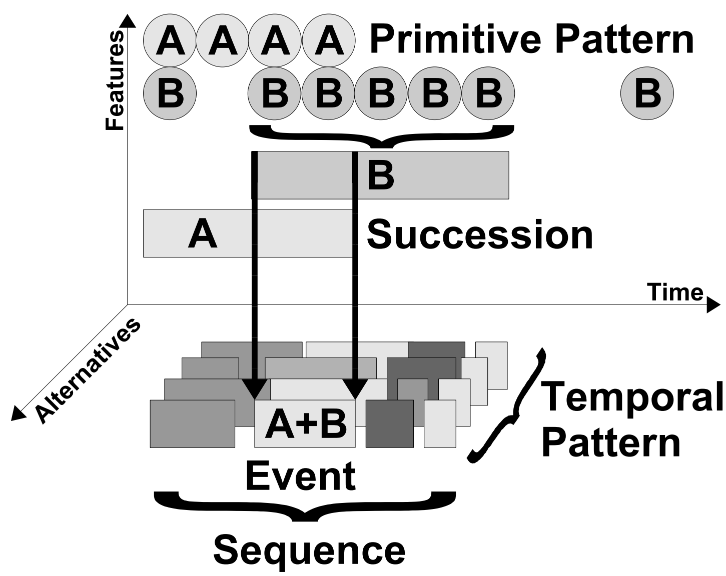

Figure 2: Hierarchy of Unification-based Temporal Grammar concepts

3.2 TSKM Framework

The starting point of the TSKM is a multivariate time series, usually but not necessarily uniformly sampled. The knowledge discovery steps of the TSKM are shown in Figure 3. First, preprocessing and feature extraction techniques should be applied where necessary (1). An expert should group the set of time series into possibly overlapping subsets, called *Aspects* (2). The series within an Aspect should be related w.r.t. the investigated problem domain. In the absence of such prior knowledge, one Aspect per time series can be used. The discretization of the time series in each aspect produces Primitive Patterns (3). Merging of these atomic states creates the Successions (4). Mining the coincidence of states from different aspects gives a series of Events (5). Typical Sequences of Events are searched in the next step (6). Similar Sequences can be merged to form the final Temporal Patterns (7). Note, that for the first two temporal description levels, each Aspect is treated individually. We want to emphasize, that the discovery process is not linear in practice, because often some parameter settings or decisions need to be refined based on partial results. This is shown by the dashed arrows. The data models and proposed algorithms for each step are described in the following Sections.

3.3 Finding Primitive Patterns

The task of finding Primitive Patterns is the reduction of the time series to a series of states. The input data is a real or vector valued time series, the output is a time series of symbols for each atomic time interval. It is important, that each symbol is accompanied by a rule and a linguistic description to complete the semiotic triple. Many discretization techniques can be used to find Primitive Patterns for univariate Aspects. Simple methods aggregate the values using histograms. Additionally down-sampling can be performed by aggregation over

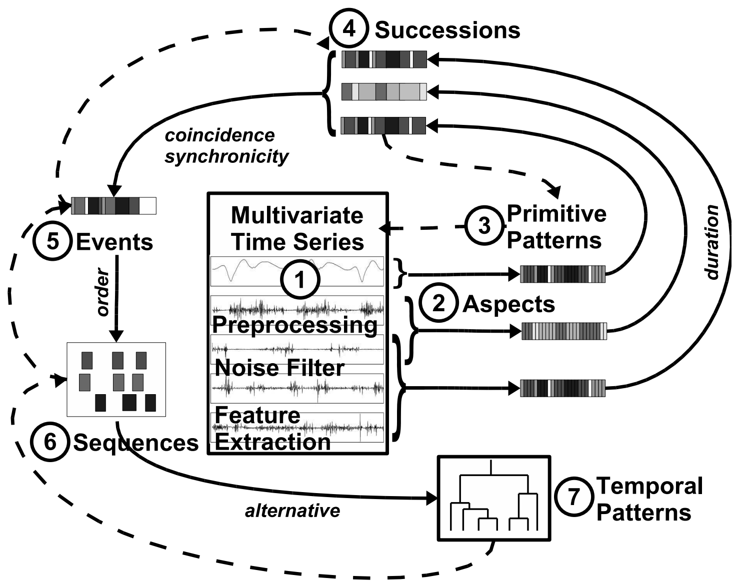

Figure 3: Circular steps of the Time Series Knowledge Mining

a time window, e.g. [LKLC03]. The symbols for the bins can easily be mapped to linguistic descriptions like *high* or *low*. A first order description method describes the current trend of a time series, e.g. [Kad99]. Second order descriptions additionally incorporate the second derivative of the signal to distinguish convex from concave trends, e.g [Höp01].

