A Matter of Time:

Machine Learning and Temporal Data Mining

Alessio Bertone

Danube University Krems
Department of Information and Knowledge Engineering (ike)

IKE-TR-2007-01

January 2007

Authors: **Alessio Bertone**

Alessio.Bertone@donau-uni.ac.at
http://www.donau-uni.ac.at/ike

Contact: **Danube University Krems**
Department of Information and Knowledge Engineering (ike)
Dr.-Karl-Dorrek-Str. 30
3500 Krems
Austria, Europe
T +43 (2732) 893 - 2453
F +43 (2732) 893 - 4450
ike@donau-uni.ac.at
http://www.donau-uni.ac.at/ike

Table of Contents

1 Introduction

This technical report aims to investigate the problems related to the application of machine learning techniques to time oriented data and to provide an overview about the usage of data mining to find interesting patterns within temporal databases.
Despite machine learning is a very well known research field and a huge amount of references related to it can be found in the literature, mostly of them deal with machine learning applied to “static” data and a very undersized subset of them deal with “dynamic” data, that is, data whose temporal aspect is treated as a special attribute.
Temporal data are very relevant in everyday life: climatic, financial and medical data are only few examples which demonstrate the importance of the temporal aspects of data. Thus the objective of this report is to start addressing questions regarding the possibility to make feasible the application of machine learning to temporal data (treating time as a special attribute), the reasonableness of such an application and finally to provide a first version of the bibliography available on that topics.
Regarding temporal data mining a larger quantity of references is available, hence in this case the aim of this report is to briefly summarize which are the potentialities, the expectations and the promises offered by this research field.

The report is organized as follows: section 2 is devoted to the description of machine learning and time oriented data; section 3 presents a summarized state of the art on temporal data mining and finally, section 4 tries to convey a discussion about what are the potential future steps towards the application of machine learning and data mining to time oriented data.

2 Machine Learning and Time

As well known to the scientific community, machine learning is the research area within artificial intelligence that studies algorithms inducing models from a set of data. Even if a huge amount of work has been devoted to the description of this research area, most of them describe the data attribute to be taken into account as “static”. This clearly clashes against real world applications, where usually the attributes are “dynamic”, since they can vary over time and therefore time appears as really relevant.

Traditionally there have been different types of Machine Learning [Chapelle et al., 2006]: unsupervised learning, supervised learning and semi-supervised learning. The goal of unsupervised learning is to find interesting structure (patterns) in a given set of examples; given a set of pairs, the goal of unsupervised learning is to learn a mapping between the two elements of a pair (using algorithms known as generative or discriminative). Finally semi-supervised learning (SSL) represents a hybrid combination between unsupervised and supervised learning: in addition to unlabeled data, some supervised information are provided, leading to an exploitation of both labelled and unlabeled data.

As a matter of fact, since labelled instances require the efforts of experienced human annotators, they are often difficult, expensive, or time consuming to obtain. On the contrary unlabeled data may be relatively easy to collect: SSL use large amount of unlabeled data along with the labelled data in order to build better classifiers. Thus it achieves higher accuracy with less human effort and currently represents a challenge for new approaches dealing with time and machine learning.

In the following some references dealing with temporal data and the three above classification are given.

Unsupervised and supervised learning

Kadous [Kadous, 2002] presents a technique for temporal classification using machine learning: it consists of extraction and parameterization of sub-events from the training instances, which allows feature construction for a subsequent learning process.

Gonzales et al. [Gonzales et al., 2000] present a supervised classification method for temporal series, which extends inductive logic programming systems with predicates to deal with time series classification tasks.

Geurts [Geurts, 2001] proposes to extend classifiers in order to allow them to detect local shift invariant properties or patterns in time series and combine the binary test classification into decision trees by using piecewise constant modelling to increase the efficiency of search for candidate patterns.

Rüping and al. [Rüping et al., 2003] outline that the use of support vector machines for different temporal learning tasks is particularly suited for high-dimensional data.

