A Multi-Objective Multi-Modal Optimization Approach for Mining Stable Spatio-Temporal Patterns.

Michèle Sebag[1], Nicolas Tarrisson[1], Olivier Teytaud[1], Julien Lefevre[2] & Sylvain Baillet[2]
(1) TAO, CNRS − INRIA − Université Paris-Sud, F-91405 - Orsay
(2) LENA, CNRS, La Pitié Salpétrière, F-75651 - Paris

Abstract

This paper, motivated by functional brain imaging applications, is interested in the discovery of stable spatio-temporal patterns. This problem is formalized as a multi-objective multi-modal optimization problem: on one hand, the target patterns must show a good stability in a wide spatio-temporal region (antagonistic objectives); on the other hand, experts are interested in finding all such patterns (global and local optima). The proposed algorithm, termed *4D-Miner*, is empirically validated on artificial and real-world datasets; it shows good performances and scalability, detecting target spatio-temporal patterns within minutes from 400+ Mo datasets.

1 Introduction

Spatio-temporal data mining is concerned with finding specific patterns in databases describing temporally situated spatial objects. Many approaches have been developed in signal processing and computer science to address such a goal, ranging from Fourier Transforms to Independent Component Analysis [Hyvarinen *et al.*, 2001], mixtures of models [Chudova *et al.*, 2003] or string kernel machines [Saunders *et al.*, 2004], to name a few. These approaches aim at particular pattern properties (e.g. independence, generativity) and/or focus on particular data characteristics (e.g. periodicity). However, in some application domains, the above properties are not relevant to extract the patterns of interest. The approach presented in this paper is motivated by such an application domain, functional brain imaging. Non invasive techniques, specifically magnetoencephalography (MEG) [Hämäläinen *et al.*, 1993], provide measures of the human brain activity with an unprecedented temporal resolution (time step 1 millisecond). This resolution enables researchers to investigate the timing of basic neural processes at the level of cell assemblies [Pantazis *et al.*, 2005]. A key task is to find out which brain areas are active and when, i.e. to detect spatio-temporal regions with highly correlated activities. It is emphasized that such regions, referred to as stable spatio-temporal patterns (STPs), are neither periodic nor necessarily independent. Currently, STPs are manually detected, which (besides being a tedious task) severely hampers the reproducibility of the results. This paper addresses the automatic detection of stable spatio-temporal patterns, i.e. maximal spatio-temporal regions with maximal correlation. This detection problem cannot be formalized as a multi-objective optimization (MOO) problem [Deb, 2001], because experts are interested in several active brain areas: an STP might be relevant though it corresponds to a smaller spatio-temporal region, with a lesser correlated activity than another STP. The proposed approach thus extends MOO along the lines of multi-modal optimization [Li *et al.*, 2002], defining a *multi-objective multi-modal optimization* framework (MoMOO). MoMOO is tackled by an evolutionary algorithm termed *4D-Miner*, using a diversity criterion to relax the multi-objective search (as opposed to diversity enforcing heuristics in MOO, e.g. [Corne *et al.*, 2000; Laumanns *et al.*, 2002]; more on this in section 2.3). To the best of our best knowledge, both the extension of evolutionary algorithms to MoMOO, and the use of multi-objective optimization within spatio-temporal data mining are new, although MOO attracts increasing attention in the domain of machine learning and knowledge discovery (see, e.g., [Ghosh and Nath, 2004; Francisci *et al.*, 2003]). Experimental validation on artificial and real-world datasets demonstrates the good scalability and performances of *4D-Miner*; sought STPs are found within minutes from medium sized datasets (450 Mo), on PC-Pentium 2.4 GHz. The paper is organized as follows. Section 2 formalizes the detection of stable spatio-temporal patterns as a multi-modal multi-objective optimization problem, and motivates the use of evolutionary algorithms [Bäck, 1995; Goldberg, 2002] for their detection. Section 3 gives an overview of *4D-Miner*. Section 4 describes the experiment goals and setting. Section 5 reports on the extensive validation of *4D-Miner* on artificial and real-world data. Section 6 discusses the approach with respect to relevant work, and the paper ends with perspectives for further research.