For Aspects spanning several time series we propose to use clustering and rule generation on the spatial attributes. If the process alternates between several regimes or states, these regions should form clusters in the high dimensional space obtained disregarding the time attribute. In [GU99] and for the identification of the skating movement phases Emergent Self-Organizing Maps (ESOM) [Ult99] have been used to identify clusters. The rules for each cluster were generated using the Sig* Algorithm [Ult91]. The ESOM enables visual detection of outliers and arbitrarily shaped clusters. Sig* aims at understandable descriptions of the Primitive Patterns. Other combinations of clustering and rule generation can be used as well, but care should be taken to obtain meaningful clusters with descriptions that can be interpreted by an expert.

3.4 Finding Successions

The input data for finding Successions is a univariate symbolic time series of Primitive Patterns, the output consist of a univariate series of labeled intervals. The merging of consecutive Primitive Patterns into a Succession is straight forward. But with noisy data there are often interruptions of a state (*Transients*). Let a Succession interval be a triple of a start point t, a duration d, and a symbol s. Let the input Successions be $S = \{(t_i, d_i, s_i)\ i = 1..n\}$ with $t_i + d_i \leq t_{i+1}$ and $s_i \neq s_{i+1}$. Let d_{max} be the maximum absolute duration and r_{max} the maximum relative duration of a Transient. For the removal of Transients we propose the *SequentialTransientFilter* algorithm shown in Figure 4.

The time complexity of the algorithm is $O(n)$. A good choice for r_{max} is 0.5, i.e. the gap is allowed to be at most half as long as the surrounding segments together. The d_{max} parameter has to be chosen w.r.t. the application. Often, some knowledge on the minimum duration of a phenomena to be considered interesting is available.

3.5 Finding Events

Events represent the concept of coincidence, thus in this step all Aspects are considered simultaneously. The input data is a multivariate series of labeled intervals (Successions) and the output is a univariate series of labeled intervals (Events). Let S be a $k \times n$ matrix containing the symbols of the Successions from k Aspects at n time points. We use $S(i)$ for the i-th column of S and $S(i) = S(j)$ for element-wise equality. Let min_d be the minimum duration of an Event. The algorithm *FullEvents* shown in Figure 5 discovers all Events where Successions from all Aspects coincide.

```
i := 2
while i < n − 1
    // check symbols and duration
    if (s_{i−1} = s_{i+1}) and (d_i ≤ d_max) and (d_i ≤ r_max ∗ (d_{i−1} + d_{i+1}))
        // merge 3 intervals
        d_{i−1} := Σ_{j=i−1}^{i+1} d_j
        ∀k ∈ {i, i + 1} d_k := 0
        i := i + 2
    else
        i := i + 1
    end if
end while
// remove zero durations
S := S \ {(t_i, d_i, s_i) ∈ S | d_i = 0}
```

Figure 4: SequentialTransientFilter algorithm

The time complexity of the algorithm is $O(n)$. The d_{max} parameter can be chosen similar to the maximum duration of Transients when finding Successions. The post-processing to identify synchronous Events is rather straight forward. For each Event the maximum difference between all start points of the participating Successions are checked against a threshold and the same is done for the end points. Additionally, the SequentialTransientFilter algorithm can be applied to the resulting Event series.

```
i := 2
while i < n
    s := i
    // search end
    while S(i)=S(i-1)
        i := i + 1
    end while
    // check duration
    if i - s ≥ min_d
        add Event on [s, i - 1]
    end if
    i := i + 1
end while
```

Figure 5: FullEvents algorithm

3.6 Finding Sequences

For the step of finding Sequences there is a large number of algorithms that could be utilized. The input data is a univariate series of labeled intervals and the output is a set of subsequences thereof. For moderately sized dataset we use a suffix trie (e.g. [Vil98]). Compared to a suffix tree, all egde labels in a trie have length one. Tries are larger, but can easier be queried for patterns with wild-cards. The trie stores all subsequences up to a maximum length and can be queried with frequency and length thresholds to find the most interesting patterns. For larger datasets more scalable and robust techniques from sequential pattern mining, e.g. [YWYH02], can be used.

