ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Recent Trends in Learning of structured and non-standard data

Frank-Michael Schleif[1], Peter Tino[1] and Thomas Villmann[2]

1- University of Birmingham
School of Computer Science,
Edgbaston B15 2TT Birmingham,
United Kingdom.

2- University of Applied Sciences,
Department of Mathematics
Mittweida, Germany

Abstract. In many application domains data are not given in a classical vector space but occur in form of structural, sequential, relational characteristics or other non-standard formats. These data are often represented as graphs or by means of proximity matrices. Often these data sets are also huge and mathematically complicated to treat requesting for new efficient analysis algorithms which are the focus of this tutorial.

1 Introduction

Computational intelligence methods for the extraction of knowledge from large structured and non-standard data is becoming more and more important. Modern measurement systems e.g. in the life sciences and cheap storage devices facilitate the quick gathering of information at large scale. Considering projects like the human genome project or next generation sequencing, huge sets of DNA sequences are recorded which are inherently structural, waiting for analysis and knowledge extraction [4]. Many and large databases store proteins, phylogenetic trees, molecular graphs, time series data, hyper-textual and XML documents from the web all in more or less structured but clearly non-standard form often omitting a vector space representation. Some typical formats are depicted in Figure 1. Standard data sets typically consist of N samples in a real D di-

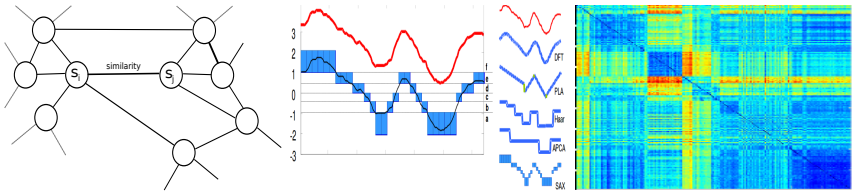

Fig. 1: Typical non-standard and structured data formats. Plot a) a graph representing relations and strength of proximity between objects Plot b) a time series and various encodings including Symbolic Aggregate approXimation. Plot c) a similarity matrix obtain from sequence scores.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

mensional vector space. Non-standard and structured data as considered in the following may deviate from this assumption in different ways. Time series data, for example often consist of vectors with different lengths but in the same D dimensional vector space. A similar problem is observed with sequential data where the N sample are typically not given in a vector space but by means of a symbol sequence of different length on a common alphabet of symbols. Also different life science data like spectral measurements are not completely standard in the way that the dimensions imply structural information and are not independent. While for example for images a vectorial representation is common it may also be better to keep the natural matrix structure to preserve neighborhood relations. Similar also video or multivolume data e.g. for functional magnetic resonance imaging (fMRI) may be better represented in tensor form. Some data are also naturally represented by means of tree or graph structures also providing a source of non-standard and structured data sets. These variations form normal data sets can be approached in different ways most often by means of an adapted representation and or a specific model and learning algorithm.

Different dedicated algorithms are already proposed for formats like timeseries, structured, sequence or matrix data [1, 2, 3, 4, 5] but still various challenges remain if it comes e.g. to large scale problems, the integration of auxiliary information, the handling of missing values or topics like semi-supervision. This tutorial provides a brief review and introduction to recent challenges and techniques in the field on *structured-* and *non-standard* data analysis.

2 Proximity matrix learning

A common strategy to process structured and non-standard data is to find a data representation such that more classical algorithms become accessible. Available similarity or dissimilarity measures for the specific formats can be used to encode the data such that a proximity matrix is obtained [6].

Considering e.g. the case of timeseries data, than such measures could be *dynamic time warping* or global alignment kernels [7]. For data like DNA or protein sequences (see Figure 2) alignment functions based on biological knowledge like the Smith-Waterman Algorithm or blast could be used [8].

