Intelligent data interpretation and case base exploration through Temporal Abstractions

A. Bottrighi, G. Leonardi, S. Montani, L. Portinale, P. Terenziani

Dipartimento di Informatica, Università del Piemonte Orientale, Alessandria, Italy

Abstract. Interpreting time series of measurements and exploring a repository of cases with time series data looking for similarities, are non-trivial, but very important tasks.

Classical methodological solutions proposed to deal with (some of) these goals, typically based on mathematical techniques, are characterized by strong limitations, such as unclear or incorrect retrieval results and reduced interactivity and flexibility.

In this paper, we describe a novel case base exploration and retrieval architecture, which supports time series summarization and interpretation by means of Temporal Abstractions, and in which multi-level abstraction mechanisms and proper indexing techniques are provided, in order to grant *expressiveness* in issuing queries, as well as *efficiency* and *flexibility* in answering queries themselves.

Relying on a set of concrete examples, taken from the haemodialysis domain, we illustrate the system facilities, and we demonstrate the advantages of relying on this methodology, with respect to more classical mathematical ones.

1 Introduction

Analysing and interpreting long and complex **time series** of measurements is a key requirement in several real world applications, in which the pattern of feature changes is critical for decision making. This consideration strongly applies to the medical domain [16], especially in those situations in which a continuous monitoring of the patient's health indicators is required - such as in intensive care units [26] or in chronic diseases [3, 23]. As a matter of fact, in these applications several process features are naturally collected in the form of time series, either automatically generated and stored by the monitoring instruments, or obtained by listing single values extracted from temporally consecutive visits/situations (as e.g. the series of glycated haemoglobin values, measured on a diabetic patient once every two months).

Interpreting time series features on screen or on paper can be tedious and prone to errors. In a medical setting, physicians may be asked to recognize *small or rare irregularities* in the series itself, or to identify *partial similarities* with previous situations: for instance, it could be important to highlight the past occurence of a similar trend behaviour in time in a given feature (referring to the same patient, or to a different one), independently of its actual values. Such

identification may be extremely relevant for patient care, but may also require a significant amount of expertise in the specific field [25]. Similar consideration can be applied to any other application field, having time series interpretation as a core business (e.g. financial analysis, wheather forecasting, etc...). In all these domains, an automated decision support and data interpretation strategy is therefore strongly desirable.

Case-based Reasoning (CBR) [1] has recently being recognized as a valuable decision support methodology in time dependent applications (see e.g. [22], and section 4), provided that proper case representation and retrieval strategies are implemented. In particular, most of the approaches proposed in the literature to this end are based on the common premise of dimensionality reduction; this allows one to reduce memory storage and to simplify time series representation, still capturing the most important characteristics of the time series itself. Dimensionality is typically reduced by means of a *mathematical transform*, able to preserve the distance between two time series (or to underestimate it). Widely used transforms are the Discrete Fourier Transform (DFT) [2], the Discrete Wavelet Transform (DWT) [7] and the Piecewise Constant Approximation (PCA)[15]. Once a series is transformed, retrieval then works in the resulting time series space.

However, mathematical transforms have several limitations. First, they can be computationally complex, and often require additional pre-processing steps, such as mean averaging and zero padding (see e.g. [25] for details). Additionally, they work well with series with relatively simple dynamics, but they can fail to characterize more complex patterns [8]. Moreover, in many cases, they operate in a black-box fashion with respect to the end users, who just input the query and collect the retrieved cases, but usually do not see (and might not understand the meaning of) the transformed time series themselves; furthermore, users are usually unable, in such approaches, to interact with the system, during the retrieval process. Finally, mathematical transforms force the user to issue a very precise query, in which both trend and value information must be provided; indeed, retrieval is usually based on standard Euclidean distance among the transformed time series points. This requirement may be hard for an end user, as it is actually the case for a physician. Moreover, such methods may incorrectly answer a query in which e.g. only trend similarity is looked for, or get lost in time series details (e.g. small oscillations), thus not retrieving cases with a basically similar behaviour.

Our research group has recently proposed [21] to exploit a different technique for dimensionality reduction and flexible retrieval, namely **Temporal Abstractions** (TA) [30, 4, 27, 20]. TA is an Artificial Intelligence (AI) methodology able to solve a *data interpretation task* [30], the goal of which is to derive high level concepts from time stamped data. Operatively, the principle of *basic* TA methods is to move from a *point-based* to an *interval-based* representation of the data [4], where the input points (*events*) are the elements of the time series, and the output intervals (*episodes*) aggregate adjacent events sharing a common behavior, persistent over time. Basic abstractions can be further subdivided into *state*

TA and *trend* TA. *State* TA are used to extract episodes associated with *qualitative levels* of the monitored feature, e.g. low, normal, high values; *trend* TA are exploited to detect specific *patterns*, such as increase, decrease or stationary behaviour, from the time series.

Through TA, huge amounts of temporal information can then be effectively mapped to a compact representation, that not only *summarizes* the original longitudinal data, but also highlights meaningful behaviours in the data themselves, which can be *interpreted* by end users as well (in an easier way if compared to mathematical methods outputs). TA-based dimensionality reduction appears to be well suited for several application domains, and in particular for medical ones (see also section 4).

In this paper, we describe the functionalities of an intelligent case base exploration and retrieval system, which implements the techniques previously described in [21], to support data interpretation through TA.