3.7 Finding Temporal Patterns

The Sequences often overlap. The last step of the framework is finding generalized Sequences, called Temporal Patterns. We propose to use clustering based on a string metric to find groups of similar Sequences. The Temporal Pattern can be generated by merging the patterns in a cluster using groups of symbols at positions where the patterns do not agree. We have successfully used hierarchical clustering based on the string edit distance with a dendrogram visualization.

4 Discovering movement patterns

We applied the TSKM to a multivariate time series from sports medicine. The goal was to identify typical muscle activation patterns during In-Line Speed Skating, which is a cyclic movement with complex inter-muscular coordination pattern. The Knowledge Discovery process started with multivariate electrical signals from kinesiological EMG and angle sensors and resulted in a high level description of the typical movement cycle. In order to find such typical patterns, observations of the movement cycles repeated in an almost identical manner are needed. We collected data from professional athletes, because highly trained individuals show less variation among the single cycles.

The conversion of the raw data to interpretable patterns, that can be presented to the domain expert, is a very difficult Knowledge Discovery task. The TSKM offers a hierarchical decomposition, where the expert can validate results on successive levels and help in choosing parameters for the algorithms. In the following chapters, we will demonstrate all the steps of the TSKM hierarchy and justify the parameter selections based on the results of the previous levels.

The knowledge gained from the muscle activation patterns can be used to further analyze the underlying mechanisms of an athletic performance and e.g. help to optimize the training process and improve performance. In other areas, as rehabilitation medicine, explanations about movement patterns can be used to detect and describe anomalies and help in the development of treatments. In robotics the abstraction to interpretable rules can help to develop more realistic movement models.

4.1 Experimental setup

The athletes performed a standardized indoor test of graded speed levels (4.56 m/s-8.72 m/s) on a large motor driven treadmill. On each speed level EMG and kinematic data were measured for 30 seconds, corresponding to 12-20 movement cycles depending on selected athlete and skating speed.

The EMG data of 7 major leg extensor and flexor muscles (*Medial Gastrocnemius*, *Gluteus Maximus*, *Rectus femoris*, *Semitendinosus*, *Tibialis Anterior*, *Tensor Fasciae Latae* and *Vastus Medialis*) of the right body side were collected using bipolar surface electrodes attached to the relevant muscles. EMG signals were pre-amplified with a gain of 2500 and band-pass filtered (low cut-off: 10 Hz, high cut-off: 700 Hz).

Lower extremity kinematics were recorded with 3 angle sensors (electrogoniometers) attached to the ankle, the knee, and the hip. An inertia switch produced a series indicating the first ground contact. All sensors were sampled at a rate of 1kHz.

4.2 Preprocessing

The angular displacement time series were low-pass filtered with a 2nd order Butterworth filter with a cutoff frequency of 20Hz to remove noise from the recording process. The numerical values were converted from the measurement scale into degrees for a better interpretation. The angular velocity was derived with a 2nd order, 5 point Savitzky-Golay differentiation filter. The regions influenced by the boundaries were removed. To prepare the signals for clustering, mean and variance were normalized per attribute.

The raw EMG signals were converted to time series describing the current energy using full-wave rectification and a moving RMS filter with a window width of 35ms. The energy values were log transformed to deemphasize very large values and achieve a more symmetric distribution.

4.3 Aspects

The 6 angular displacement and angular velocity time series were grouped into one Aspect to describe the current leg position. We will refer to this as the *Movement* Aspect. For each EMG series a univariate Aspect was created, to be able to distinguish between the single muscles. An alternative would be to group all single joint muscles and multiple joint muscles into two Aspects, respectively. The same is possible by distinguishing between flexor and extensor muscles.

4.4 Primitive Patterns

The multivariate angle Aspect was discretized using clustering with ESOM and the accompanying visualizations U-Matrix[Ult92], P-Matrix[Ult03a], and U*-Matrix[Ult04a]. The cluster descriptions were found with the Sig* rule generation algorithm.