2 Position of the problem ; notations, criteria

2.1 Notations and definitions

Let N be the number of measure points. To the i-th measure point is attached a spatial position[1] $M_i = (x_i, y_i, z_i)$ and a

[1] MEG measure points actually belong to a 2D shape (the surface of the skull). However, the approach equally handles 2D or 3D spatio-temporal data.

temporal activity $C_i(t), t \in [1, T]$. Let $I = [t_1, t_2] \subset [1, T]$

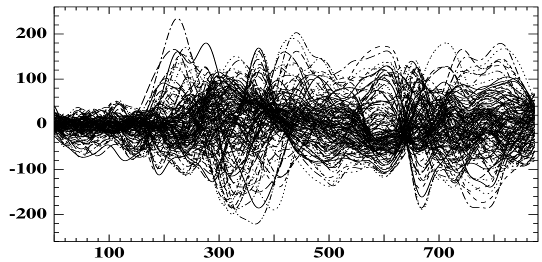

Figure 1: Magneto-Encephalography Data (N = 151, T = 875)

be a time interval, and let $\bar{C}_i^I$ denote the average activity of the i-th measure point during I. The I-alignment $\sigma_I(i,j)$ of measure points i and j over I is defined as:

$$\sigma_I(i,j) = < i,j >_I \times \left(1 - \frac{|\bar{C}_i^I - \bar{C}_j^I|}{|\bar{C}_i^I|}\right), \quad \text{where}$$

$$< i,j >_I = \frac{\sum_{t=t_1}^{t_2} C_i(t).C_j(t)}{\sqrt{\sum_{t=t_1}^{t_2} C_i(t)^2 \times \sum_{t=t_1}^{t_2} C_j(t)^2}}$$

As the sought patterns do not need to be spheric, ellipsoidal distances are considered. Only axis-parallel ellipsoids will be considered throughout the paper. For each weight vector $w = (a,b,c)$ ($w \in \mathbb{R}^{+3}$), distance d_w is defined on the measure points as:

$$d_w(i,j) = \sqrt{a(x_i - x_j)^2 + b(y_i - y_j)^2 + c(z_i - z_j)^2}$$

2.2 Multi-objective formalization

A pattern $X = \{i, I, w, r\}$ is characterized by a center point $i, i \in [1, N]$, a time interval I, an ellipsoidal distance d_w, and a radius $r > 0$. The spatial neighborhood $\mathcal{B}(i, w, r)$ is defined as the set of measure points j such that $d_w(i,j)$ is less than r.

The spatial amplitude of X, noted $a(X)$, is the number of measure points in $\mathcal{B}(i, w, r)$. The temporal amplitude of X, noted $\ell(X)$, is the number of time steps in interval I. The spatio-temporal alignment of X, noted $\sigma(X)$ is defined as the average, over all measure points in $\mathcal{B}(i, w, r)$, of their I-alignment with respect to the center i :

$$\sigma(X) = \frac{1}{a(X)} \sum_{j \in \mathcal{B}(i,w,r)} \sigma_I(i,j)$$

Interesting spatio-temporal patterns, according to the expert's first specifications, show maximal spatial and temporal amplitudes together with a maximal spatio-temporal alignment. Specifically, a solution pattern is such that i) increasing its spatial or temporal amplitude would cause the spatio-temporal alignment to decrease; ii) the only way of increasing its spatio-temporal alignment is by decreasing its spatial or temporal amplitude. It thus comes naturally to formalize the STP detection in terms of multi-objective optimization problem (see [Deb, 2001] and references therein; [de la Iglesia *et al.*, 2003; Bhattacharyya, 2000] for related data mining approaches).

