Discovering interpretable muscle activation patterns with the Temporal Data Mining Method

Fabian Mörchen, Alfred Ultsch, and Olaf Hoos

Data Bionics Research Group & Department of Sports Medicine
University of Marburg, D-35032 Marburg, Germany

1 Introduction

In this paper we demonstrate how interpretable temporal patterns can be discovered within raw EMG measurements collected from tests in professional In-Line Speed Skating. This representation of complex muscle coordination opens 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).

2 Temporal Data Mining Method

The time series Knowledge Discovery framework *Temporal Data Mining Method* (TDM) [1] is a framework of methods and algorithms to mine rules in multivariate time series (MVTS). The patterns are expressed with the hierarchical temporal rule language *Unification-based Temporal Grammar* (UTG) [2]. The whole process is data driven and the resulting rules can be used for interpretation and classification.

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.

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. An expert is needed to interpret the rule and complete the triple with a meaningful label.

After preprocessing and feature extraction the time series should be grouped into possibly overlapping subsets, called *Aspects*, related w.r.t. the investigated problem domain. The remaining TDM steps correspond to the hierarchy levels of the UTG and are described below along with the application.

3 Discovering movement patterns

The goal in analysing a multivariate time series from sports medicine 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 (3 major leg muscles), angle sensors (hip, knee, ankle), and an inertia switch indicating the ground contact. The analysis resulted in a high level description of the typical movement cycle.

One Aspect per muscle and one for the ground contact were created. The angles formed a multivariate Aspect including the angle velocities. After some preprocessing, e.g. noise filtering, the Aspects needed to be discretized to meaningful state labels, called *Primitive Patterns*.

The multivariate angle Aspect was discretized using spatial clustering with Emergent Self-Organizing Maps (ESOM) [3] and the accompanying visualizations U-Matrix, P-Matrix, and U*-Matrix[4]. The cluster descriptions were found with the rule generation algorithm Sig*[5]. Even though these clusters were found using no time information, the resulting Primitive Patterns were continuous in time. An expert was able to identify and label the clusters as movement phases, e.g. *forward gliding*.

The univariate muscle activity series were discretized using log-normal mixture models. Consulting the expert, 3 states corresponding to *low*, *medium*, and *high* were chosen. The Aspect for the inertia sensor was discretized into two states, namely *on* and *off*.

The union of consequtive Primitive Patterns compose a *Succession* representing the concept of duration. Short, physiologically not plausible, interruptions of otherwise persisting states were filtered out.

The coincidence of several Successions from the different Aspects is an *Event*. The minimum length for Events was chosen to be 50ms, because any co-ordinated muscle activation in a complex movement has at least this duration. The minimum count for an event was set to 10, slightly below the number of movement cycles present in the time series (12-20).

With these setting 5 Events listed in Table 1 were found. The Event labels were given by the expert in order to summarize the set of Succession labels.

Table 1. Events

Symbol	Label	Count	Duration (ms)					Muscles		
			Min	Mean	Max	Trigger	Move	GA	VM	GM
E1	active gliding	29	53	150	476	on	fwd. gliding	medium	high	high
E4	relaxation	16	119	280	375	on	stabilization	medium	high	high
E5	anticipation	15	217	274	336	off	prep. foot	low	low	low
E8	weight transfer	10	57	91	160	on	fwd. gliding	high	high	high
E10	initial gliding	11	51	93	170	off	leg swing	low	low	low

Within the Events we searched for typical *Sequences*, representing the concept of order. The minimum length of a sequence was set to 3 and the minimum count to 5. The resuling Sequences are listed in Table 2.

Table 2. Events

Symbol	Name	Count	Duration (ms) Min	Mean	Max	UTG
S22	contraction & relaxation	12	1378	1848	3310	E1→E4→E5
S24	movement cycle	8	1297	1596	1848	E1→E4→E5→E8
S16	movement cycle +	5	1547	2143	3773	E1→E4→E5→E8→E10

The last TDM step joins several similar Sequences based on the concept of alternative. Due to the small number of Sequences found, and the striking similarity among them, a single Temporal Pattern was constructed manually from the rules describing the 3 Sequences and was labeled as the *Total Movement Cycle in In-line Speed-Skating*.

4 Summary

Based on EMG and kinematic measurements the TDM successfully identified functional movement patterns in In-Line Speed Skating. We were able to make the difficult transition from the raw measurement data to interpretable patterns via increasingly complex knowledge representations. Each parameter in the discovery process was carefully selected based on previous results. The ultimate goal of Knowledge Discovery was reached, because the model was interpreted by the domain expert and found novel and interesting. Note that it can also be automatically evaluated on new data. We plan on repeating the method with data from other skaters running at various speeds and investigate possible differences among the resulting patterns.

References

1. Mörchen, F., Ultsch, A.: Discovering temporal knowlegde in multivariate time series. In: GfKl 2004, Dortmund, Germany. (2004)
2. Ultsch, A.: Unification-based temporal grammar. Technical Report 37, Philipps-University Marburg, Germany (2004)
3. Ultsch, A.: Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series. In Oja, E., Kaski, S., eds.: Kohonen Maps. (1999) 33–46
4. Ultsch, A.: U*-matrix: a tool to visualize clusters in high dimensional data. Technical Report 36, Philipps-University Marburg, Germany (2004)
5. Ultsch, A.: Connectionistic models and their integration in knowledge-based systems (german) (1991)