It can be assumed that the multivariate time series contains states corresponding to movement phases. This implies a very difficult problem for clustering algorithms, because the movement is continuous and there is no clear separation between the states. Clustering with the well known k-Means algorithm is not advisable here, because it will always find exactly as many clusters as given, but it will not be clear whether the cluster boundaries are justified by the data. When using ESOM, possible cluster boundaries based on distance and/or density information can be visually detected. Also, not every point needs to be assigned to a class, allowing for transition points. In Figure 6 you can see the tiled display [Ult03a] of the U-Matrix for this dataset based on an ESOM with 16k neurons arranged on a 100×164 toroid grid [1]. The gray shades in the background visualize the local distance structure among the prototype vectors of the neurons. Data points (gray squares) that are projected in a common valley (dark) surrounded by mountains (bright) are very similar and form a cluster. We manually identified six clusters.

Even though these clusters were found using no time information, the result-

[1] The toroid grid and the tiled display show every neuron and every data point four times to avoid the disconnected display of clusters

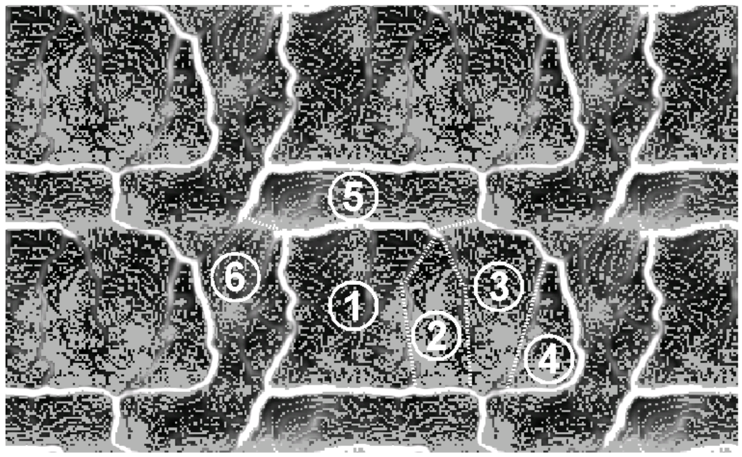

Figure 6: U-Matrix with 6 clusters for ESOM of Movement Aspect

ing Primitive Pattern are almost continuous in time, as shown in Figure 7. The top three time series show the angular speed and the other three time series the angles of hip, knee, and ankle. The shaded intervals at the very bottom show the cluster assignment of each time point (PP). The second line of intervals shows the Succession intervals after removing short interruptions (Suc). The vertical lines show the correspondence of these intervals to the multivariate movement time series used for clustering. Using this visualization and the rules describing the clusters, an expert was able to identify and label the clusters as the following movement phases: stabilization (*stabil*), forward gliding (*glide*), pre-acceleration of center of gravity (*cogacc*), preparation of foot contact (*prepft*), foot placement (*ftplc*), push-off (*push*), leg swing (*swing*). The *stabil* and *cogacc* phases were originally one cluster, but it was split due to the clear temporal separation between them, corresponding to two different functional meanings.

The univariate muscle activity series were discretized using log-normal mix-

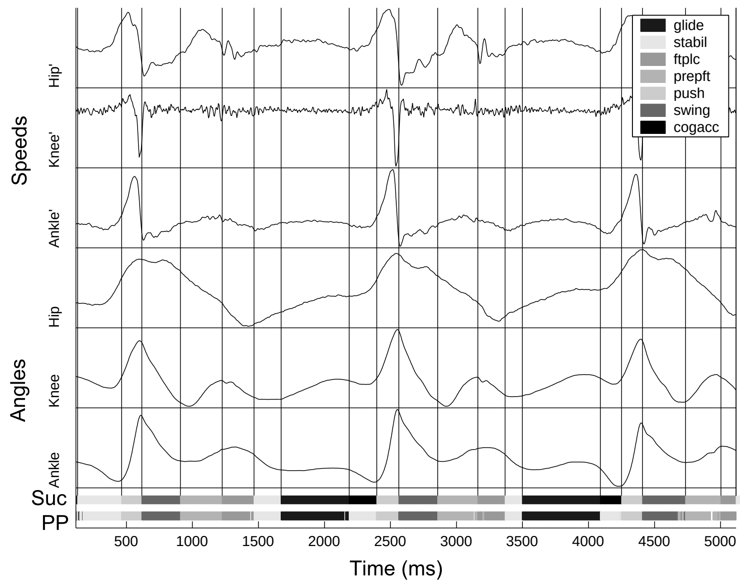

Figure 7: Primitive Patterns and Successions for 5s of the Movement Aspect

ture models. Consulting the expert, 3 states corresponding to *low*, *medium*, and *high* were chosen. Figure 8 shows one such mixture model with the bin boundaries. The dotted line shows a mixture of three Gaussians to approximate the empirical density (estimated with Pareto Density Estimation [Ult03b]) depicted with a continuous line. The boundaries between the states were chosen as the mean plus one standard deviation of the leftmost and the mean minus one standard deviation of the rightmost mixture component, respectively

The Aspect for the inertia sensor was discretized into two states, namely *on* and *off*, by a custom algorithm similar to thresholding at 95% of the maximum.