A very generic measure is the Kolmogorov-Complexity which can be approximated by the normalized compression distance [9] often used to analyze textual documents. Given the obtained measure is *metric*, standard approaches like a kernel PCA, kernel k-Means, Support Vector Machine or Laplacian Eigenmap [5, 10] can be used. For non-metric measures [11] the situation is slightly more complicated because the obtained non-metric dissimilarities or indefinite kernels are less common and only few methods are available [12, 13, 14, 15, 16] or transformation methods or embedding methods are needed [6, 17].

The most common approaches to convert a non-metric similarity matrix $\mathbf{S}$, as obtained e.g. from a sequence alignment, to a positive semi definite matrix are based on eigenvalue corrections like: *flipping, clipping, vector-representation, shift correction*. The underlying idea is to remove negative eigenvalues in the eigenspectrum of the matrix $\mathbf{S}$. One may also try to learn an alternative psd

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

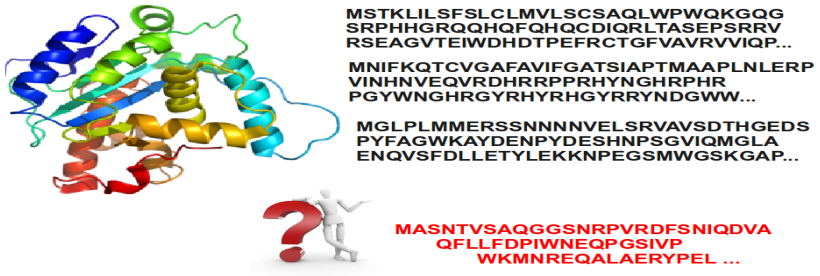

Fig. 2: Example of a protein sequence identification task. A new sequence is compared to a database of known annotated sequences using an alignment algorithm and the meta information of the predicted most similar protein sequences may provide information about the new unknown sequence.

kernel representation with maximum alignment to the original non-psd kernel matrix [18, 6, 19] or split the proximities into positive and negative eigenvalues as discussed in [20, 21], both with high computational costs.

But the appropriate encoding of the data is not the only point to care about but also the number of points. Obviously for a quadratic proximity matrix $\mathbf{S}^{N\times N}$, with N as the number of samples larger data sets e.g. $N > 10000$ become very challenging. Given the rank of the matrix is not to large efficient approximation strategies like the Nyström approximation [22] or random approximation strategies [23] can be used, which can also be used for indefinite proximity matrices [24, 25].

Most of the aforementioned approaches rely on symmetric matrices but recently also some strategies for asymmetric analysis of proximity matrices were proposed [26, 27].

3 Matrix completion and collaborative filtering

Non-standard data are also often given as an *incomplete* set of (weighted) relations between objects, which could be represented by a very sparse graph. Common examples are e.g. recommendation systems. The users of such a system (continuously) generate training data by ranking items (e.g. bought books) and other users may rank similar or the same items. The system tries to infer rules to predict items which are of interest for a user based on this knowledgebase most often employing meta information, like the fact that a specific set of items belongs to a common group. Matrix completion, collaborative filtering and low-rank matrix estimation are concepts used in this line [28, 29, 30] and are very challenging as standard settings have been shown to be NP-hard [31].

For matrix completion the data matrix Z is typically $N \times M$ with $M < N$. The rows can be e.g. customers and the columns refer to votes of articles the customers may have been bought. In general this matrix is very sparse because a customer bought only few of the M items. A typical sparsity is in the range of 99% zero entries e.g. for the famous Netflix database. The objective is now to fill