The three a, ℓ and σ criteria induce a partial order on the set of patterns, referred to as Pareto domination.

Definition 1. (Pareto domination)
Let $c_1, ...c_K$ be K real-valued criteria defined on domain Ω, and let X and Y belong to Ω.
X Pareto-dominates Y*, noted $X \succ Y$, iff $c_k(X)$ is greater or equal $c_k(Y)$ for all $k = 1..K$, and the inequality is strict for at least one k_0.*

$$(X \succ Y) \iff \left\{ \begin{array}{l} \forall k = 1..K, (c_k(X) \geq c_k(Y)) \quad \textit{and} \\ \exists k_0 \; s.t. \; (c_{k_0}(X) > c_{k_0}(Y)) \end{array} \right.$$

The set of non-dominated solutions after a set of criteria is classically referred to as *Pareto front* [Deb, 2001].

2.3 Multi-modal multi-objective formalization

However, the above criteria only partially account for the expert's expectations: a STP might have a lesser spatio-temporal alignment and amplitude than another one, and still be worthy, provided that it corresponds to another active brain area. Therefore, not all sought STPs belong to the Pareto front.

Multi-modal optimization, interested in finding global *and local* optima of the fitness landscape, has been extensively studied in the EC literature [Li *et al.*, 2002]. Based on the multi-modal framework, a new framework extending multi-objective optimization and referred to as *multi-modal multi-objective optimization* (MoMOO), is thus proposed. Formally, let us first define a relaxed inclusion relationship, noted p-inclusion.

Definition 2. (p-inclusion)
Let A and B be two subsets of set Ω, and let p be a positive real number ($p \in [0,1]$). A is p-included in B, noted $A \subset_p B$, iff $|A \bigcap B| > p \times |A|$.

Defining adequately the *support* of a candidate solution X (see below) multi-modal Pareto domination can be defined as follows:

Definition 3. (multi-modal Pareto domination)
With same notations as in Def. 1, X mo-Pareto dominates Y*, noted $X \succ_{mo} Y$, iff the support of Y is p-included in that of X, and X Pareto-dominates Y.*

$$X \succ_{mo} Y \iff [(Supp(Y) \subset_p Supp(X)) \textit{ and } (X \succ Y)]$$

In the case of STPs, the support of $X = (i, I, w, r)$ is naturally defined as $Supp(X) = \mathcal{B}(i, d_w, r) \times I$, with $Supp(X) \subset [1, N] \times [1, T]$.

In contrast with [Corne *et al.*, 2000; Laumanns *et al.*, 2002] who use diversity-based heuristics for a better sampling of the Pareto front, diversity is thus used in MoMOO to redefine and extend the set of solutions.

2.4 Discussion

Functional brain imaging sets two specific requirements on spatio-temporal data mining. First, the expert's requirements are subject to change. Much background knowledge is involved in the manual extraction of stable spatio-temporal patterns, e.g. about the expected localization of the active brain areas. A flexible approach, accommodating further criteria and allowing the user to customize the search (in particular, tuning the thresholds on the minimal spatial or temporal amplitudes, or spatio-temporal alignment) is thus highly desirable.

Second, the approach must be scalable with respect to the data size (number of measure points and temporal resolution). Although the real data size is currently limited, the computational cost must be controllable in order to efficiently adjust the user-supplied parameters; in other words, the mining algorithm must be an *any-time algorithm* [Zilberstein, 1996].

The approach proposed is therefore based on evolutionary algorithms (EAs); these are widely known as stochastic, population-based optimization algorithms [Bäck, 1995; Goldberg, 2002] that are highly flexible. In particular, EAs address multi-modal optimization [Li *et al.*, 2002] and they can be harnessed to sample the whole Pareto front associated to a set of optimization criteria, with a moderate overhead cost compared to the standard approach (i.e., optimizing a weighted sum of the criteria gives a single point of the Pareto front) [Deb, 2001]. Last, the computational resources needed by EAs can be controlled in a straightforward way through limiting the number of generations and/or the population size.