More precisely, our system allows for *multi-level abstractions*, according to two *dimensions*, namely a taxonomy of (trend or state) symbols, and a variety of time granularities. This functionality provides **(1)** great **expressiveness** in issuing queries, which can be defined at any level of detail in both dimensions. Moreover, we provide end users with **(2)** a significant degree of **flexibility** in the retrieval process, since they are allowed to interact with the system, and progressively reduce/enlarge the retrieval set, depending on their current needs. Finally, we grant for **(3)** computational **efficiency**, since query answering takes advantage of *multi-dimensional orthogonal index structures* for focusing and early pruning.

In the next sections, by means of a set of concrete examples taken from the haemodialysis domain, we illustrate such facilities, and we demonstrate the advantages of relying on this methodology, with respect to more classical mathematical ones.

The paper is organized as follows. Section 2 describes the tool functionalities. Section 3 provides some experimental results in the haemodyalisis domain. Section 4 compares our tool with related works, and section 5 addresses our concluding remarks.

2 System functionalities

In this section, we describe the system functionalities introduced in section 1.

Our tool is composed by two main modules:

- a GUI module, which is used to interact with the system;
- a TA ENGINE module, which is responsible for building the orthogonal index structures, and for navigating them in order to support efficient and flexible retrieval.

In particular, our system allows for *multi-level abstractions*. We can abstract time series data at different levels of detail, and users can issue more general as well as more specific queries, according to two *dimensions*, which are formalized

by means of two taxonomies: (i) a taxonomy of (trend or state) *TA symbols*, and (ii) a taxonomy of *time granularities*.

Since our tool has been designed and implemented in order to be domain independent and to obtain the maximal generality, all the information which regards the description of the domain application must be taken as an input by the TA ENGINE. This holds for the two taxonomies; obviously, depending on the application domain, the definition of symbols or granules can change and the taxonomies can become wider or higher. In the rest of the paper, we exemplify it by describing a specific application to the haemodialysis domain.

The TA symbol taxonomy is organized as a conventional *is-a* taxonomy. For our experiments in the haemodialysis domain (see section 3), we have defined the taxonomy of trend symbols shown in figure 1, to which we will refer henceforth. The taxonomy is organized as follows: the symbol *Any* is specialized into D (decrease), S (stationary), I (increase), and D is further specialized into D_S (strong decrease) and D_W (weak decrease), according to the slope; I is further specialized into I_W (weak increase) and I_S (strong increase). Moreover we have introduced two special symbols, i.e. U (Unknown), M (Missing). U means that no TA symbols capture the signal behaviour (e.g. due to noise problems), while M means that data are missing (e.g. because the available monitoring instruments are not returning any data).

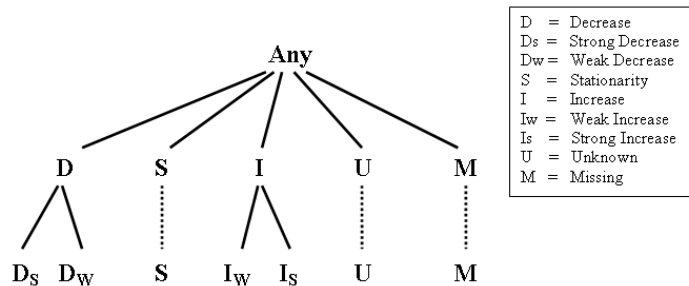|	
D	= Decrease
Ds	= Strong Decrease
Dw	= Weak Decrease
S	= Stationarity
I	= Increase
Iw	= Weak Increase
Is	= Strong Increase
U	= Unknown
M	= Missing

Fig. 1. The symbol taxonomy used in the experiments

The time granularity taxonomy allows one to describe the episodes at increasingly more abstract levels of temporal aggregation, starting from the bottom level; in our example domain, at the bottom level we find 5 minutes long granules, while at top level we aggregate episodes up to 4 hours long granules (see figure 2).

In order to abstract along the temporal dimension, a special function, called the *up* function, is needed in order to scale up from one level to the coarser one in the taxonomy. Again, since abstracting from one granularity to a coarser one is a highly domain-dependent procedure, the definition of the *up* function should be provided by domain experts and is required as an input to the TA ENGINE. Analogously, experts need to provide a proper *distance* function, to
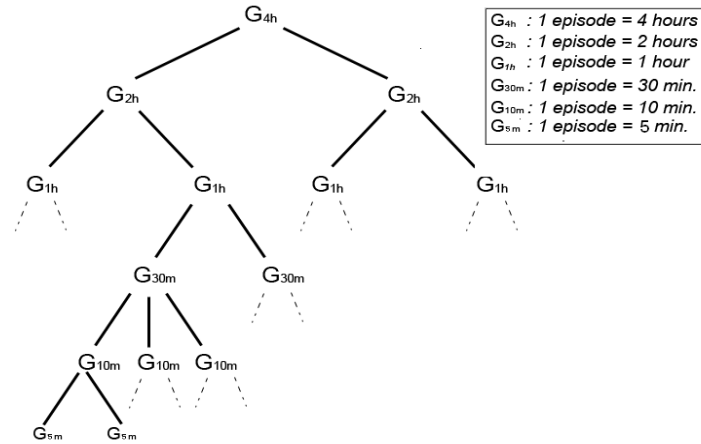
The legend box in the figure reads:

G_{4h} : 1 episode = 4 hours
G_{2h} : 1 episode = 2 hours
G_{1h} : 1 episode = 1 hour
G_{30m}: 1 episode = 30 min.
G_{10m}: 1 episode = 10 min.
G_{5m} : 1 episode = 5 min.