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

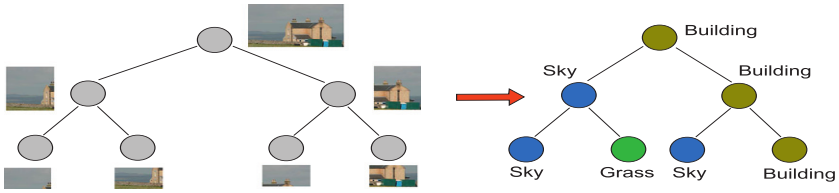

Fig. 3: Example of a tree transduction from a multi resolution image to a structural semantic representation (taken from [38])

the missing values for an individual customer to predict its potential votes. To phrase this setting as a learning problem one assumes that Z lies in a much lower dimensional manifold and can be represented e.g. as $Z \approx V_{N\times k} \cdot G_{k\times M}$ with $k \ll \min\{N, M\}$. G may be considered to be a grouping matrix with a small number of groups and $g_{lj} \in G$ as the relative score of item j in group l. The values $v_{il} \in V$ define the affinity of customer i to the group l. Accordingly the modeled score for customer i on item j is the $\sum_{l=1}^{k} V_{il} \times G_{lj}$ of group affinities times group scores. This problem can be formulated as a constrained optimization problem and different variants were proposed to make it feasible e.g. for larger datasets.

Current work in this field tries to provide more specific analysis about minimal requirements to reconstruct a low rank matrix [32] or considers alternative norms and optimization schemes to obtained better results [33]. A very recent survey on Bayesian techniques for low rank matrix estimation was provided in [29].

4 Models for structured and non-standard data

Since already more than a decade learning and mining of structured and non-standard data has becoming more and more into the focus of the research community. Starting with early work around recursive neural networks [34], successful methods like the Self-Organizing-Map for structured data (SOM-SD) and variants [35, 36, 37] or very recent approaches for tree-structured data [38, 39, 40] have successfully tackled different data analysis problems for structured data as shown in [41]. An example how to generate a structural data representation model of a complex set of input data is depicted in Figure 3 (details in [38]). Recent work on representing biological sequence data by means of effective tree kernels can be found in [44]. An efficient strategy to integrate auxiliary information for spectral data in the model process is proposed in [45] whereas an example of non-trivial metric adaptation for microarray and spectral data is given in [46].

As another source of non-standard data and models one may also consider multiway array data as a generalization of matrices [47]. Such data typically occur in the line of medical fMRI and EEG measurements but also three dimensional video sequences or sensory data in the context of robotics are of this type. The corresponding tensor based algorithms can often better use the multilinear

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

structural properties of these datasets.

Time series data, as a specific form of non-standard data, have raised continuous interest in all times [1] but with new measurement systems (e.g. long term measurements in life sciences) and data sources in the Internet (e.g. click statistics, network logging data, high-frequency trading) [42] not only the amount of data has raised but also the dimensionality and the properties like varying sampling rate, sparsity and length have changed. These challenges motivated new dedicated methods like the Generalized Topographic Model Through Time (GTM-TT) which can effectively deal with multi-dimensional timeseries [48, 49] and was also extended by metric adaptation and supervision concepts [50, 51] and more recent approaches are based on reservoir computing and kernel techniques [52].

One prominent approach for large sets of time series data is Symbolic Aggregate approXimation (SAX) [53] which was generalized in [54] to take specific data properties into account. Also the approach proposed in [55] considers time series data but improved the learning of a regression model by employing privileged information about the data. Another recent approach linking to the former section was provided in [56] where multi-dimensional time series data are mapped into a parameter space of a hidden markov model employing a discriminative fisher kernel. A semi-supervised hidden random field based approach for timeseries was recently proposed in [57].

5 Conclusions

In this tutorial we briefly reviewed challenges and approaches common in the field of non-standard and structured data analysis. The more recent proposals in these domains focus on the analysis of large scale problems and the effective integration of meta information. The different strategies strongly depend on the chosen or given data representation. For matrix data the integration of meta information may be realized by learning an aligned matrix within an optimization problem [18]. Large scale problems for matrix data are address by matrix approximation approaches [22, 24, 23] or by learning algorithms which are also effective on few available proximity data like core set techniques coupled with probabilistic sampling [58, 59]. More recently also sparse probabilistic models where proposed which do not rely on *metric* proximity data [15]. Also for structural and sequence data large scale datasets remain challenging and approximations approaches like SAX are very popular [53]. But also here the integration of meta data by learning appropriate kernel representation [36, 52], also for streaming data [60], or by explicitly optimizing alignment function during training [61] or other metric adaptation approaches [46] show promising new directions. Also the explicit analysis of tensor data by means of dedicated methods [47] gets more interest.