Manganaris [Manganaris et al., 1994] [Manganaris., 1997] suggest to use Bayesian network to classify time series according to their features, starting from pre classified examples.

Hsu and al. [Hsu et al., 1998] describe a system for learning heterogeneous time series using artificial neural networks and Bayesian networks in order to alleviate the prediction task during critical events ("crisis monitoring").

Semi-supervised learning

As previously said, Semi Supervised Learning exploits both labelled and unlabeled data. There are a plenty of methods which have been proposed and can be divided into five classes according to their assumptions: SSL with generative models, low density separation, graph-based models, co-training methods and self-training methods (a complete description can be found in [Chapelle et al., 2006], [Zhou, 2005]).

In the following some references about SSL and different models are proposed.

Nigam [Nigam, 2001] describes generative model to deal with text classification tasks providing high accuracy in the results.

Bruce [Bruce, 2001] presents a Bayesian network approach to semi supervised learning, so that the parameters of a probability model are estimated using Bayesian techniques and then used to perform the classification task in the field of word-sense disambiguation.

Amini and al. [Amini, 2003] propose discriminating algorithms using both labelled and unlabeled data for text classification and text summarization tasks. Similarly Vittaut and al. [Vittaut et. al, 2002] extends these algorithms from text classification to email spam detection.

Wei and al. [Wei et. al, 2006] apply a nearest neighbour with Euclidean distance classifier (and a stopping criterion) for building time series classifiers to deal with text classifications tasks or yoga motions.

3 Temporal Data Mining

Data Mining can be defined as a step of the Knowledge Discovery process which consists of the search for patterns of interest in a particular representational form or a set of such representations, such as classification rules and trees, clustering or association rules [Fayyad et al., 1996]. The adoption of time in the data mining tasks leads to what is known as Temporal Data Mining.

To incorporate time in the data mining process gives us the ability to detect activities rather than just states, i.e. the ability to find behavioural aspects of communities of objects rather then just describing their states at a certain point in time.

Temporal data mining covers several different topics, such as the discovery of frequent, or interesting, sequences in a time series (e.g., finding customers whose spending pattern over time are similar to a given spending profile), the detection of relations between different sequences (e.g., in the stock market analysis we are interested in finding rules that describe relations between different stocks) and so on.

Temporal data mining is used in applications from various fields, such as medicine, finance, engineering and meteorology. Applications in temporal data mining often combine several different methods to reach a solution.

Even if time series analysis and temporal data mining might appear similar they mainly differ for two reasons [Laxman et al., 2006]: firstly, the size and nature of data sets they deal with. Temporal data mining methods may analyse data sets that are prohibitively large for conventional time series modelling techniques. Moreover the sequences may be nominal-valued or symbolic (rather than being real or complex-valued), and techniques to model the time series such as autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) are inapplicable. Secondly, the information that one wants to extract from the data. As a matter of fact the scope of temporal data mining overcomes the standard forecast or control applications of time series analysis. Moreover in several (temporal) data mining applications, what correlations are available and among what variables (and events) are usually unexpected, as well as hidden or unexpected trends or patterns, which are actually the ultimate goals of temporal data mining.

The discovery of relations between sequences (and sub-sequences) of events can be divided into three phases: the representation and modelling of the data sequence, the definition of similarity measures between sequences and the application of models and representations to the actual mining problems. For this division and the terms used I referred to [Antunes et al., 2001] , [Laxman et al., 2006].

Representation of Temporal Sequences

The problem of how to represent temporal sequences is essential above all when dealing with time series, as it is extremely difficult to perform efficiently a direct manipulation of continuous, high-dimensional data.