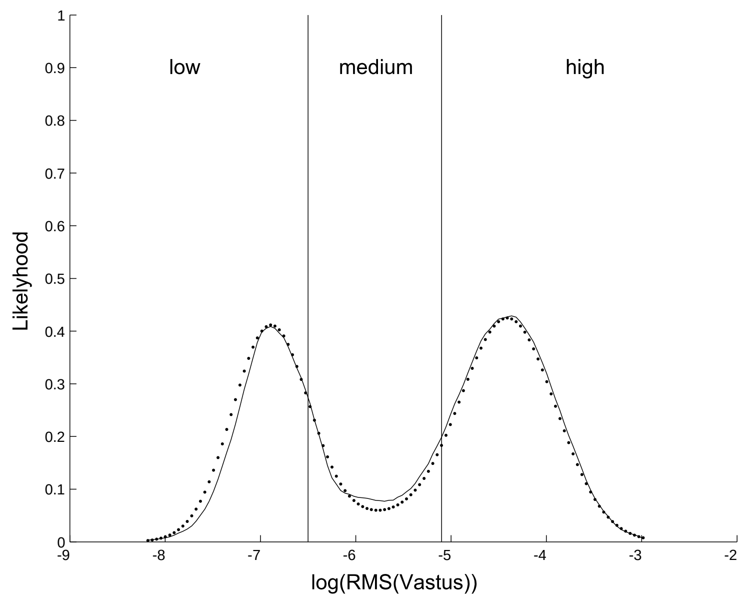

Figure 8: Mixture model for discretizing an EMG Aspect

4.5 Successions

Since the Movement Aspect describes longer movement phases, relatively large transients (up to 100ms) were filtered out. You can see the resulting Successions for a period of 5 seconds in Figure 7 in the upper interval bar at the bottom (*Suc*). Muscle activation lasting less than 50ms was also filtered out, because such short bursts are physiologically not plausible and likely to be caused by measurement inaccuracy.

4.6 Events

The visual analysis of the Successions of the muscle Aspects showed that some muscles were somewhat synchronous with the movement phases, while others showed more chaotic, apparently unrelated behavior. This correlates with biomechanical and neurophysiological background knowledge, because some muscles are directly involved in forward propulsion, while others mainly stabilize an optimized position of the body's center of gravity with respect to the supporting area, e.g. the in-line skate. For the discovery of Events, which might be functionally related to movement, we selected 3 muscles that are mainly responsible for forward propulsion: *Medial Gastrocnemius* (GA), *Vastus Medialis* (VM), and *Gluteus Maximus* (GM). Considering these 3 muscles together with the ground contact and the movement phases results in 5 Aspects.

The minimum length for Events was chosen to be 50ms, because any coordinated muscle activation in a complex movement has at least this duration. We only searched for Events involving all 5 Aspects. The minimum count for an event was set to 10. This is slightly below the number of movement cycles present in the time series (12-20).

With these setting 5 Events were found. Table 1 lists the Events with their respective counts and corresponding state labels for the Aspects. Figure 9 show the Successions of the 5 Aspects in the top rows and the instances of the 5 Events in the bottom rows for a complete 30 second time period. The Event labels were given by the expert in order to summarize the set of Succession labels.

Event E1 characterizes the *active* gliding phase which precedes the main propulsive phases cogacc and push. The occurrence of 29 events may be explained by the fact that in most movement cycles in elite in-line speed-skaters there are 2 active parts during forward gliding indicated by the bimodular activ-

Table 1: List of interesting Events with Succession labels.