Fig. 2. The time granularities taxonomy used in the experiments

compute the distance between two cases. We allow the maximal generality in the *up* and *distance* function definitions, provided that they verify some very general constraints in order to avoid ambiguities (see [21] for the details).

Cases are stored in a specific database. Their features, originally in the form of time series, have been preprocessed through TA to transform them in sequences of abstractions at the lowest level of detail according to both taxonomies [1].

To perform a query, the TA ENGINE module takes advantage of a forest of multi-dimensional indexing trees which is built off-line; the engine also relies on the time and symbols taxonomies and on the *up* function, for the generalization task. Each tree indexes a portion of the case base, accordingly to the trend symbol(s) specified in its root, i.e. at the highest level of granularity in time and in the taxonomy of symbols. For instance, the tree in figure 3 indexes all the cases whose feature shows a global increasing trend. Each node can expand in two dimensions: a leading dimension (e.g. time) and a secondary one (e.g. symbols). In the leading dimension, the children of the current node will index cases compatible with sequences of trends obtained at the same symbol level, but at a lower time granularity; moving in the secondary dimension, the tree will index nodes partitioning the space at the same time granularity, but at a lower symbol level. The root node *I* in figure 3, for instance, is refined along the time dimension from the 4 hours granularity to the 2 hours granularity, so that the nodes *II*, *IS* and *SI* stem from it, provided that $up(I, I) = I$, $up(I, S) = I$ and $up(S, I) = I$. Moreover the root node *I* is refined in the nodes I_W and I_S according the symbol taxonomy. Indexes can be incomplete and can grow on demand, on the basis of the available data and of the most frequently issued queries. The domain expert can also define which are the leading and the secondary dimensions before generating the tree. For further details, see [21].

---

[1] In the following, for the sake of clarity, we will focus on a single time series feature.
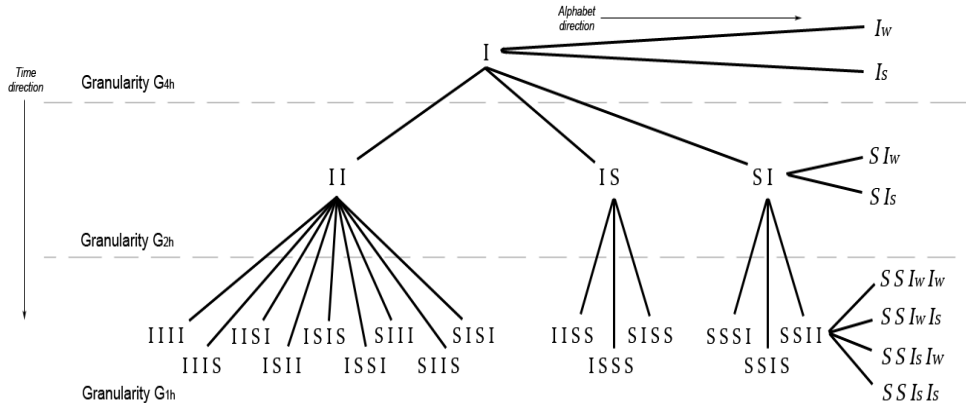
Fig. 3. Part of a multi-level orthogonal index structure

Although the use of a symbol taxonomy and/or of a temporal granularity taxonomy has been already advocated in other works (e.g. in a data warehouse context, see [34]), to the best of our knowledge our system is the first approach attempting to fully exploit the advantages of taxonomical knowledge in order to support (i) case navigation and (ii) flexible case retrieval. For what concerns navigation (i.e. case exploration), our system allows to enter a proper index in the forest and to visit it, visualizing children (in both dimensions), siblings and father of every node. In this way the users can easily analyze cases, whose features can be abstracted as identical or very similar TA strings. The retrieval task asks the user to provide a sequence of TA symbols as a query, specified at any level of detail both in time and in the symbol taxonomy. The system will then elaborate the query and retrieve all the cases compatible with the query. If results are not satisfactory, the user can interact with the system and navigate the multi-dimensional index(es), generalizing or restricting the initial query in either the time or symbol dimensions, in order to find a larger or narrower set of cases, depending on her/his current needs.

Details of query answering are illustrated in section 3 by means of specific examples in the haemodyalisis domain (for other additional technical details about the framework see also [21]).

3 Experimental results

We executed our experiments in the haemodialysis domain, using a dataset containing data belonging to 37 real patients, for a total of 1475 haemodialysis sessions (i.e. cases), collected at the Vigevano hospital, Italy. In particular, we focused our attention on the analysis of the blood pressure feature. In this section, we report on the difference between the results obtained with a classical method based on a mathematical transform (i.e. DFT), and the ones obtained using our TA-based flexible retrieval.

Diastolic pressure trends are an important indicator of the haemodialysis session performance. The desired overall trend for this feature is a slight, constant decrease from the start to the end of the session, due to the reduction of metabolites and water from blood. Clinical studies state that intradialytic increase in diastolic pressure (DP hanceforth) can complicate the management of hypertension in haemodialysis patients. Furthermore, DP increasing trend is associated with greater 2-year mortality in these patients [11]. It is therefore important to study this trend, in order to verify which patients are affected by intradialytic increases in DP, and in what haemodialysis sessions this problem arises.

