Inferring Definite-Clause Grammars to Express Multivariate Time Series

Gabriela Guimarães[1], Luís Moniz Pereira[1]

CENTRIA (Centre for Artificial Inteligence),
Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
{gg, lmp}@di.fct.unl.pt
http://centria.di.fct.unl.pt/

Abstract. In application domains such as medicine, where a large amount of data is gathered, a medical diagnosis and a better understanding of the underlying generating process is an aim. Recordings of temporal data often afford an interpretation of the underlying pattens. This means that for diagnosis purposes a symbolic, i.e. understandable and interpretable representation of the results for physicians, is needed. This paper proposes the use of definitive-clause grammars for the induction of temporal expressions, thereby providing a more powerful framework than context-free grammars. An implementation in Prolog of these grammars is then straightforward. The main idea lies in introducing several abstraction levels, and in using unsupervised neural networks for the pattern discovery process. The results at each level are then used to induce temporal grammatical rules. The approach uses an adaptation of temporal ontological primitives often used in AI-systems.

1 Introduction

In several application domains, such as medicine, industrial processes, meteorology, often a large amount of data is recorded over time. The main aim lies in performing a diagnosis of the observed system. For example, consider an EEG recording to diagnose different sleep stages, or a chemical plant that goes through different process states, or the development of hail cells that possibly originate severe hailstorms, and so on. In all these cases, several types of processes are observed and problem specific diagnoses are searched for. Human beings, after a training phase, often develop the ability to recognise complex patterns in multivariate time series. The reason lies in their background knowledge, and in their experience to deal with standard and non-standard situations, thereby being able to make a diagnosis analysing the time series at different time scales.

The identification of complex temporal patterns is very hard to handle with technical systems. Classical approaches in the field of pattern recognition (PR) are very useful for feature extraction, where no temporal context has to be considered [5,21]. In order to interpret temporal patterns in time series, temporal dependencies between the primitive patterns (features) have to be taken into account. Syntactic PR views complex patterns as sentences of primitive patterns. Thus, techniques for syntactic PR strongly rely on the theory of formal languages [6]. New approaches in adaptive PR and neurocomputing have recently been developed [3, 18], and enable a connection between the two approaches. In this paper we will show a way to extend adaptive PR-methods with Artificial Intelligence (AI) techniques.

Complex patterns in time series, as considered here, have to be seen in a temporal context. This requires context sensitive knowledge. And it means that context-free grammars are not powerful enough to parse context dependency in temporal series languages. Therefore, a powerful extension of context-free grammars, the so called definitive clause grammars (DCGs), is suitable. The advantage of DCGs, besides their context-dependency, lies in an easy implementation of their rules as logic statements [22]. Such an implementation enables an efficient parsing using a theorem prover like Prolog, or better still, XSB-Prolog, which can handle left recursion by means of tabling.

In section 2 related work is presented. Section 3 describes the main properties of DCGs and introduces the inference mechanism. An example in medicine to illustrate the extracted rules is given in section 4. Conclusions are presented in section 5.

2 Related work

Approaches for the extraction of a rule-based description from time series in the form of grammars or automata usually employ a pre-classification of the signals, i.e. the time series are segmented and transformed into sequences of labeled intervals. The approaches differ in the way segmentation is performed or how rules are induced from the labeld time series.

Concerning the segmentation problem, approaches have been proposed where the main patterns in the time series are pre-defined, for instance already having a classification of P-waves or QRS-complexes of an ECG signal [14], or otherwise classified using simple algorithms, like the simple waveform detection operations of local minimum or negative slope [2], or of zero-crossings in the first derivatives, in order to segment the time series into increasing/decreasing and convex/concave parts [12], or of frequent episodes from a class of episodes [16]. Other approaches use more elaborate methods for segmentation, such as information-theoretic neural networks with changeable number of hidden layers, associated with different values of the corresponding input attribute applied to [15]. The connections represent associations rules between conjunctions of input attributes and the target attribute.

A strongly related approach that also uses SOMs in combination with recurrent neural networks for the generation of automata is presented in [7]. It was used to predict daily foreign exchange rates. One-dimensional SOMs are used to extract elementary patterns form the time series. This approach, however, is limited to univariate time series. SOMs are again used for knowledge discovery of time series satellite images [13]. The images are classified by a two-stage SOM and described in regard to season and relevant features, such as typhoons or high-pressure masses. Time-dependent association rules are then extracted using a method for finding frequently co-occurring term-pairs from text. The rules are stored in a database, which then allows for high-level queries.