The literature provides different solutions to it: firstly, it is possible to use the data with only minimal transformation, either keeping it in its original form ([Agrawal et al., 1995], [Lin et al., 1998])or using windowing and piecewise linear approximations to obtain manageable sub-sequences ([Das et al., 1997], [Guralnik et al., 1999]). Secondly, it is possible to map the data into a more manageable space, either continuous (e.g., using a Discrete Fourier Transform [Agrawal et al., 1993] or Discrete Wavelet Transform [Chan et al., 1999]) or discrete (e.g., using ad hoc languages plus a blurry matching [Agrawal et al., 1995b], [Roddick et al., 2001], clustering [Das et al., 1998] or self organizing maps [Giles et al., 2001], [Guimarães et al., 2000]). Thirdly, it is possible to use a generative model, in which a statistical or deterministic model is obtained from the data and can be applied to answer more complex questions ([Ge et al., 2000], [Mannila et al., 1995], [Bettini et al., 1998], [Guralnik et al., 1998]).

A part from this cases, it is possible to deal with transactional databases with timing information as proposed in [Agrawal et al., 1995c], [Han et al., 2000].

Finally, all the methods should be able to discover and represent the sub-sequences of a sequence, e.g., using sliding window over the sequence to define a new sub-sequence, composed by the elements inside the window ([Faloutsos et al., 1994]).

Similarity Measures for Sequences

The second phase in the discovery of relations between sequences (and sub-sequences) of events is the definition of a similarity measure between sequences, taking also into account outliers, data noise, amplitude differences, time axis distortion problems and so on.

The literature provides different proposals: for the time-domain continuous representations, Euclidean distances ([Agrawal et al., 1995]) and Dynamic Time Warping ([Berndt et al., 1996], [Keogh et al., 1999], [Yi et al., 1998]) are most commonly adopted; in case of transformation based methods, Euclidean distances ([Agrawal et al., 1993], [Chan et al., 1999])or approaches finding relevant sub-sequences in the original data ([Faloutsos et al., 1994]) are used; in case of similarity measure in discrete space, ad hoc languages plus a blurry matching ([Agrawal et al., 1995b]) or algorithms for string editing to define a distance over the space of strings of discrete symbols ([Huang et al., 1999], [Mannila et al., 1997])can be applied. Finally for generative models, in case deterministic models, verifying if a sequence matches a given model will often provide a choice between only two possible values (yes/no), in case of stochastic models, it is possible to obtain a number indicating the probability that a given sequence was generated by a given model, without using complex similarity measures between sequences.

Mining Operations (or tasks)

One of the main data mining task is the discovery of association rules to capture correlations between attributes. In such cases, the conditional probability of the consequent occurring given the antecedent, is referred to as *confidence* of the rule and a well known algorithm to discover association rules is the apriori algorithm ([Agrawal et al., 1994]). Other methods also consider cyclic rules ([Özden et al., 1998]), that is rules that occur at regular time intervals and calendrical association rules, related to different time units ([Ramaswamy et al., 1998]).

In the task of classification each sequence have to belong to a finite and predefined set of classes, for instance in application such as speech recognition ([Juang et al., 1993], [O'Shaughnessy, 2000]), gesture recognition ([Darrell et al., 1993]), handwritten word recognition ([Kundu et al., 1988], [Tappert et al., 1990]) and so on.

Clustering of sequences or time series is another data mining task. It consists of grouping a collection of time series (or sequences) based on their similarity. Examples of clustering applications are web activity logs, financial data and so on. The fundamental problem with clustering is to find the number of clusters to represent the different sequences and initialise their parameters. However supposing to know the number of clusters along with some Markov models ([Smyth et al., 1999]) or adopting hierarchical clustering, like COBWEB ([Ketterlin et al., 1997], [Fisher et al., 1987]), to cluster temporal sequences databases, these problems can be faced with.

Finally, I would like to mention is the prediction task. Time-series prediction consists of forecasting (typically) future values of the time series based on its past samples. In order to perform this task, a predictive model for the data is needed. Different predictive models can be applied assuming the time series to stationary ([Box et al., 1994], [Chatfield, 1996], [Hastie et al., 2001]), nonstationary (like the autoregressive integrative moving average) or locally stationary (where the series is divided into smaller frames within each of which, the stationarity condition can be assumed to hold).