Symbol	Label	Count	UTG Trigger	Move	GA	VM	GM
E1	active gliding	29	on	glide	medium	high	high
E4	relaxation	16	on	stab	medium	high	high
E5	anticipation	15	off	prepft	low	low	low
E8	weight transfer	10	on	glide	high	high	high
E10	initial gliding	11	off	swing	low	low	low

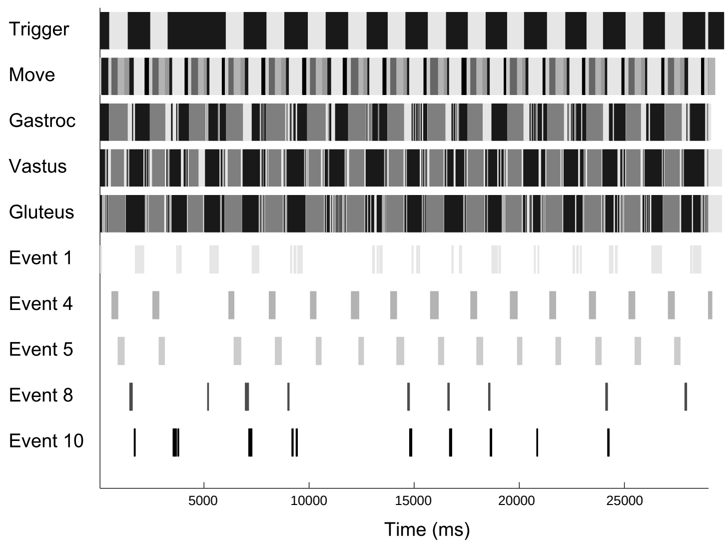

Figure 9: Occurences of Successions and interesting Events

ity of Gluteus and Vastus muscle, both showing a high activity at the beginning and at the end of the gliding phase. Events E4 and E5 mark the corresponding time frames where the major propulsive muscles are able to *relax* and the next supporting phase is *anticipated*, which is the case in each movement cycle. Event E8 seems to mirror the *weight transfer* from the left to the right leg while the final part of the push-off phase of the left side occurs. Event E10 characterizes the *initial* part of the gliding phase.

4.7 Sequences

We analyzed the Sequence of the 5 Events using a suffix trie. The minimum length of a sequence was set to 3, i.e. each Sequence must contain at least 3 Events. The minimum count was set to 5. This resulted in 3 Sequence classes listed in Table 2 with the count and an abbreviated UTG rule. The corresponding instances are depicted in Figure 10. The top five rows show the instanced of the Events (same as the bottom five rows in Figure 9). The three rows on the bottom show the time interval where the three Sequences appear. Especially the Sequence in the middle row shows a strong regularity.

Table 2: List of interesting Sequences and short UTG notation.

Symbol	Name	Count	UTG
S22	contraction & relaxation	12	E1→E4→E5
S24	movement cycle	8	E1→E4→E5→E8
S16	movement cycle +	5	E1→E4→E5→E8→E10

Sequence S22 includes the preceding and following time frames of the propulsive phases (cogacc and push) and mirrors the *contraction and relaxation* sequence that occurs in every cyclic movement. In 2 cases a separation between 2 movement cycles obviously could not be done. Sequences S24 and S16 cover the time frame for a total *movement cycle* and are quite similar. Sequence S16 additionally covers a small part of every following movement cycle.

4.8 Temporal Patterns

Due to the small number of Sequences found, and the striking similarity among them, a single Temporal Pattern was constructed manually from the rules de-

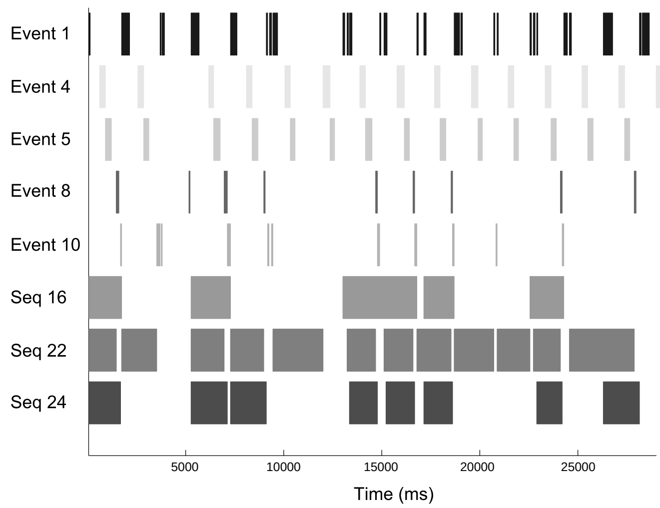

Figure 10: Occurences of interesting Events and Sequences

scribing the Sequences. Figure 11 shows the UTG rule of the final Temporal Pattern, that was labeled as the *Total Movement Cycle in In-line Speed-Skating*.

```
A TemporalPattern is a ’total movement cylce’ [1297...3773 ms]
if ’S22’ [1378...3310 ms]
or ’S24’ [1297...1848 ms]
or ’S16’ [1547...3773 ms]
```