3 Inferring Definitive Clause Grammars from Multivariate Time Series at distinct Abstraction Levels

The induction of grammatical rules is an important issue in pattern recognition. It comprehends extraction, identification, classification, and description of patterns in data gathered from real and simulated environments. In pattern recognition this is handled at different levels, by handling primitive and complex patterns differently.

Primitive patterns are characterised and described by features. They are regarded as a whole and associated to a given class. Complex patterns always consist in a structural and/or hierarchical alignment of primitive patterns. In statistical pattern recognition, primitive patterns are identified using statistical methods [5, 21], and recently neural networks are also used [3,18]. No temporal constraints are considered here. This means pattern recognition is performed at a low-level, a data processing level.

Syntactical pattern recognition approaches, however, assume that primitive patterns have already been identified and thus are represented at a symbolic level. Primitive patterns are also building blocks of complex patterns. Here, the main goal lies in identifying and describing structural or hierarchical, and in our case temporal, relations among the primitive patterns. Methods from the theory of formal languages in computer science are suitable for this task, through regarding complex patterns as words and primitive patterns as characters of the language. The main aim is always to describe a large amount of complex patterns using a small number of primitive patterns and grammatical rules.

Definitive clause grammars (DCGs) are a powerful extension of context-free (cf-) grammars and therefore suitable for inducing temporal relations. Most applications of DCGs have been for many years in natural language parsing systems [4]. A good introduction to this formalism can be found in [20]. The use of DCGs for time series was for the first time proposed in [10].

Basically, DCGs are built up from cf-rules. In order to provide context-dependency, a DCG extends a cf-grammar by augmenting non-terminals with arguments. DCGs extend cf-grammars in three important ways [20]:

- DCGs provide *context-dependency* in a grammar, such that a word category in a text may depend on the context in which that word occurs in the text.
- DCGs allow arbitrary *tree structures* that are built up in the course of parsing, providing a representation of meaning of a text.
- DCGs allow extra conditions.

The advantage of DCGs in dealing with context-dependency lies in their efficient implementation of DCG-rules as logic statements by definitive clauses or *Horn clauses*. Now the problem of parsing a word of a language is reduced to a problem of proving a theorem in terms of a Prolog interpreter. In DCGs nonterminals are written as Prolog atoms and terminals as facts.

Inducing DCGs for multivariate time series not only affords a hierarchical and temporal decomposition of the patterns at different abstraction levels, but also an explicit temporal knowledge representation. At distinct levels, special unsupervised neural networks in an hierarchical alignment [9] allow for a successive and step-wise mining of the patterns, such that the obtained results can be converted into grammatical rules more easily. In this paper only a brief description of the abstraction levels is given. For a more detailed description of the method see [11].

The input to our system are multivariate time series sampled at equal time steps. As a result, we obtain the discovered temporal patterns as well as a linguistic description of the patterns (see Fig. 1), which can be transformed into a definite-clause grammar employed for parsing. Next, a description of the different abstraction levels is given.

Features The feature extraction process exercises a pre-processing of all time series. Pre-processing can be applied to one (e.g. FFT) or more then one time series (e.g. cross correlation). A feature is then the value of a function applied to a selection of time series with a time lag.

Primitive patterns Each primitive pattern (pp) is associated with a single point in time, forming an inseparable unit. pp´s are identified by clustering algorithms or unsupervised neural networks using features as input, and without taking time into consideration. A pp is then assigned to one of the clusters, i.e. a pp-class. Time points not associated with a pp-class are a kind of transition points or transition periods if they last long between succeeding pp´s of the same pp-class. A pp-channel is the allocation of the whole time lag with pp´s and transitions periods (i.e. a sequence of characters).

We want to point out that it is possible and even desirable to perform several feature selections for the generation of several pp-channels. The selection depends highly on the application and reduces strongly the complexity, since not all time series are considered at the same time.

Successions Temporally succeeding pp´s of the same pp-class are successions, each having a specific duration. The concept of duration and temporal relation is introduced here for the first time.

Events Here the concept of approximate simultaneity, i.e. states ocurring more or less at the same time, is introduced. An event is identified by temporal overlapping sequences at distinct pp-channels. Recurring events then belong to the same event class. Regions not associated with an event-class are regarded as transitions periods. Since the duration of events belonging to the same class may differ, event classes have a minimal and a maximal duration in the context of a sequence.