3 Overview of *4D-Miner*

This section describes the *4D-Miner* algorithm designed for the detection of stable spatio-temporal patterns.

3.1 Initialization

Following [Daida, 1999], special care is devoted to the initialization of this evolutionary algorithm. The extremities of the Pareto front, where STPs display a high correlation (respectively, a low correlation) on a small spatio-temporal region (resp. a wide region), do not fulfill the expert's expectations. Accordingly, some user-supplied thresholds are set on the minimal spatio-temporal amplitude and alignment.

In order to focus the search on relevant STPs, the initial population is generated using the initialization operator, sampling patterns $X = (i, w, I, r)$ as follows:

- Center i is uniformly drawn in $[1, N]$;
- Weight vector w is set to $(1, 1, 1)$ (initial neighborhoods are based on Euclidean distance);
- Interval $I = [t_1, t_2]$ is such that t_1 is drawn with uniform distribution in $[1, T]$; the length $t_2 - t_1$ of I_j is drawn according to a Gaussian distribution $\mathcal{N}(min_\ell, min_\ell/10)$, where min_ℓ is a time length threshold[2].
- Radius r is deterministically computed from a user-supplied threshold min_σ, corresponding to the minimal I-alignment desired.

$$r = min_k\{d_w(i,k) \ s.t. \ \sigma^I_{i,k} > min_\sigma)\}$$

All patterns X whose spatial amplitude $a(X)$ is less than a user-supplied threshold min_a are non-admissible; they will not be selected to generate new offspring. The user-supplied thresholds thus govern the proportion of usable individuals in the initial population.

[2] In case t_2 is greater than T, another interval I is sampled.

The computational complexity is in $\mathcal{O}(P \times N \times min_\ell)$, where P is the population size, N is the number of measure points and min_ℓ is the average length of the intervals.

3.2 Variation operators

From parent $X = (i, w, I, r)$, mutation generates an offspring by one among the following operators:

- replacing center i with another measure point in $\mathcal{B}(i, w, r)$;
- mutating w and r using self-adaptive Gaussian mutation [Bäck, 1995];
- incrementing or decrementing the bounds of interval I;
- generating a brand new individual (using the initialization operator).

The crossover operator, starting from parent X, first selects the mate pattern $Y = (i', w', I', r')$ by tournament selection, Y minimizing the sum of the Euclidean distance between i and i', and the distance between the center of I and I' among K patterns uniformly selected in the population, where K is set to $P/10$ in all the experiments. The offspring is generated by:

- replacing i with the center i' of the mate pattern Y;
- replacing w (resp. r) using an arithmetic crossover of w and w' (respectively r and r');
- replacing I by the smallest interval containing I and I'.

An offspring is said admissible iff it satisfies the user-supplied thresholds mentioned in the initialization step.

3.3 Selection scheme

A Pareto archive is constructed with size L (set to 10 $\times P$ in the experiments).

A steady-state scheme is used; at each step, an admissible parent X is selected among K uniformly drawn individuals in the population, by retaining the one which is dominated by the fewest individuals in the archive (Pareto tournament [Deb, 2001]). A single offspring Y is generated from X by applying a variation operator among the ones mentioned above.

Offspring Y is evaluated by computing criteria $a(Y), \ell(Y), \sigma(Y)$. Y is rejected if it is mo-Pareto dominated in the population; otherwise, it replaces a non-admissible individual in the population if any; if none it replaces an individual selected after anti-Pareto tournament (the individual out of K randomly selected ones in the population, that is dominated by the most individuals in the archive).

The archive is updated every P generations, replacing the mo-Pareto dominated individuals in the archive with individuals selected from the population after Pareto tournament.

4 Experimental setting and goal

This section presents the goal of the experiments, describes the artificial and real-world datasets used, and finally gives the parameters of the algorithm and the performance measures used to evaluate the algorithm.