Acknowledgment

This work was funded by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Program (PIEF-GA-2012-327791).

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

References

[1] A. Douzal-Chouakria and C. Amblard. Classification trees for time series. *Pattern Recognition*, 45(3):1076–1091, 2012.

[2] M. Kim. Probabilistic sequence translation-alignment model for time-series classification. *IEEE Transactions on Knowledge and Data Engineering*, 26(2):426–437, 2014.

[3] A. Passerini. Kernel methods for structured data. *Intelligent Systems Reference Library*, 49:283–333, 2013.

[4] A. Gonzalez and R. Knight. Advancing analytical algorithms and pipelines for billions of microbial sequences. *Current Opinion in Biotechnology*, 23(1):64–71, 2012.

[5] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis and Discovery*. Cambridge University Press, 2004.

[6] Yihua Chen, Eric K. Garcia, Maya R. Gupta, Ali Rahimi, and Luca Cazzanti. Similarity-based classification: Concepts and algorithms. *JMLR*, 10:747–776, 2009.

[7] Marco Cuturi. Fast global alignment kernels. In Getoor and Scheffer [62], pages 929–936.

[8] Dan Gusfield. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, 1997.

[9] T. Geweniger, F.-M. Schleif, A. Hasenfuss, B. Hammer, and T. Villmann. Comparison of cluster algorithms for the analysis of text data using kolmogorov complexity. In *In Proceedings of the ICONIP 2008*, pages CD–Publication, 2008.

[10] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[11] W. Xu, R.C. Wilson, and E.R. Hancock. Determining the cause of negative dissimilarity eigenvalues. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6854 LNCS(PART 1):589–597, 2011.

[12] Elzbieta Pekalska and Bernard Haasdonk. Kernel discriminant analysis for positive definite and indefinite kernels. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(6):1017–1032, 2009.

[13] J. Yang and L. Fan. A novel indefinite kernel dimensionality reduction algorithm: Weighted generalized indefinite kernel discriminant analysis. *Neural Processing Letters*, pages 1–13, 2013.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

[14] P. Kar and P. Jain. Supervised learning with similarity functions. volume 1, pages 215–223, 2012.

[15] Huanhuan Chen, Peter Tino, and Xin Yao. Probabilistic classification vector machines. *IEEE Transactions on Neural Networks*, 20(6):901–914, 2009.

[16] H. Sun and Q. Wu. Least square regression with indefinite kernels and coefficient regularization. *Applied and Computational Harmonic Analysis*, 30(1):96–109, 2011.

[17] R.C. Wilson, E.R. Hancock, E. PÈ©kalska, and R.P.W. Duin. Spherical embeddings for non-euclidean dissimilarities. pages 1903–1910, 2010.

[18] Yihua Chen, Maya R. Gupta, and Benjamin Recht. Learning kernels from indefinite similarities. In *In Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada, June 14-18, 2009*, page 19, 2009.

[19] Wu-Jun Li, Zhihua Zhang, and Dit-Yan Yeung. Latent wishart processes for relational kernel learning. *JMLR - Proceedings Track*, 5:336–343, 2009.

[20] Elsbieta Pekalska and Bernard Haasdonk. Kernel discriminant analysis for positive definite and indefinite kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):1017–1032, 2009.

[21] E. Pekalska and R. Duin. *The dissimilarity representation for pattern recognition.* World Scientific, 2005.

[22] Mu Li, James T. Kwok, and Bao-Liang Lu. Making large-scale nyström approximation possible. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel*, pages 631–638, 2010.

[23] Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

[24] F.-M. Schleif and A. Gisbrecht. Data analysis of (non-)metric proximities at linear costs. In *Proceedings of SIMBAD 2013*, pages 59–74, 2013.