Sequences Recurrent sequences of events are the main structures in the time series, and describe a temporal order over the whole multivariate time series. Transition periods between sequences occur just as well, and also having a minimal and a maximal duration. Probabilistic automata can be used for the identification of sequences of events, where transition probabilities between events are identified and described.

Temporal patterns Finallly, the concept of similarity results in the identification of temporal patterns. Similar sequences are sequences with a small variation of events in different sequences. This aggregation enables once again a simplification of the DCGs. String exchange algorithms are suitable for the identification of temporal patterns. Temporal patterns are the final result of the whole temporal mining process and describe the main temporal structures in the multivariate time series.

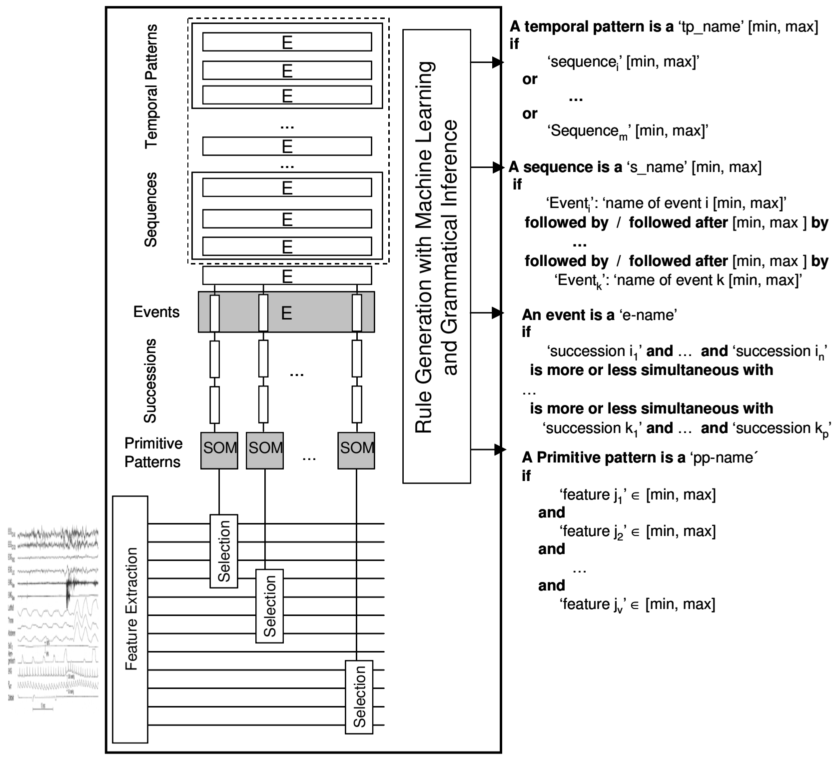

Fig. 1. A method with several abstraction levels for temporal pattern detection and for inferring Definite-Clause Grammars at distinct levels

Using the terminology of formal languages, primitive patterns can be regarded as characters used for forming words, or even complex words, in our case forming successions of characters or single ones, representing events. Sequences and temporal patterns are then composed by a sequence of events, like words form a sentence in a natural or a computer language.

As mentioned before, ML-algorithms are used to induce a rule-based and symbolic description of the pp´s. A parser for these rules can easily be implemented in Prolog [23]. A grammatical specification of events, sequences and temporal patterns presupposes that temporal dependences can be grammatically described, thus leading to the use of DCGs at higher abstraction levels. Before starting the induction process, however, an explicit temporal knowledge representation is needed. In AI a temporal reference is usually made up of a set of temporal elements, called ontological primitives (op). The main concepts for op´s are time points [17], time intervals [1], or a combination of both. For an overview to the main concepts on temporal reasoning, concerning logical formalisms in time in AI, ontological primitives, and concepts related with reasoning about action, see [24].