4 Conclusion and Discussion

This technical report investigated the current state of the art on machine learning with respect to the application to time oriented data, and temporal data mining.

The first remark which can be outlined is that even if the adoption of machine learning algorithm to deal with time oriented data seems meaningful, only few works have been devoted to this problem. However they have been tested onto ad hoc and very simple examples, the focus was into obtaining interpretable rules rather than accuracy, and above all they seemed related to the application they have been thought for, rather than to be proposed in a more general context.

More recently some proposals tried to exploit semi-supervised learning for the classification of time series: even if it is really interesting, but preliminary, I would stress some limitations of the approach. First of all, the examples presented are very simple; and the application to the classification of yoga motions seems to be too much related to the context rather than to be a wider and more general proposal and in case of application to the medical field, it seams not reasonable to train the algorithm for each patience. Secondly, the visualization proposed results understandable, but too easy and with no interaction and in case of application to e.g., medical data, the aid that could arise from it would be very limited. Thirdly, even if in the context of time series classification, the problem of dealing with multidimensional data (multiple time series) seems to be ignored. Even if some dimension reduction techniques can be applied, some relevant information could be missed during this phase.

The good idea could be to make us of the huge amount of unlabelled data as semi-supervised machine learning suggests, but within a more general context, where this process is a part of a larger framework and it is connected with further steps e.g. visualization transformation/ abstraction, so that any application (e.g., medical, financial, …) is only an instance of the framework itself and not a not extendible case study. For instance, a generative models approach for SSL (according to which the knowledge of the structure of data can be incorporated into the model itself, saving temporal and non temporal information) could be applied between the data transformation and visualization transformation phases of the TimeViz framework.

A similar approach could be used to deal with time oriented data along with temporal data mining, taking also into account during the modelling of the representation that even for sub-sequences a comparison with respect different granularities should be available and that the matching could be blurry or in case of uncertainty even fuzzy.

5 Bibliography

[Chapelle et al., 2006] Chapelle, O., Zien, A., & Sch¨olkopf, B. (Eds.). (2006c). Semi-supervised learning.MIT Press.

[Kadous, 2002] Mohammed Waleed Kadous. Temporal Classification: Extending the Classification Paradigm to Multivariate Time Series. PhD thesis, School of Computer Science & Engineering, University of New South Wales, 2002.

[Gonzales et al., 2000] Carlos Alonso González, Juan J. Rodríguez Diez: Time Series Classification by Boosting Interval Based Literals. Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial 11: 2-11 (2000)

[Geurts, 2001] Geurts, P.: Pattern extraction for time series classification. Lecture Notes in Computer Science 2168 (2001) 115–127

[Rüping et al., 2003] Rüping, S. and Morik, K. (2003), "Support Vector Machines and Learning about Time", in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)

[Manganaris et al., 1994] S. Manganaris and D. Fisher Learning Time Series for Intelligent Monitoring . Third Intl. Symp. on AI, Robotics, and Automation for Space p71-74, 1994.

[Manganaris., 1997] S. Manganaris, Supervised classification with temporal data, PhD Thesis, Vanderbilt University, 1997.

[Hsu et al., 1998] W. H. Hsu, N. D. Gettings, V. E. Lease, Y. Pan, and D. C. Wilkins. Crisis Monitoring: Methods for Heterogeneous Time Series Learning. In Proceedings of the International Workshop on Multistrategy Learning (MSL-98). Milan, Italy, June, 1998.

[Zhou, 2006] Zhou X., Semi-supervised learning literature survey http://www.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf

[Nigam, 2001] Kamal Nigam. Using Unlabeled Data to Improve Text Classification. Doctoral Dissertation, Computer Science Department, Carnegie Mellon University. Technical Report CMU-CS-01-126. 2001.

[Bruce, 2001] Bruce, R. (2001). Semi-supervised learning using prior probabilities and EM. Presented at the International Joint Conference of AI Workshop on Text Learning: Beyond Supervision, Seattle, Washington, 2001.