4.1 Goals of experiments and datasets

The initial goal is to provide the expert with a usable STP detection algorithm. The real datasets have been collected from people observing a moving ball. Each dataset involves 151 measure points and the number of time steps is 875. As can be noted from Fig. 1, which shows a representative dataset, the amplitude of the activities widely vary along the time dimension. The other goal is to assess the scalability and the performances of *4D-Miner*, which is done using artificial datasets.

The artificial datasets are generated as follows. N measure points are located in uniformly selected locations in the 3D domain $[0, 1]^3$. Activities are initialized from random cumulative uniform noise, with $C_i(t + 1) = C_i(t) + \epsilon$, and ϵ is drawn according to $U(0, \tau)$. Every target STP $S = (i, w, I, r, \delta)$ is defined from a center i, a weight vector w, a time interval I, a radius r, and a fading factor δ, used to bias the activities as detailed below. The activity C_S of the STP S is the average activity in the spatio-temporal region $\mathcal{B}(i, w, r) \times I$. Thereafter, activities are biased according to the target STPs: for each measure point j, for each STP S such that j is influenced by S ($d_w(i, j) < r$), the activity $C_j(t)$ is smoothed as

$$C_j(t) = (1 - e^{-\alpha_i(j,t)})C_j(t) + e^{-\alpha_i(j,t)}C_S$$
$$\text{where} \quad \alpha_i(j, t) = d_w(i, j) + \delta \times d(t, I)$$

and $d(t, I)$ is the distance of t to interval I (0 if t belongs to I, otherwise the minimum distance of t to the bounds of I). The scalability of *4D-Miner* is assessed from its empirical

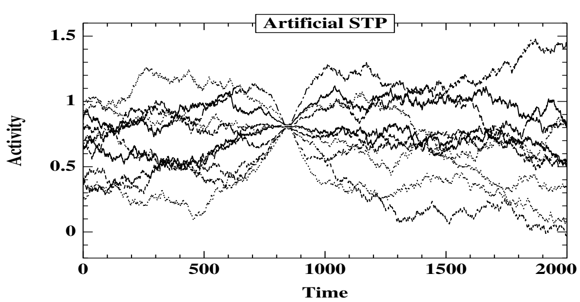

Figure 2: An artificial STP (T = 2000, N = 2000)

complexity depending on the number N of measure points and the number T of time steps.

The performances of *4D-Miner* are measured using the standard criteria in information retrieval, recall and precision. The recall is the fraction of target STPs that are identified by *4D-Miner*, i.e. p-included in an individual in the archive; the precision is the fraction of relevant individuals in the archive (p-included in a target STP). For each experimental setting (see below), the recall and precision are averaged over 21 independent runs. The number of target STPs is set to 10 in all experiments. Each STP influences a number of measure points varying in [10,20], during intervals uniformly selected in $[1, T]$ with length varying in [15, 25], and δ uniformly selected in $[0, .05]$. Each problem instance thus includes STPs with various levels of difficulty; the detection is hindered as the spatio-temporal support (r and I) is comparatively low and the δ parameter increases (the regularity is not visible outside interval I, as in Fig 2).

4.2 Experimental setting

The experiments reported in the next section considers a population size $P = 200$, evolved along 8000 generations (8,000 fitness evaluations per run). A few preliminary runs were used to adjust the operator rates; the mutation and crossover rates are respectively set to .7 and .3.

The number N of measure points ranges in $\{500, 1000, 2000, 4000\}$. The number T of time steps ranges in $\{1000, 2000, 4000, 8000\}$.

The thresholds used in the initialization are: $min_a = 5$ (minimum number of curves supporting a pattern); $min_\ell = 5$ (minimum temporal amplitude of a pattern); $min_\sigma = .1$ (minimum spatio-temporal alignment of two curves in a pattern).