In this approach, a representation formalism related to Allen´s interval calculus is proposed. In the context of semi-automatic temporal pattern extraction Allen´s conception, with its 14 precedence relations, however, is far too complex and strict. For our purposes, a simpler forrmalism to describe an approximate simultaneity of events is needed, subsuming 10 of Allen´s precedence relation into a single one. Consequently, just a few op´s are needed to give a full description of the main concepts related to temporal patterns in multivariate time series. This leads to a simple and concise representation formalism built up by the following op´s:

- *and* for inclusion of features describing a primitive pattern
- *is more or less simultaneous with* describing an approximate simultaneity of successions
- *followed by* describing directly succeeding events
- *followed by … after* describing succeeding events after a transition period
- *or* for alternative (temporal) sequences

4 An example

This approach was applied to a sleep disorder with high prevalence, called sleep-related breathing disorders (SRBDs). For the diagnosis of SRBDs the temporal dynamics of physiological parameters such as sleep-related signals (EEG, EOG, EMG), concerning the respiration (airflow, ribcage and abdominal movements, oxygen saturation, snoring) and circulation related signals (ECG, blood pressure), are recorded and evaluated. Since the main aim is to identify different types of sleep related breathing disorders, mainly apnea and hypopnea, only the signals concerning the respiration have been considered [19]. Severity of the disorder is calculated by counting the number of apnea and hypopneas per hour of sleep, named respiratory disturbance index (RDI). If the RDI exceeds 40 events per hour of sleep, the patient has to be referred to therapy.

The different kinds of SRBDs are identified through the signals ´airflow´, ´ribcage movements´ and ´abdominal movements´, ´snoring´ and ´oxygen saturation´, as shown in Fig. 2, where a distinction between amplitude-related and phase-related disturbances is made. Concerning the amplitude-related disturbances, disturbances with 50%, as well as disturbances with 10-20%, of the baseline signal amplitude may occur. Phase-related disturbances are characterised by a lag between ´ribcage movements´ and ´abdominal movements´. An interruption of ´snoring´ is present at most SRBDs as well as a drop in ´oxygen saturation´.

For this experiment, 25 Hz sampled data have been used from three patients having the most frequent SRBDs. One patient even exhibited multiple sleep disorders.

In this paper we present an excerpt of the grammatical rules extracted from the results of the self-organizing neural networks at distinct abstraction levels, in order to demonstrate how the algorithm for the generation of DCGs works. These rules can be transformed into Prolog rules and parsed at a symbolic level with a Prolog interpreter.

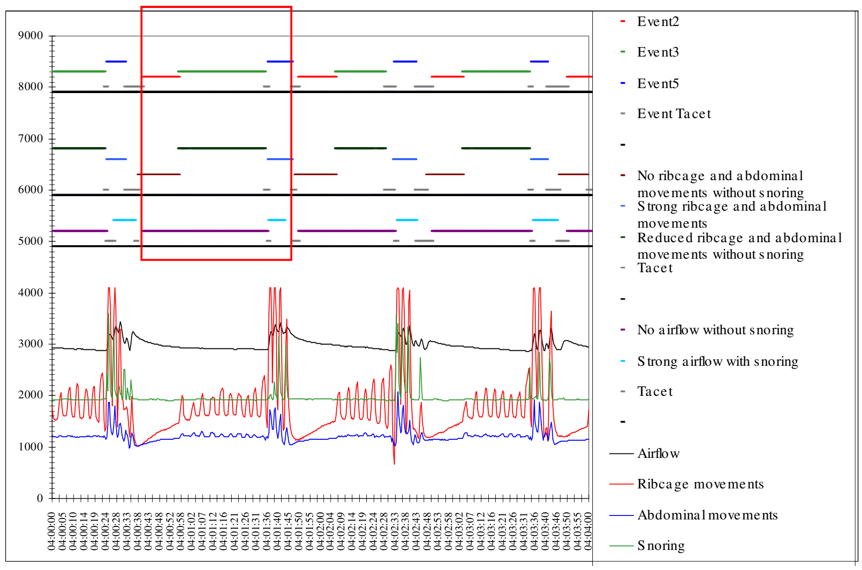

Fig. 2. Identified temporal pattern from a patient with SRBDs

For the extraction of primitive pattern rules, the ML-algorithm sig* [23] was used, which generates rules for each class based on its most significant features. For instance,

```
a pp-class is a 'A4' if
        'strong airflow' ∈ [0.37, 1]
  and   'airflow' = 0
  and   'snoring intensity' ∈ [0.15, 1]
a pp-class is a 'B6' if
        'intense abdominal movements' ∈ [0.19, 1]
  and   'reduced ribcage movements' ∈ [0, 0.84]
  and   'intense ribcage movements' ∈ [0, 1]
```