Figure 11: The final Temporal Pattern for muscle activation

5 Discussion

Based on EMG and kinematic measurements the TSKM successfully identified functional movement patterns in In-Line Speed Skating. An understandable high level description was extracted from the raw signals. The ultimate goal of Knowledge Discovery was reached, because the model can be interpreted by the domain expert and can be automatically evaluated on new data. The transition from multivariate times series to a high level rule was done via increasingly complex knowledge representations.

The Temporal Pattern found in the skating data provided new insights for the expert. The symbolic representation offers better interpretation capabilities on the interactions of different skeletal muscles than the raw EMG data. We identified the most important cyclical motion phases. The rule describing this phase can be expanded to provide more details. A the level of Temporal Patterns there is a Sequence of Events allowing some variations. Each Event is associated with a rule listing the coinciding muscle and movement states in form of the underlying Successions. Each movement Succession is linked to a Primitive Pattern with a rule describing the range of hip, knee, and ankle angles observed during this state. We plan to compare the patterns between several skaters and running speeds to investigate possible differences. Based on background knowledge about the performance of the individual skaters this can lead to strategies for individualized training optimization.

In the TSKM process each parameter was carefully selected based on previous results. The Primitive Patterns and Successions, described the Movement

Aspect with the main movement phases and the intensity of muscle activation of the primarily propulsive lower limb muscles. The discovered Events related the muscle activity and the kinematics. They characterize typical functional aspects of the In-Line Speed Skating movement like the necessity of weight transfer from one supporting leg to the other or the active gliding process. The order of the different Events in time formed Sequences that characterize overall movement phenomena like the alternating contraction-relaxation procedure in cyclic movements or the total movement cycle itself, which is in our case as well described by the discovered Temporal Pattern.

In the current set of patterns, there is no Event for the main propulsive phases (push & cogacc). This might be explained by the fact that the search for Events was only done considering all 5 Aspects. Using smaller Events, however, increases the number of Events found and makes mining Sequences more problematic because Events can overlap. In order to identify e.g. the movement phases glide, cogacc and push-off as different Events a further refinement of the discretization of the extensor muscle activation (e.g. high/very high) might be helpful as all of the 3 extensor muscles are highly active in all 3 phases. An extension of the analysis to the antagonistic flexor muscles and the use of wavelet-transforms (e.g. [SB01]) to determine muscle activation with a better time-frequency resolution could also lead to interesting results.

6 Conclusion

We have briefly described our temporal rule language and presented the accompanying time series knowledge extraction framework. The TSKM builds up a hierarchy of UTG constructs that introduces the temporal concepts *duration*, *coincidence*, *synchronicity*, and *order* at successive levels. Rules from each level are accompanied by linguistic descriptions, thus partial results can be interpreted and filtered by experts.

The framework was successfully applied in the search for interpretable muscle activation patterns. The investigation was based on data recorded during In-Line Speed Skating describing muscle activity and the movement phases. Each parameter in the discovery process was carefully selected based on previous results. We were able to make the difficult transition from the raw measurement data to interpretable patterns via increasingly complex knowledge representations. At the highest representation level, we have identified a pattern describing the typical movement cycle. Based on the success of the TSKM, we plan on repeating the method with data from other skaters running at various speeds and investigate possible differences among the resulting patterns.

References

[All83] James F. Allen. Maintaing knowledge about temporal intervals. *Comm. ACM, 26(11)*, pages 832–843, 1983.

[CC93] J. P. Clarys and J. Cabri. Electromyography and the study of sports movements: a review. *J Sports Sci.*, 11(5):378–448, 1993.

[CC97] J. P. Clarys and J. Cabri. The use of surface electromyography in biomechanics. *Journal of Applied Biomechancis*, 13:135–163, 1997.

[Cha01] T. Chau. A review of analytical techniques for gait data. part 2: neural network and wavelet methods. *Gait Posture*, 13(2):102–20, 2001.