[25] F.-M. Schleif. Proximity learning for non-standard big data. In *Proceedings of ESANN 2014*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2014. D-Side Publications.

[26] H. Choi. Data visualization for asymmetric relations. *Neurocomputing*, 124:97–104, 2014.

[27] M. Strickert, Kerstin Bunte, F.-M. Schleif, and E. Huellermeier. Correlation-based neighbor embedding. *NeuroComputing*, page to appear, 2014.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

[28] S.T. Aditya, O. Dabeer, and B.K. Dey. A channel coding perspective of collaborative filtering. *IEEE Transactions on Information Theory*, 57(4):2327–2341, 2011.

[29] P. Alquier. Bayesian methods for low-rank matrix estimation: Short survey and theoretical study. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8139 LNAI:309–323, 2013.

[30] R. Mazumder, T. Hastie, and R. Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11:2287–2322, 2010.

[31] N. Gillis and F. Glineur. Low-rank matrix approximation with weights or missing data is np-hard. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1149–1165, 2011.

[32] E.J. Candes and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

[33] F. Nie, H. Wang, X. Cai, H. Huang, and C. Ding. Robust matrix completion via joint schatten p-norm and lp- norm minimization. pages 566–574, 2012.

[34] Paolo Frasconi, Marco Gori, and Alessandro Sperduti. On the efficient classification of data structures by neural networks. In *IJCAI*, pages 1066–1071. Morgan Kaufmann, 1997.

[35] Markus Hagenbuchner, Alessandro Sperduti, and Ah Chung Tsoi. A self-organizing map for adaptive processing of structured data. *IEEE Transactions on Neural Networks*, 14(3):491–505, 2003.

[36] Fabio Aiolli, Giovanni Da San Martino, Markus Hagenbuchner, and Alessandro Sperduti. Learning nonsparse kernels by self-organizing maps for structured data. *IEEE Transactions on Neural Networks*, 20(12):1938–1949, 2009.

[37] C. Ferles and A. Stafylopatis. Self-organizing hidden markov model map (sohmmm). *Neural Networks*, 48:133–147, 2013.

[38] Davide Bacciu, Alessio Micheli, and Alessandro Sperduti. An input-output hidden markov model for tree transductions. *Neurocomputing*, 112:34–46, 2013.

[39] K. Rieck, T. Krueger, U. Brefeld, and K.-R. MÃ$\frac{1}{4}$ller. Approximate tree kernels. *Journal of Machine Learning Research*, 11:555–580, 2010.

[40] G. Da San Martino, N. Navarin, and A. Sperduti. A tree-based kernel for graphs. pages 975–986, 2012.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

[41] G. Da San Martino and A. Sperduti. Mining structured data. *IEEE Computational Intelligence Magazine*, 5(1):42–49, 2010.

[42] M.J. Cafarella, A. Halevy, and J. Madhavan. Structured data on the web. *Communications of the ACM*, 54(2):72–79, 2011.

[43] L. Schietgat. Graph-based data mining for biological applications. *AI Communications*, 24(1):95–96, 2011.

[44] C. J. Bowles and J. M. Hogan. Weighted tree kernels for sequence analysis. In Michel Verleysen, editor, *Proceedings of the 22. European Symposium on Artificial Neural Networks ESANN 2014*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2014. D-Side Publications.

[45] K. Domaschke, A. Rossberg, and T. Villmann. Utilization of chemical structure information for analysis of spectra composites. In Michel Verleysen, editor, *Proceedings of the 22. European Symposium on Artificial Neural Networks ESANN 2014*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2014. D-Side Publications.

[46] M. Lange, D. ZÃ$\frac{1}{4}$hlke, O. Holz, and T. Villmann. Applications of l_p-norms and their smooth approximations for gradient based learning vector quantization. In Michel Verleysen, editor, *Proceedings of the 22. European Symposium on Artificial Neural Networks ESANN 2014*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2014. D-Side Publications.