These pp-classes were named A4: *strong airflow with snoring* and B6: *intense ribcage and abdominal movements*. For the other pp-classes rules were extracted as well, and meaningful names were given. These names can be used at the next level for the description of the event-classes. For instance,

```
an event-class is a 'Event5 'if
         ('strong airflow with snoring'
    or    'reduced airflow with snoring'
    or    'transition period')
  is more or less simultaneous with
        'intense ribcage and abdominal movements'
```

This event was named *strong breathing without snoring*. The names of the event-classes are then used at the next level for the descriptions of the sequences or temporal patterns.

```
a sequence is a 'Sequence1' [40 sec, 64 sec] if
       'Event2': 'no airflow with no chest and abdominal wall
```

```
         movements and without snoring' [13 sec, 18 sec]
   followed by
        'Event3': 'no airflow with reduced chest and no
        abdominal wall movements and without snoring' [20 sec,
        39 sec]
   followed after [0.5 sec, 5 sec] by
        'Event5':'strong breathing with snoring' [6 sec,
        12 sec]
```

The rules are simple and understandable for domain experts, since they provide a linguistic description of their domain. Experts can stay with their thought pattern. The domain expert can identify the above mentioned sequence as an *mixed apnoe* and Event5 as an *hypopnoe*. Other temporal patterns were identified, namely *obstructive hypopnoe*, *mixed obstructive apnoe*, and *obstructive snoring*.

Next, a small excerpt of the DCG for the above mentioned temporal pattern is given. **Rules**

```
succession(S,D) --> succ(S), op, duration(D), cp.
…
transition(T,D) --> trans(T), op, duration(D), cp.
…
succes('E5',D1) --> succession('A4',D) ; succession('A1',D) ;
                  transition(T,D).
succes('E5',D2) --> succession('B6',D).
…
event('E5',D) --> succes('E5',D1), simultaneity,
               succes('E5',D2),range('E5',LR,UR),
               {D is (D1+D2)/2, D<UR, D>LR}.
…
sequence('S1',D) --> event('S1',D1), followedby,
            event('S1',D2),
            followedafter, transition(T,D3),
            event('S1',D4),{uplimit('S1',UD),
            lowlimit('S1',LD), D is D1+D2+D3+D4, D<UD, D>LD}.
…
duration(D) --> [D],{number(D)}.
range(D) --> [D],{number(D)}.
uplimit('S1',<value>).
lowlimit('S1',<value>).
…
Facts
trans(T) --> [transition,period].
op --> ['['].
cp --> [']',sec].
and --> [and].
or --> [or].
followedafter --> [followed,after].
followedby -->  [followed,by].
simultaneity --> [is,more,or,less,simultaneous,with].
succ('A4') --> [strong,airflow,with,snoring].
succ('A1') --> [reduced,airflow,with,snoring].
succ('B6') --> [intense,ribcage,and,abdominal,movements].
```

A structured and complete evaluation of the discovered temporal knowledge at the different abstraction levels was made by questioning an expert. All events and temporal patterns presented to the physician described the main properties of SRBDs. All of the four discovered temporal patterns described very well the domain knowledge. For one of the patterns new knowledge was even found.

5 Conclusion

The recognition of temporal patterns in time series requires the integration of several methods, as statistical and signal processing pattern recognition, syntactic pattern recognition as well as new approaches like AI-methods and special neural networks. The main idea of this approach lies in introducing several abstraction levels, such that a step-wise discovery of temporal patterns becomes feasible. The results of the unsupervised neural networks are used to induce grammatical rules. Special grammars, named DCGs, have been used here, since they are a powerful extension of context-free grammars. The main advantage in using DCGs lies in augmenting non-terminals with arguments, such as temporal constraints, as required here.

If no temporal relations have to be considered, for instance for the generation of a rule-based description of the primitive patterns, then Machine Learning algorithms can be used straightforwardly. The main advantage of our approach lies in the generation of a description for multivariate time series at different levels. This permits a structured interpretation of the final results, where an expert can navigate between rules at the same level and, if needed, zoom in to a rule at a lower level or zoom out to a rule at a higher level. This procedure provides an understanding of the underlying process, first at a coarse and later on at more and more finer granulation.

Acknowledgment We would like to thank Prof. Dr. J. H. Peter and Dr. T. Penzel, Medizinische Poliklinik, of the Philipps University of Marburg, for providing the data.

