Case-Based Decision Support in Time Dependent Medical Domains

Stefania Montani

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

Abstract. Medical applications often require to capture the observed phenomenon dynamics, in order to properly support physicians' decision making. Case-based Reasoning (CBR), and more specifically case-based retrieval, is recently being recognized as a valuable decision support methodology in these domain. However, adopting CBR in this field is non trivial, since the need for describing the process dynamics impacts both on case representation and on the retrieval activity itself. In this work, we survey different methodologies introduced in the literature in order to implement medical CBR applications in time dependent domains, with a particular emphasis on *time series* representation and retrieval. Among the others, a novel approach, which relies on Temporal Abstractions, is analysed in depth.

1 Introduction

Several real world applications require to capture the evolution of the observed phenomenon over time, in order to describe its behaviour, and to exploit this information for future problem solving. Despite the fact that this requirement characterizes many different domains, like e.g. financial and physics ones, it appears to be particularly relevant in medical applications, where the physician typically needs to recall the clinical history that led the patient to the current condition, before prescribing a therapy. In several medical applications, (many) process features are naturally collected in the form of **time series**, either automatically generated and stored by control instruments (as e.g. in Intensive Care Unit monitoring), or obtained by listing single values extracted from temporally consecutive situations (as e.g. the series of glycated haemoglobin values, measured on a diabetic patient once every two months). Analysing long and complex time series of measurements at a screen or on paper can be tedious and prone to errors for physicians; an automated decision support strategy is therefore strongly needed in these domains.

Case-based Reasoning (CBR) [1], a reasoning paradigm that exploits the knowledge collected on previously experienced situations, known as *cases*, is recently being recognized as a valuable knowledge management and decision support methodology in time dependent applications (see e.g. [8]). Cases are composed of a problem description part, and a solution part. CBR operates by *retrieving* past cases that are similar to the current one and by *reusing* past successful solutions after, if necessary, properly *adapting* them; the current case can

then be *retained* and put into the system knowledge base, called the *case base*. *Purely retrieval* systems are also very valuable decision support tools, especially when automated adaptation strategies can hardly be identified, as in medicine. As a matter of fact, various case-based retrieval works dealing with cases with time series features have been recently published, in the medical domain (see section 2), as well as in different ones (see e.g.[7]).

However, adopting case-based retrieval can be non trivial in these applications [8]. In particular, pre-processing techniques are required in order to simplify the representation of cases with time series features, and to optimize the retrieval activity. Most of the approaches proposed in the literature to this end are founded on the common premise of **dimensionality reduction**, which allows to reduce memory occupancy, still capturing the most important characteristics of time series. Dimensionality is typically reduced by means of a **mathematical transform**, able to preserve the distance between two time series (or to underestimate it). On the other hand, our research group has recently proposed [8,3] to exploit a different technique for dimensionality reduction, namely **Temporal Abstractions** (TA) [11]. These techniques will be described in section 2. Finally, section 3 will address conclusions and future work.

2   Methods for Dimensionality Reduction

Mathematical methods. A wide literature exists about time series dimensionality reduction (see the survey in [6]). Dimensionality is typically reduced by adopting a mathematical transform that preserves the distance between two time series (or underestimates it). Widely used transforms are the Discrete Fourier Transform (DFT) [2], and the Discrete Wavelet Transform (DWT) [4].

The choice of the most cost-effective transformation to apply should be done on the basis of the application at hand. For instance, dimensionality reduction in medical CBR by resorting to DFT has been proposed in [9], where the application domain is the one of haemodialysis. In [9] only the first DFT coefficients are kept for similarity calculation, and Retrieval is implemented as a multi-step procedure. In particular, the most similar cases with respect to each one of the time series features, considered individually, are first extracted. To guarantee that retrieved cases have the required level of similarity on each time series feature, the intersection of the sets of returned cases is then computed. On the result set, global distance is finally calculated, by computing a weighted average; returned cases are ordered in terms of overall distance.

A significant application of DWT-based dimensionality reduction in medical CBR, on the other hand, can be found in [10]. The tool is able to classify Respiratory Sinus Arrhythmia (RSA), i.e. the respirations affecting the heart rate: persons do sometimes have physiological or psychological (e.g. stress-related) disorders, that appear as dysfunctions in the RSA patterns. The system uses CBR as the method for classification of dysfunctional patterns within the RSA. In the system, RSA time series are first dimensionally reduced by means of DWT; then the most similar clusters of previous cases are retrieved from the case base, in

order to support disorder classification. The authors were also able to compare the use of DWT to the one of DFT, since they originally applied DFT to the same domain; in RSA classification, they have obtained an increased retrieval rate when resorting to wavelets.

Temporal Abstractions. Mathematical methods for time series retrieval are widely accepted in the scientific community, and have been experimentally tested in several domains. Nevertheless, they have a number of limitations. For instance, they can be computationally complex, and their output is often not easily interpretable by end users (e.g. physicians). Additionally, they work well with signals with relatively simple dynamics, but they can fail to characterize more complex patterns, as the ones often encountered in medicine. The study of an alternative way to deal with time series retrieval is therefore well justified.

The idea of relying on Temporal Abstractions (TA) methods for time series dimensionality reduction and retrieval support, originally introduced by our research group [8], starts to be reported in the literature, especially dealing with medical applications (see e.g. [5,3]). TA is an Artificial Intelligence methodology able to move from a *point-based* to an *interval-based* representation of the data, where: (i) the input points are the elements of the discretized time series; (ii) the output intervals aggregate adjacent points sharing a common behavior, persistent over time. More precisely, the method described above should be referred to as *basic* TA. 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.