To afford this task, we compared the approach described in this paper with a more classical approach we implemented in the past within the system RHENE [23]. That approach was based on DFT for dimensionality reduction, and on spatial indexing (through TV-trees) for further improving retrieval performances.

As a first consideration, it is worth highlighting the greater *expressiveness* of the TA-based approach, in which we can provide the query as a sequence of symbols describing the abstractions to be searched for, at any level of detail in both dimensions. On the other hand, mathematical methods require to provide a whole time series as a query, built to match the characteristics (e.g. trend) we are interested in. This can be a very complex operation for an unexperienced user. In RHENE we mitigated this difficulty by allowing the user to provide an existing case, whose shape basically showed the desired characteristics, as a query. As an example, we will consider the query case belonging to patient 49, in session 177, taken from our case library. The shape of this signal represents the clinical problem we are going to investigate: in particular, DP basically mantains a stable trend for the first part of the dialysis session, then it slightly increases until the end of the session. The plot is shown in figure 4.
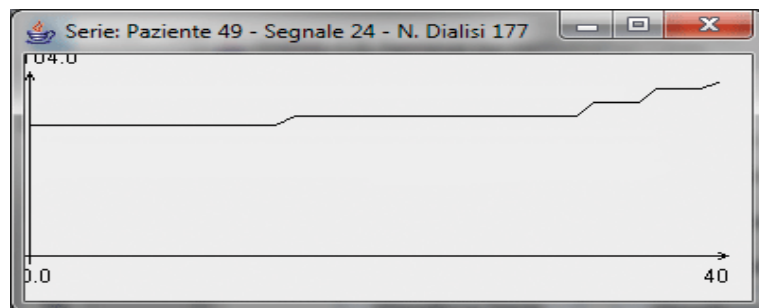

Fig. 4. The query case

The nearest 8 results obtained through DFT are shown in figure 5.

Looking at figure 5, we can see that not all the retrieved cases are compatible with the shape we were searching for. In particular, we have obtained 2 results showing the correct trends (patient 43, session 313, and patient 46, session 272);
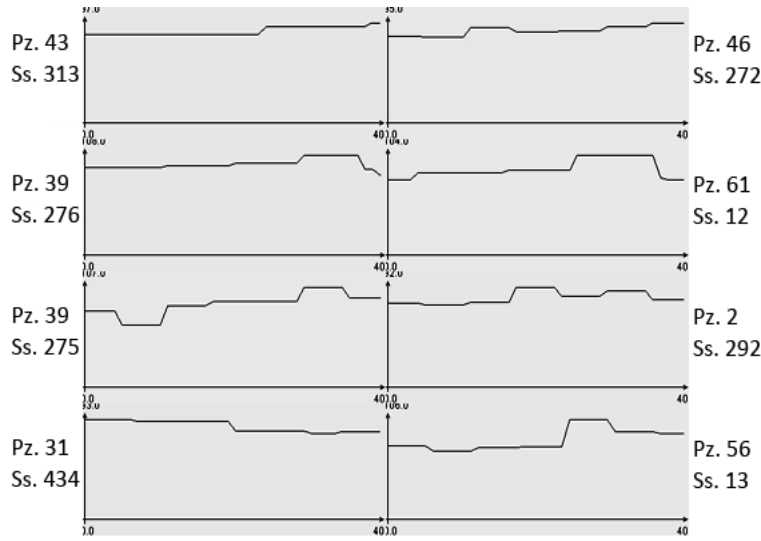
Fig. 5. The best 8 results using DFT

3 results showing an increasing trend, but with some instability (patient 39, session 276; patient 61, session 12, and patient 39, session 275), and a result which is almost stable, but with frequent episodes of increasing and decreasing trends (patient 2, session 292). One result even shows an overall decreasing trend (patient 31, session 434).

It can be noticed that the results whose trends are not looking similar to the query, are compatible with the retrieval strategy that has been adopted: using the Euclidean distance over the series points, the system retrieves cases that minimize the overall alignment around the points of the query. Considering this point of view, the result showing an overall decreasing trend is correctly retrieved, because its plot is not far from the plot of the query, comparing their values one by one. Figure 6 shows this situation.

Adopting the approach introduced in this paper, on the other hand, the query can be simply issued as a string of trend symbols. The DP problems we are looking for (taking place in the case belonging to patient 47, session 177) can be expressed as the following string: $SSI_W SSI_W I_W I_W$, if choosing a time granularity of 30 minutes. In order to answer this query, the system first progressively generalizes it in the symbol taxonomy dimension, while keeping time granularity fixed. Then, it generalizes the query in the time dimension as well. The generalization steps are shown in figure 7, and allow to select the index in figure 3 for optimizing retrieval. Following the generalization steps backwards, the user can enter the index from its root, and descend along it, until the node which fits the original query time granularity is reached. If an orthogonal index stems from this node, the user can descend along it, always following the query generalization steps backwards, until the same level of detail in the symbol tax-
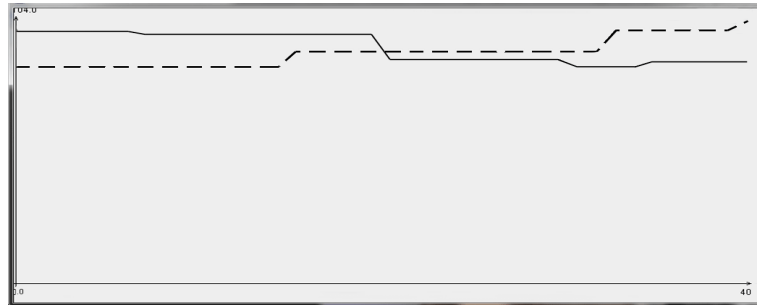
Fig. 6. Comparison between the query case (dotted line) and a result with a decreasing trend