References

1. Allen, J.:Towards a General Theory of Action and Time. Artificial Intelligence 23 (1984) 123-154
2. Bezdek, J.C.: Hybrid modeling in pattern recognition and control. Knowledge-Based Systems 8, Nr 6 (1995) 359-371
3. Bishop, C.M.: Neural Networks for Pattern Recognition. Clarendow Press, Oxford (1995)
4. Bolc, L.: Natural Language Parsing Systems. Springer Verlag, New York (1987)
5. Duda, O., Hart, P.E.: Pattern Classification and Scene Analysis. John Wiley and Sons, Inc. New York (1973)
6. Fu, S.: Syntactic Pattern Recognition and Applications. Prentice-Hall, Englewood-Cliffs, N.J (1982)
7. Giles, C.L., Lawrence, S., Tsoi, A.C.: Rule Inference for Financial Prediction using Recurrent Neural Networks. In: Proceedings of IEE/IAFE Conf. on Computational Intelligence for Financial Engineering (CIFEr), IEEE, Piscataway, NJ (1997) 253-259
8. Gonzalez, R.C., Thomason, M.G.: Syntactic Pattern Recognition, Addison-Wesley (1978)
9. Guimarães, G.: Temporal knowledge discovery with self-organizing neural networks. In: Part I of the Special issue (Guest Editor: A. Engelbrecht): Knowledge Discovery from Structured and Unstructured Data, The International Journal of Computers, Systems and Signals (2000) 5-16
10. Guimarães, G.; Ultsch, A.: A Symbolic Representation for Patterns in Time Series using Definitive Clause Grammars. In: Klar,R., Opitz,R. (eds.): Classification and Knowledge Organization, 20th Annual Conf. of the Gesellschaft für Klassifikation (GFKl´96), March 6 - 8, Springer (1997) 105-111

11. Guimarães, G., Ultsch, A.: A Method for Temporal Knowledge Conversion. In: Hand, D.J., Kok, J.N., Berthold, M.R. (Eds.): Advances in Intelligent Data Analysis (IDA´99), The Third Symposium on Intelligent Data Analysis, August 9-11, Amsterdam, Netherlands, Lecture Notes in Computer Science 1642, Springer (1999) 369-380
12. Höppner, F.: Learning Dependencies in Multivariate Time Series. In: Proc. of the ECAI'02 Workshop on Knowledge Discovery in (Spatio-) Temporal Data, Lyon, France, (2002) 25-31
13. Honda, R., Takimoto, H., Konishi, O.: Semantic indexing and temporal rule discovery for time-series satellite images. In: Proceedings of the International Workshop on Multimedia Data Mining in conjunction with ACM-SIGKDD Conference, Boston, MA, 82-90, 2000
14. Koski, A., Juhola, M. Meriste, M.: Syntactic recognition of ECG signals by attributed finite automata. Pattern Recognition, The Journal of the Pattern Recognition Society 28, Issue 12, December (1995) 1927-1940
15. Last, M., Klein, Y., Kandel, A.: Knowlegde Discovery in time series databases. In: IEEE Transactions on Systems. Man and Cybernetics, Part B Cybernetics, Vol. 31, No. 1, (2001) 160-169
16. H. Mannila, H. Toivonen and I. Verkamo: Discovery of frequent episodes in event sequences. Data Mining and Knowledge Discovery 1, Nr. 3 (1997) 259-289
17. McDermott, D.: A Temporal Logic for Reasoning about Processes and Plans. Cognitive Science 6 (1982) 101-155
18. Pao, Y.-H.: Adaptive Pattern Recognition and Neural Networks. Addison-Wesley, New York (1994)
19. Penzel, T., Peter, J.H.: Design of an Ambulatory Sleep Apnea Recorder. In: H.T. Nagle, W.J. Tompkins (eds.): Case Studies in Medical Instrument Design, IEEE, New York (1992) 171-179
20. Pereira, F., Warren, D.: Definitive Clause Grammars for Language Analysis - A Survey of the Formalism and a Comparison with Augmented Transition Networks. Artificial Intelligence 13 (1980) 231-278
21. Tou, J.T., Gonzalez, R.C.: Pattern Recognition Principles, Addison-Wesley (1974)
22. Sterling, L., Shapiro, E.: The Art of Prolog. MIT Press (1986)
23. Ultsch, A.: Knowledge Extraction from Self-organizing Neural Networks. In: Opitz, O., Lausen, B., Klar, R. (eds.): Information and Classification, Berlin, Springer (1987) 301-306
24. Vila, L.: A Survey on Temporal Reasoning in Artificial Intelligence. Ai Communications 7, Nr 1 (1994) 4-28