[Cla00] J. P. Clarys. Electromyography in sports and occupational settings: an update of its limits and possibilities. *Ergonomics*, 43(10):1750–62, 2000.

[GU99] Gabriela Guimaraes and Alfred Ultsch. A method for temporal knowledge conversion. In *Advances in Intelligent Data Analysis, Proceedings of the 3rd Int. Symp., Amsterdam, The Netherlands, Springer, Berlin*, pages 369–380, 1999.

[HK93] S. H. Holzreiter and M. E. Kohle. Assessment of gait patterns using neural networks. *J Biomech.*, 26(6):645–51, 1993.

[Höp01] Frank Höppner. Discovery of temporal patterns – learning rules about the qualitative behaviour of time series. In *Proceedings of the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases, Lecture Notes in Artificial Intelligence 2168, Springer*, 2001.

[Höp02] Frank Höppner. Learning dependencies in multivariate time series. *Proceedings of the ECAI'02 Workshop on Knowledge Discovery in (Spatio-) Temporal Data, Lyon, France*, pages 25–31, 2002.

[Kad99] Mohammed Waleed Kadous. Learning comprehensible descriptions of multivariate time series. In *Proceedings 16th International Conf. on Machine Learning*, pages 454–463. Morgan Kaufmann, San Francisco, CA, 1999.

[KF00] Po-Shan Kam and Ada Wai-Chee Fu. Discovering temporal patterns for interval-based events. In Yahiko Kambayashi, Mukesh K. Mohania, and A. Min Tjoa, editors, *Second International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2000)*, volume 1874, pages 317–326, London, UK, 2000. Springer.

[LKK01] M. Last, Y. Klein, and A. Kandel. Knowledge discovery in time series databases. *IEEE Transactions on Systems, Man, and Cybernetics 31B(2001).*, 2001.

[LKLC03] Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Bill Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, DMKD 2003*, pages 2–11, 2003.

[MU04a] F. Mörchen and A. Ultsch. Discovering temporal knowlegde in multivariate time series. In *Proceedings GfKl 2004, Dortmund, Germany*, 2004.

[MU04b] F. Mörchen and A. Ultsch. Mining hierarchical temporal patterns in multivariate time series. In *Proceedings KI 2004, Ulm, Germany*, 2004.

[SB01] Karlsson S. and Gerdle B. Mean frequency and signal amplitude of the surface EMG of the quadriceps muscles increase with increasing torque - a study using the continuous wavelet transform. *J Electromyogr Kinesol*, 11(2):131–140, 2001.

[Ult91] A. Ultsch. Connectionistic models and their integration in knowledge-based systems (german). Technical Report Report Nr. 396, University Dortmund, Germany, 1991.

[Ult92] A. Ultsch. Self-organizing neural networks for visualization and classification. In *Proceedings Conf. Soc. for Information and Classification, Dortmund, April 1992*, 1992.

[Ult96] A. Ultsch. A unification-based grammar for the description of complex patterns in multivariate time series (german), personal communication, 1996.

[Ult99] A. Ultsch. Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series. In E. Oja and S. Kaski, editors, *Kohonen Maps*, pages 33–46, 1999.

[Ult03a] A. Ultsch. Maps for the visualization of high dimensional data spaces. In *Proceedings WSOM, Japan*, 2003.

[Ult03b] A. Ultsch. Pareto density estimation. In *Proceedings GfKl 2003, Cottbus, Germany*, 2003.

[Ult04a] A. Ultsch. U*-matrix: a tool to visualize clusters in high dimensional data. Technical Report 36, Philipps-University Marburg, Germany, 2004.

[Ult04b] A. Ultsch. Unification-based temporal grammar. Technical Report 37, Philipps-University Marburg, Germany, 2004.

[Vil98] Jaak Vilo. Discovering frequent patterns from strings. Technical Report C-1998-9, Department of Computer Science, University of Helsinki, 1998.

[VVVBB01] J. Van Vaerenbergh, R. Vranken, L. Briers, and H. Briers. A neural network for recognizing movement patterns during repetitive self-paced movements of the fingers in opposition to the thumb. *J Rehabil Med.*, 33(6):256–9, 2001.

[YWYH02] Jiong Yang, Wei Wang, Philip S. Yu, and Jiawei Han. Mining long sequential patterns in a noisy environment. In *Proceedings of the 2002 ACM SIGMOD international conference on Management of data*, pages 406–417. ACM Press, 2002.