onomy is reached, as in the original query, or at any level desired, depending on the results obtained from time to time.
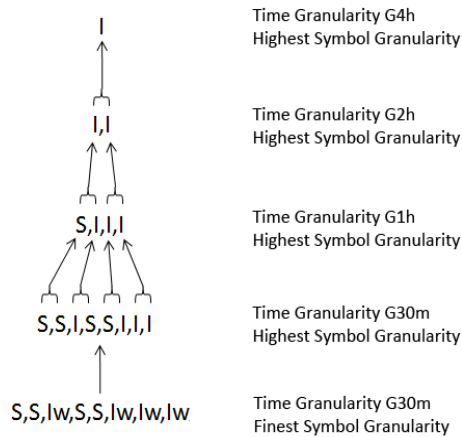

	Time Granularity G4h
I	Highest Symbol Granularity
I,I	Time Granularity G2h
	Highest Symbol Granularity
S,I,I,I	Time Granularity G1h
	Highest Symbol Granularity
S,S,I,S,S,I,I,I	Time Granularity G30m
	Highest Symbol Granularity
S,S,lw,S,S,lw,lw,lw	Time Granularity G30m
	Finest Symbol Granularity

Fig. 7. The generalization steps for the example query

In the example, 3 cases could be retrieved from node $SSI_W SSI_W I_W I_W$, which exactly matched the query. In order to enlarge the retrieval set, the user can *interactively* relax the query constraints, being guided by the index structure. In this case, we could generalize the initial requirement to a higher level of detail both in the taxonomy of symbols, and in time granularity, thus reaching node $SIII$ at 1 hour granularity level. All the cases indexed by the selected node still have a zero distance from the (relaxed) query; this is because the *up* function preserves the distance and node $SIII$ is the abstraction at 1 hour-granularity of the 30 minutes-granularity query (see details in [21]) . However, the user can also visualize the cases indexed by the siblings of this node, grouped into different

sets, depending on the distance from the query itself. The distance from the query to a sibling of this node, for example the node $SIII$, is computed using a suitable distance measure between abstraction sequences as illustrated in [21]. Figure 8 shows the 19 nearest cases retrieved, indexed by node $SIII$ (at distance=0, group A in Figure 8) and by its nearest sibling $SISI$ (at distance=1.5, group B). Please note that, among the cases retrieved at distance=0, we find the case used as the query of the DFT method (patient 49, session 177).

A) Results from the selected node: S,I,I,I

Patient	Session	TA string at ground level
49	177	S,S,S,S,S,S,S,Iw,S,S,S,S,S,S,S,Iw,S,S,S,Iw,S,S,S,Iw
2	284	Dw,S,Dw,Iw,S,S,S,S,S,Iw,S,S,Dw,S,Iw,Iw,S,S,S,S,Iw,S,M,M
41	333	S,S,S,S,S,S,S,S,Is,S,S,S,S,S,S,S,S,Is,S,S,S,S,Is,S
46	264	S,S,S,S,S,S,S,Iw,S,S,S,S,S,Iw,S,S,S,S,Iw,S,S,S,M,M

B) Results from the sibling node: S,I,S,I – distance from query: 1.5

Patient	Session	TA string at ground level
2	280	S,S,Is,Ds,S,Dw,S,S,Iw,S,S,Iw,S,S,S,Iw,S,S,S,M,M,M,M,M
27	424	S,S,S,S,S,Is,S,S,S,Iw,S,S,S,S,Iw,S,S,S,S,S,S,S,Is,M
27	442	S,S,S,S,S,S,S,S,S,S,Is,S,S,S,S,S,S,Iw,S,S,S,S,M,M
30	451	S,S,S,S,S,S,S,S,Iw,S,S,S,S,S,Iw,S,S,S,S,S,S,S,S,S
36	363	Iw,S,S,S,Ds,S,S,S,S,S,Iw,S,Iw,S,S,S,S,S,S,S,M,M,M,M
43	313	S,S,S,S,S,S,S,S,S,S,S,Iw,S,S,S,S,S,S,Iw,S,M,M,M,M
46	229	S,S,S,S,S,S,S,Iw,S,S,S,S,S,S,S,S,Is,S,S,S,S,S,M,M
46	234	S,S,S,S,S,S,Iw,S,S,S,S,S,S,S,Iw,S,S,S,M,M,M,M,M
46	250	S,S,S,S,S,S,S,Iw,S,S,S,S,S,S,S,S,S,Iw,S,S,S,S,M
48	205	S,S,S,S,S,S,S,Iw,S,S,S,S,S,S,S,S,S,Iw,S,M,M,M,M,M
48	230	S,S,S,S,S,S,S,S,S,S,S,Iw,S,S,S,S,S,Iw,S,S,S,S,M,M
48	241	S,S,S,S,S,S,Iw,S,S,S,S,S,S,S,Iw,S,S,S,M,M,M,M
53	34	S,S,S,S,S,S,Iw,S,S,S,S,S,S,S,S,S,Iw,S,S,S,S,M,M
53	55	S,S,S,S,S,S,Iw,S,S,S,S,S,Iw,S,S,S,S,M,M,M,M,M,M
57	21	S,S,S,S,S,S,S,Iw,S,S,S,S,S,S,Is,S,S,S,S,S,S,S,S,M

Fig. 8. Results using TA-based flexible retrieval

For comparison with the DFT method, we also provide the plots of 8 out of the best retrieved cases (3 at distance=0, and 5 at distance=1.5 from the query) in figure 9. All cases match the required trend. However, observe that only one case (patient 43, session 313) was also obtained by DFT, see figure 5.