In particular, we are currently working at the definition of a time series retrieval framework, which allows for *multi-level abstractions*, according to two *dimensions*, namely a taxonomy of (trend or state) TA symbols, and a variety of time granularities. TA symbols can be organized in a taxonomy, in order to provide different levels of detail in the description of episodes. On the other hand, time granularities allow one to describe episodes at increasingly more abstract levels of temporal aggregation. Our framework also takes advantage of a forest of *multi-dimensional orthogonal index structures*, allowing for early pruning and focusing during the retrieval process, each of which orthogonally spans both the time and the symbol taxonomy dimensions. The root node of each index structure is represented by a string of symbols, defined at the highest level in the symbol taxonomy and in the time granularity taxonomy. An example, taking as a root the I (increase) symbol, is provided in figure 1. Here, the root node I is refined along the time dimension from the 4 hours to the 2 hours granularity, so that the nodes II, IS and SI stem from it (provided that IS and SI generalize to I when switching to a coarser time granularity level, see figure 1; S stands for stationary). Orthogonally, each node can be specialized in the symbol taxonomy dimension; for instance, SI can be specialized into $S\ I_W$ (stationary, weak increase), $S\ I_S$ (stationary, strong increase).
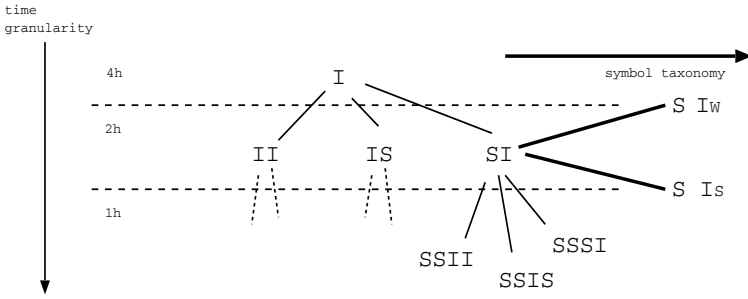
Fig. 1. An example multi-level orthogonal index structure

Users are allowed to issue queries at any level of detail. To answer the query, we first progressively generalize it in the symbol taxonomy dimension, while keeping time granularity fixed. Then, we generalize the query in the time dimension as well. Generalization allows us to identify the root of one index in the forest. Following the generalization steps backwards, we are then able to descend along the index, until we reach the node which fits the original query time granularity. If an orthogonal index stems from this node, we can descend along it as well, always following the query generalization steps backwards. We stop when we reach the same detail level in the symbol taxonomy as in the original query. If the query detail level is not represented in the index, because the index is not complete, we stop at the most detailed possible level. We then return all the cases indexed by the selected node.

Interactive and progressive query relaxation and refinement are supported as well in our framework. Query relaxation (as well as refinement) can be repeated several times, until the user is satisfied with the obtained results.

Some first experiments on the use of our framework were executed in the field of haemodialysis. In this domain, we compared our TA-based approach with the DFT-based one adopted in [9]. The experiments proved that, when focusing on the shape of time series (rather than on values), the use of TA provides more reliable results. Moreover, our TA-based method is computationally more efficient. The interested reader may find additional details about the framework and about the experiments in [3].

3   Conclusions

Time series retrieval is a critical issue in all medical domains in which the observed phenomenon dynamics have to be dealt with. Case-based Reasoning (CBR), and more specifically case-based retrieval, is recently being recognized as a valuable decision support methodology in these domains, as testified by the growing number of works in the field. In this paper, we have analyzed the main methodologies used to implement case-based retrieval as a decision support strategy in these applications. Among the others, we have described a framework

in which time series dimensionality is reduced by means of TA. The framework supports multi-level abstractions, both along the time dimension, and along a symbol taxonomy one, thus increasing the flexibility of retrieval. Query answering is interactive and is made faster by the use of orthogonal index structures. In our opinion, flexibility and interactivity represent a relevant advantage of our approach with respect to more classical techniques, in which end users are typically unable to intervene in the retrieval process, that often operates in a black-box fashion. In the future, we plan to extensively test our framework in several different application domains, thus validating its significance, and studying ways of making it more and more efficient and usable.

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.N.: 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, Berlin (1993)
3. Bottrighi, A., Leonardi, G., Montani, S., Portinale, L., Terenziani, P.: Intelligent data interpretation and case base exploration through temporal abstractions. In: Bichindaritz, I., Montani, S. (eds.) ICCBR 2010. LNCS (LNAI), vol. 6176, pp. 36–50. Springer, Heidelberg (2010)
4. Chan, K.P., Fu, A.W.C.: Efficient time series matching by wavelets. In: Proc. ICDE 1999, pp. 126–133. IEEE Computer Society Press, Washington (1999)
5. 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, Heidelberg (2008)
6. Hetland, M.L.: A survey of recent methods for efficient retrieval of similar time sequences. In: Last, M., Kandel, A., Bunke, H. (eds.) Data Mining in Time Series Databases. World Scientific, London (2003)
7. Jaere, M.D., Aamodt, A., Skalle, P.: Representing temporal knowledge for case-based prediction. In: Craw, S., Preece, A.D. (eds.) ECCBR 2002. LNCS (LNAI), vol. 2416, pp. 174–188. Springer, Heidelberg (2002)
8. Montani, S., Portinale, L.: Accounting for the temporal dimension in case-based retrieval: a framework for medical applications. Computational Intelligence 22, 208–223 (2006)
9. 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)
10. Nilsson, M., Funk, P., Xiong, N.: Clinical decision support by time series classification using wavelets. In: Chen, C.S., Filipe, J., Seruca, I., Cordeiro, J. (eds.) Proc. Seventh International Conference on Enterprise Information Systems (ICEIS 2005), pp. 169–175. INSTICC Press (2005)
11. Shahar, Y.: A framework for knowledge-based temporal abstractions. Artificial Intelligence 90, 79–133 (1997)