[Amini, 2003] Massih-Reza Amini, Patrick Gallinari: The use of unlabeled data to improve supervised learning for text summarization. SIGIR 2002: 105-112

[Vittaut et. al, 2003] Jean-Noël Vittaut, Massih-Reza Amini, Patrick Gallinari: Learning Classification with Both Labeled and Unlabeled Data. ECML 2002: 468-479

[Wei et. al, 2006] Li Wei and Eamonn Keogh (2006). Semi-Supervised Time Series Classification. In Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), pp. 748 - 753, Philadelphia, PA, U.S.A., August 20-23, 2006.

[Laxman et al., 2006] S. Laxman and P.S. Sastry, A survey of temporal data mining, *Sadhana* Vol. 31, Part 2, April 2006, pp. 173–198.

[Antunes et al., 2001] C. M. Antunes and A. L. Oliveira. *Temporal data mining: An overview*. In KDD Workshop on Temporal Data Mining, pages 1--13, San Francisco, CA, 26 August 2001.

[Fayyad et al., 1996] U. Fayyad, P. Smyth, G. Piatetetsky-Shapiro, R. Uthurusamy, *Advances in knowledge discovery and data mining*, AAAI Press/The MIT Press, 1996.

[Agrawal et al., 1995] Agrawal, R., Lin, K., Sawhney, H., Shim, K.: Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. VLDB (1995) 490-501

[Lin et al., 1998] Lin, L., Risch, T.: Querying Continuous Time Sequences. VLDB (1998) 170-181

[Das et al., 1997] Das, G., Gonopulos, D., Mannila, H.: Finding Similar Time Series. PKDD (1997) 88-100

[Guralnik et al., 1999] Guralnik. V., Srivastava. J.: Event Detection from Time Series Data. KDD (1999) 33-42

[Agrawal et al., 1995b] Agrawal, R., Giuseppe, P., Edward, W.L., Zait, M.: Querying Shapes of Histories. VLDB (1995) 502-514

[Roddick et al., 2001] Roddick, J., Spiliopoulou, M.: A Survey of Temporal Knowledge Discovery Paradigms and Methods. In IEEE Transactions of Knowledge and Data Engineering, vol. 13 (2001).

[Das et al., 1998] Das, G., Mannila, H., Smyth, P.: Rule Discovery from Time Series. KDD (1998) 16-22

[Giles et al., 2001] Giles, C., Lawrence, S., Tsoi, A.C.: Noisy Time Series Prediction using Recurrent Neural Networks and Grammatical Inference. Machine Learning, vol. 44. (2001) 161-184

[Guimarães et al., 2000] Guimarães, G.: The Induction of Temporal Grammatical Rules from Multivariate Time Series. ICGI (2000) 127-140

[Ge et al., 2000] Ge, X., Smyth, P.. Deformable Markov Model Templates for Time Series Pattern Matching. KDD (2000) 81-90

[Mannila et al., 1995] Mannila, H., Toivonen, H., Verkamo, I.: Discovering Frequent Episodes in Sequences. KDD (1995) 210-215

[Bettini et al., 1998] Bettini, C., Wang, S.X., Jagodia, S., Lin, J.L. Discovering Frequent Event Patterns with Multiple Granularities in Time Sequences. IEEE Transactions on Knowledge and Data Engineering, vol. 10, nº 2 (222-237). 1998.

[Guralnik et al., 1998] Guralnik. V., Wijesakera. D., Srivastava. J.: Pattern Directed Mining of Sequence Data. KDD (1998) 51-57

[Agrawal et al., 1995c] Agrawal, R., Srikant, R.: Mining sequential patterns. ICDE (1995) 3-14

[Han et al., 2000] Han, J., Pei, J., Yin, Y.: Mining Frequent Patterns without Candidate Generation. ACM SIGMOD Int. Conf. on Management of Data (2000) 1-12