From these experimental results, some observations can be drawn:

– considering the 19 best retrieved cases obtained through DFT (the number was chosen for comparison with the results in figure 8), we verified that only 1 case was retrieved by both approaches. This is reasonable, since the two methods work in a different way, giving a different semantics to the query. When adopting mathematical methods, we look for cases whose point-to-point distance from the query is globally minimized, which means that we are searching for the best overall alignment of the resulting case over the plot of the query. The main focus is on values, rather than on the shape. Using TA, instead, we focus on the shape, rather than on values. The choice of the best method to be adopted depends on the domain, and on the specific user needs. Interestingly, the integration of the two different methods in a single system would allow the user to explore the same data from different points of view. Integrating the TA-based approach in RHENE will be object of our future work;
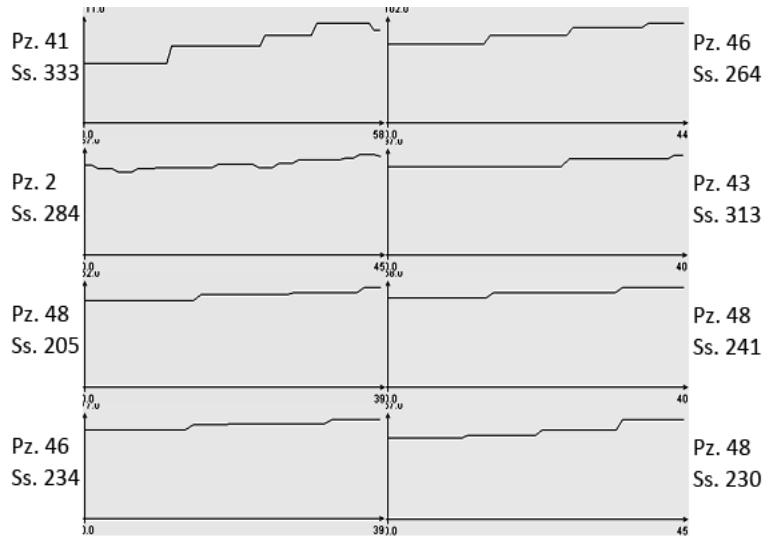
Fig. 9. Plots of the best results obtained using TA-based flexible retrieval

– considering the dynamics of the actions performed to answer a query, it is clear that mathematical methods work in a black box fashion. The user selects the query case, sets the parameters of the query (e.g. the value of K for K-Nearest Neighbour, or a maximum distance for a range query), and then the system performs its retrieval task. Using our TA-based method, instead, the user can drive the query in an interactive fashion, taking control of the retrieval task by navigating the indexing structure, to find the proper level of generalization or refinement of the query. The system is able to help the user in navigating the structure and in selecting the proper navigation direction, providing him/her with quantitative and qualitative information about the cases indexed by sons and siblings of the currently visited node (i.e. the number of indexed cases, the sequence of abstractions representing the cases and the distance from the sequence of abstractions representing the node currently visited by the user).
– the results of the query performed using our TA-based method are easily interpretable compared to the results obtained by mathematical methods. For each retrieved case, we obtain both its distance from the query, as well as its representation as a sequence of abstractions. This sequence gives an immediate understanding of the shape of the result; therefore, in most cases it is not necessary to look at the original plot.

In conclusion, the semantics underlying the two retrieval methods are different; each of them can be more or less suitable for a given search need. However, the use of TA allows the user to perform a flexible and interactive retrieval, with easy-to-understand queries and results. Furthermore, when the focus is on the shape of time series, the use of TA provides better results (see figure 9).

Our TA-based method also proved to be computationally more efficient, in answering the query, than the mathematical method in these experiments. This is because the mathematical method requires to reduce the original data using the DFT transform, to perform the query exploiting the advantages of the indexing tree, and then to post-process the results, in order to filter out the so-called "false alarms" [2]. Even if these steps produce a computationally sub-linear retrieval, the computation to be performed usually requires non neglectible time. The use of TA in the proposed multi-dimensional indexing structure allows the query to be answered by executing the generalization steps from the input TA sequence (using the *up* function), and by using them to navigate the multi-dimensional indexing structure. These steps are very fast; the retrieval performed in our experiments (using an Intel Core 2 Duo T9400 processor running at 2.53 GHz, equipped with 4 Gb of DDR2 ram) took, on average, only 41 milliseconds. On the other hand, the mathematical method required 3276 milliseconds. Since verification is in an early stage, further tests will be performed in the future, on databases of different sizes, to evaluate the quality of the results and the computational efficiency, also comparing this method with other retrieval techniques.