[47] Qibin Zhao, Guoxu Zhou, Tülay Adali, Liqing Zhang, and Andrzej Cichocki. Kernelization of tensor-based models for multiway data analysis: Processing of multidimensional structured data. *IEEE Signal Process. Mag.*, 30(4):137–148, 2013.

[48] Christopher M. Bishop, Markus Svensén, and Christopher K. I. Williams. Developments of the generative topographic mapping. *Neurocomputing*, 21(1-3):203–224, 1998.

[49] Iván Olier and Alfredo Vellido. Advances in clustering and visualization of time series using gtm through time. *Neural Networks*, 21(7):904–913, 2008.

[50] Andrej Gisbrecht and Barbara Hammer. Relevance learning in generative topographic mapping. *Neurocomputing*, 74(9):1351–1358, 2011.

[51] Frank-Michael Schleif, Bassam Mokbel, Andrej Gisbrecht, Leslie Theunissen, Volker Dürr, and Barbara Hammer. Learning relevant time points for time-series data in the life sciences. In Alessandro E. P. Villa, Wlodzislaw Duch, Péter Érdi, Francesco Masulli, and Günther Palm, editors, *ICANN (2)*, volume 7553 of *Lecture Notes in Computer Science*, pages 531–539. Springer, 2012.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

[52] Huanhuan Chen, Fengzhen Tang, Peter Tino, and Xin Yao. Model-based kernel for efficient time series analysis. In Inderjit S. Dhillon, Yehuda Koren, Rayid Ghani, Ted E. Senator, Paul Bradley, Rajesh Parekh, Jingrui He, Robert L. Grossman, and Ramasamy Uthurusamy, editors, *KDD*, pages 392–400. ACM, 2013.

[53] Alessandro Camerra, Themis Palpanas, Jin Shieh, and Eamonn J. Keogh. isax 2.0: Indexing and mining one billion time series. In Geoffrey I. Webb, Bing Liu, Chengqi Zhang, Dimitrios Gunopulos, and Xindong Wu, editors, *ICDM*, pages 58–67. IEEE Computer Society, 2010.

[54] H. T. Kruitbosch, I. Giotis, and M. Biehl. Segmented shape-symbolic time series representation. In Michel Verleysen, editor, *Proceedings of the 22. European Symposium on Artificial Neural Networks ESANN 2014*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2014. D-Side Publications.

[55] F. Tang, P. Tino, P.A. Gutierrez, and H. Chen. Support vector ordinal regression using privileged information. In Michel Verleysen, editor, *Proceedings of the 22. European Symposium on Artificial Neural Networks ESANN 2014*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2014. D-Side Publications.

[56] Laurens van der Maaten. Learning discriminative fisher kernels. In Getoor and Scheffer [62], pages 217–224.

[57] M. Kim. Semi-supervised learning of hidden conditional random fields for time-series classification. *Neurocomputing*, 119:339–349, 2013.

[58] Ivor Wai-Hung Tsang, András Kocsor, and James Tin-Yau Kwok. Large-scale maximum margin discriminant analysis using core vector machines. *IEEE Transactions on Neural Networks*, 19(4):610–624, 2008.

[59] Ivor W. Tsang, András Kocsor, and James T. Kwok. Simpler core vector machines with enclosing balls. In *Machine Learning, Proceedings of the Twenty-Fourth International Conference (ICML 2007), Corvallis, Oregon, USA, June 20-24, 2007*, pages 911–918, 2007.

[60] B. Li, X. Zhu, L. Chi, and C. Zhang. Nested subtree hash kernels for large-scale graph classification over streams. pages 399–408, 2012.

[61] Bassam Mokbel, B. Paassen, and B. Hammer. Adaptive distance measures for sequential data. In Michel Verleysen, editor, *Proceedings of the 22. European Symposium on Artificial Neural Networks ESANN 2014*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2014. D-Side Publications.

[62] Lise Getoor and Tobias Scheffer, editors. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*. Omnipress, 2011.