[Faloutsos et al., 1994] Faloutsos, C., Ranganathan, M., Manolopoulos, Y.: Fast Subsequence Matching in Time-Series Databases. ACM SIGMOD Int. Conf. on Management of Data (1994) 419-429

[Berndt et al., 1996] Berndt, D., Clifford, J.: Finding Patterns in Time Series: a Dynamic Programming Approach. In: Fayyad, U., Shapiro, G., Smyth, P., Uthurusamy, R. (eds.): Advances in Knowledge Discovery and Data Mining. AAAI Press. (1996) 229-248

[Keogh et al., 1999] Keogh, E., Pazzani, M.: Scaling up Dynamic Time Warping to Massive Datasets. Proc. Principles and Practice of Knowledge Discovery in Databases (1999) 1-11

[Yi et al., 1998] Yi, B., Jagadish, H., Faloutsos, C.: Efficient Retrieval of Similar Time Sequences Under Time Warping. ICDE (1998) 201-208

[Agrawal et al., 1993] Agrawal, R., Faloutsos, C., Swami, A.: Efficient similarity search in sequence databases. FODO - Foundations of Data Organization and Algorithms Proceedings (1993) 69-84

[Chan et al., 1999] Chan, K., Fu, W.: Efficient Time Series Matching by Wavelets. ICDE (1999) 126-133

[Huang et al., 1999] Huang, Y., Yu, P.: Adaptive Query Processing for Time Series Data. KDD (1999) 282-286

[Mannila et al., 1997] Mannila, H., Ronkainen, P.: Similarity of Event Sequences. In TIME (1997). 136-139

[Agrawal et al., 1994] Agrawal R, Srikant R 1994 Fast algorithms for mining association rules in large databases. In Proc.20th Int. Conf. on Very Large Data Bases, pp 487–499

[Özden et al., 1998] Özden, B., Ramaswamy, S., Silberschatz, A.: Cyclic association rules. ICDE (1998) 412-421

[Ramaswamy et al., 1998] Ramaswamy, S., Mahajan, S., Silberschatz, A.: On the Discovery of Interesting Patterns in Association Rules. VLDB (1998) 368-379

[Juang et al., 1993] Juang B H, Rabiner L 1993 Fundamentals of speech recognition, Englewood Cliffs, NJ: Prentice Hall

[O'Shaughnessy, 2000] O'Shaughnessy D 2003 Speech communications: Human and machine, Piscataway, NJ: IEEE Press

[Darrell et al., 1993] Darrell T, Pentland A 1993 Space-time gestures. In Proc. 1993 IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition (CVPR'93), pp 335–340

[Kundu et al., 1988] Kundu A, He Y, Bahl P 1988Word recognition and word hypothesis generation for handwritten script: A Hidden Markov Model based approach. In Proc. 1988 IEEE Comput. Soc. Conf. on ComputerVision and Pattern Recognition (CVPR'88), pp 457–462

[Tappert et al., 1990] Tappert C C, Suen C Y, Wakahara T 1990 The state of the art in on-line handwriting recognition. IEEE Trans. Pattern Anal. Machine Intell. 12: 787–808

[Smyth et al., 1999] Smyth, P.: Probabilistic Model-Based Clustering of Multivariate and Sequential Data. Artificial Intelligence and Statistics (1999) 299-304

[Ketterlin et al., 1997] Ketterlin, A.: Clustering Sequences of Complex Objects. KDD (1997) 215-218

[Fisher et al., 1987] Fisher, D.. Knowledge Acquisition Via Incremental Conceptual Clustering. Machine Learning vol. 2 (1987) 139-172

[Box et al., 1994] Box G E P, Jenkins GM, Reinsel G C 1994 Time series analysis: Forecasting and control, Singapore,Pearson Education Inc.Chatfield 1996;

[Chatfield, 1996] Chatfield C 1996 The analysis of time series 5th edn., New York, Chapman and Hall

[Hastie et al., 2001] Hastie T, Tibshirani R, Friedman J 2001 The elements of statistical learning: Data mining, inference and prediction, New York, Springer-Verlag