4 Discussion

CBR has recently been recognized as a valuable decision support technique in time-dependent applications. As a matter of fact, various CBR works dealing with cases with time series features have been recently published in different domains (e.g. robot control [28], process forecast [29], process supervision [9], pest management [6], prediction of faulty situations [13]). Applications in medicine have also been reported, relying on classical mathematical dimensionality reduction techniques, such as DFT [23] and DWT [24]. Moreover, general (e.g. logic-based) frameworks for case representation in time dependent domains have been proposed [22, 12, 18, 5].

As regards TA, they have been extensively resorted to in the literature, especially in the medical field, from diabetes mellitus [31, 3], to artificial ventilation of intensive care units patients [19](see also the survey in [33]). However, they have been typically adopted with the aim to solve a data interpretation task, and never as a flexible navigation/retrieval support facility. The goal of our proposal is to try to fill this gap, by supporting data interpretation, as well as case exploration and retrieval; this idea appears to be significantly innovative in the recent literature panorama.

As a final consideration, observe that TA are not the only methodology for reducing dimensionality by transforming a time series into a sequence of symbols. Actually a wide number of symbolic representations of time series have been introduced in the past decades (see [8] for a survey). However, some of them require a extremely specific and hard to obtain domain knowledge [14], since they a priori partition the signal into intervals, naturally provided by the underlying system dynamics, which divide the overall time period into distinct physical phases (e.g. respiration cycles in [10]). Many other approaches to sym-

bolizations are weakened by other relevant issues, like e.g. the fact that the conversion to symbolic representation requires to have access to all the data since the beginning, thus making it not exploitable in a context of data streaming. Rather interestingly, Lin [17] has introduced an alternative to TA, capable to deal with such issues, in which intervals are first obtained through PCA [15], and subsequently labeled with proper symbols. In particular this contribution allows distance measures to be defined on the symbolic approach that lower bound the corresponding distance measures defined on the original data. Such a feature permits to run some well known data mining algorithms on the transformed data, obtaining identical results with respect to operating on the original data, while gaining in efficiency. Despite these advantages, the approach in [17] is not as simple as TA, which allow a very clear interpretation of the qualitative description of the data provided by the abstraction process itself. As a matter of fact, such a description is often closer to the language of end users (e.g. of clinicians [32]), and easily adapts to domains where data values that are considered as normal at one time, or in a given context, may become dangerously abnormal in a different situation (e.g. in medicine, due to disease progression or to treatments obsolescence). And, of course, the ease and flexibility at which knowledge can be managed and understood by experts is an aspect that impacts upon the suitability and the usefulness of decision support systems in practice.

5   Conclusions

Time series interpretation and retrieval is a critical issue in all domains in which the observed phenomenon dynamics have to be dealt with, like in many medical applications. In this paper, we have described an architecture for the exploration of cases containing time-varying features, which allows time series data interpretation and summarization by means of Temporal Abstractions, a well known data interpretation AI technique.

In our system, we allow end users to issue queries at any level of detail, according to two abstraction dimensions, thus granting a significant degree of *expressiveness*. Abstracted data can also be easily navigated, and similar cases can be retrieved in a *quick*, *flexible* and *interactive* way, relying on proper orthogonal index structures, whose exploitation has been described by means of a concrete case study.

In the future, we plan to extensively test our tool, by analysing its retrieval results, and by evaluating its computational performances and its scalability on large case bases, compared to other, more classical techniques.

References

1. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations and systems approaches. AI Communications 7, 39–59 (1994)
2. Agrawal, R., Faloutsos, C., Swami, A.: Efficient similarity search in sequence databases. In: Lomet, D. (ed.) Proc. 4th Int. Conf. of Foundations of Data Organization and Algorithms. pp. 69–84. Springer-Verlag, Berlin (1993)

3. Bellazzi, R., Larizza, C., Magni, P., Montani, S., Stefanelli, M.: Intelligent analysis of clinical time series: an application in the diabetes mellitus domain. Artificial Intelligence in Medicine 20, 37–57 (2000)
4. Bellazzi, R., Larizza, C., Riva, A.: Temporal abstractions for interpreting diabetic patients monitoring data. Intelligent Data Analysis 2, 97–122 (1998)
5. Bichindaritz, I., Conlon, E.: Temporal knowledge representation and organization for case-based reasoning. In: Proc. TIME-96. pp. 152–159. IEEE Computer Society Press, Washington, DC (1996)
6. Branting, L., Hastings, J.: An empirical evaluation of model-based case matching and adaptation. In: Proc. Workshop on Case-Based Reasoning, AAAI 1994 (1994)
7. Chan, K., Fu, A.: Efficient time series matching by wavelets. In: Proc. ICDE 99. pp. 126–133. IEEE Computer Society Press, Washington, DC (1999)
8. Daw, C., Finney, C., Tracy, E.: Symbolic analysis of experimental data. Review of Scientific Instruments 2002-07-22 (2001)
9. Fuch, B., Mille, A., Chiron, B.: Operator decision aiding by adaptation of supervision strategies. In: Case-Based Reasoning Research and Development, LNAI. pp. 23–32. Springer-Verlag, Berlin (1995)
10. Funk, P., Xiong, N.: Extracting knowledge from sensor signals for case-based reasoning with longitudinal time series data. In: Perner, P. (ed.) Case-Based Reasoning in Signals and Images. pp. 247–284. Springer (2008)
11. Inrig, J.K., Patel, U.D., Toto, R.D., Szczech, L.A.: Association of blood pressure increases during hemodialysis with 2-year mortality in incident hemodialysis patients: A secondary analysis of the dialysis morbidity and mortality wave 2 study. American Journal of Kidney Diseases 54(5), 881 – 890 (2009)
12. Jaczynski, M.: A framework for the management of past experiences with time-extended situations. In: Proc. ACM conference on Information and Knowledge Management (CIKM) 1997. pp. 32–38. ACM Press, New York, NY (1997)
13. Jaere, M., Aamodt, A., Skalle, P.: Representing temporal knowledge for case-based prediction. In: Craw, S., Preece, A. (eds.) Proc. European Conference on Case Based Reasoning (ECCBR) 2002, in: Lecture Notes in Artificial Intelligence 2416. pp. 174–188. Springer-Verlag, Berlin (2002)
14. Kadar, S., Wang, J., Showalter, K.: Noise-supported travelling waves in sub-excitable media. Nature 391, 770–772 (1998)
15. Keogh, E.: Fast similarity search in the presence of longitudinal scaling in time series databases. In: Proc. Int. Conf. on Tools with Artificial Intelligence. pp. 578–584. IEEE Computer Society Press, Washington, DC (1997)
16. Keravnou, E.: Modeling medical concepts as time objects. In: Proceedings AIME 1995, LNAI 934. pp. 67–90. Springer-Verlag, Berlin (1995)
17. Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. In: Proc. of ACM-DMKD. San Diego (2003)
18. Ma, J., Knight, B.: A framework for historical case-based reasoning. In: Ashley, K., Bridge, D. (eds.) Proc. International Conference on Case Based Reasoning (ICCBR) 2003, in: Lecture Notes in Artificial Intelligence 2689. pp. 246–260. Springer-Verlag, Berlin (2003)
19. Miksch, S., Horn, W., Popow, C., Paky, F.: Utilizing temporal data abstractions for data validation and therapy planning for artificially ventilated newborn infants. Artificial Intelligence in Medicine 8, 543–576 (1996)
20. Montani, S., Bottrighi, A., Leonardi, G., Portinale, L.: A cbr-based, closed loop architecture for temporal abstractions configuration. Computational Intelligence 25(3), 235–249 (2009)

21. Montani, S., Bottrighi, A., Leonardi, G., Portinale, L., Terenziani, P.: Multi-level abstractions and multi-dimensional retrieval of cases with time series features. In: McGinty, L., Wilson, D. (eds.) Proc. International Conference on Case-Based Reasoning (ICCBR) 2009, Lecture Notes in Artificial Intelligence 5650. pp. 225–239. Springer-Verlag, Berlin (2009)
22. Montani, S., Portinale, L.: Accounting for the temporal dimension in case-based retrieval: a framework for medical applications. Computational Intelligence 22, 208–223 (2006)
23. Montani, S., Portinale, L., Leonardi, G., Bellazzi, R., Bellazzi, R.: Case-based retrieval to support the treatment of end stage renal failure patients. Artificial Intelligence in Medicine 37, 31–42 (2006)
24. Nilsson, M., Funk, P., Olsson, E., vonScheele, B., Xiong, N.: Clinical decision-support for diagnosing stress-related disorders by applying psychophysiological medical knowledge to an instance-based learning system. Artificial Intelligence in Medicine 36, 159–176 (2006)
25. Nilsson, M., Funk, P., Xiong, N.: Clinical decision support by time series classification using wavelets. In: Chen, C., Filipe, J., Seruca, I., Cordeiro, J. (eds.) Proc. Seventh International Conference on Enterprise Information Systems (ICEIS'05). pp. 169–175. INSTICC Press (2005)
26. Palma, J., Juarez, J., Campos, M., Marin, R.: A fuzzy approach to temporal model-based diagnosis for intensive care units. In: de Mantaras, R.L., Saitta, L. (eds.) Proc. European Conference on Artificial Intelligence (ECAI) 2004. pp. 868–872. IOS Press, Amsterdam (2004)
27. Portinale, L., Montani, S., Bottrighi, A., Leonardi, G., Juarez, J.: A case-based architecture for temporal abstraction configuration and processing. In: Proc. IEEE International Conference on Tools with Artificial Intelligent (ICTAI). pp. 667–674. IEEE Computer Society, Los Alamitos (2006)
28. Ram, A., Santamaria, J.: Continuous case-based reasoning. In: Proc. AAAI Case-Based Reasoning Workshop. pp. 86–93 (1993)
29. Rougegrez, S.: Similarity evaluation between observed behaviours for the prediction of processes. In: Topics in Case-Based Reasoning, LNAI 837. pp. 155–166. Springer-Verlag, Berlin (1994)
30. Shahar, Y.: A framework for knowledge-based temporal abstractions. Artificial Intelligence 90, 79–133 (1997)
31. Shahar, Y., Musen, M.: Knowledge-based temporal abstraction in clinical domains. Artificial Intelligence in Medicine 8, 267–298 (1996)
32. Stacey, M.: Knowledge based temporal abstractions within the neonatal intesive care domain. In: Proc. CSTE Innovation Conference, University of Western Sidney (2005)
33. Terenziani, P., German, E., Shahar, Y.: The temporal aspects of clinical guidelines. In: Teije, A.T., Miksch, S., Lucas, P. (eds.) Computer-based Medical Guidelines and Protocols: A Primer and Current Trends (2008)
34. Xia, B.: Similarity search in time series data sets. Tech. rep., School of Computer Science, Simon Fraser University (1997)