For computational efficiency, the p-inclusion is computed as: X is p-included in Y if the center i of X belongs to the spatial support of Y, and there is an overlap between their time intervals.

The maximal size of the datasets ($T = 8000,\ N = 4000$) is 456 Mo. Computational runtimes are given on PC-Pentium IV, 2.4 GHz. *4D-Miner* is written in C^{++}.

5 Experimental validation

This section reports on the experiments done using *4D-Miner*.

5.1 Experiments on real datasets

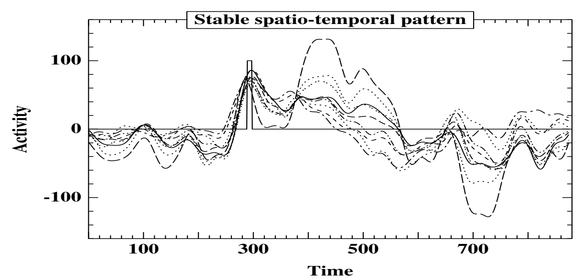

Figure 3: A stable spatio-temporal pattern, involving 8 measure points within interval [289,297], alignment .2930

Typical STPs found in the real datasets are shown in Figs. 3 and 4, showing all curves belonging to the STP plus the time-window of the pattern. The runtime is less than 1 minute (PC Pentium 1.4 GHz). As discussed in section 2.3, many relevant STPs are Pareto-dominated: typically the STP shown in Fig 3 is dominated by the one in Fig 4.

These patterns are considered satisfactory by the expert. All experiments confirm the importance of the user-defined thresholds, controlling the quality of the initial population. Due to the variability of the data, these threshold parameters are adjusted for each new experiment.

The coarse tuning of the parameters can be achieved based on the desired proportion of admissible individuals in the initial population. However, the fine-tuning of the parameters could not be automatized up to now, and it still requires running *4D-Miner* a few times. For this reason, the control of the computational cost through the population size and number of generations is one of the key features of the algorithm.

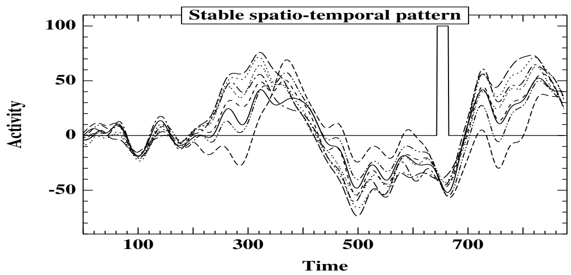

Figure 4: Another stable spatio-temporal pattern involving 9 measure points within interval [644,664], alignment .3963

5.2 Multi-objective multi-modal optimization

The good scalability of *4D-Miner* is illustrated in Fig. 5. The empirical complexity of the approach is insensitive to the number of time steps T and linear in the number N of measure points. This computational robustness confirms the analysis (section 3.1), the evaluation of each pattern has linear complexity in N. On-going work is concerned with exploiting additional data structures inspired from [Wu *et al.*, 2004] in order to decrease the complexity in N. Table 1 reports the

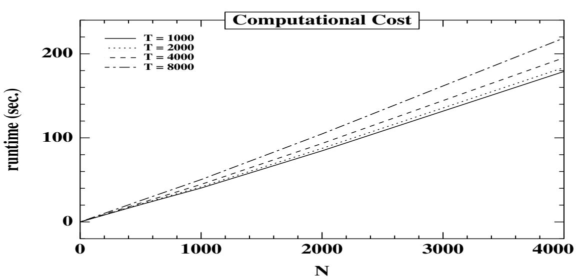

Figure 5: Computational cost *vs* N, for $T = 1000, 2000, 4000, 8000$. (in seconds, on PC 2.4 GHz)

recall achieved by *4D-Miner* over the range of experiments, averaged over 21 independent runs, together with the standard deviation. On-line performances are illustrated on Fig

N	T			
	1,000	2,000	4,000	8,000
500	98 ± 5	93 ± 9	92 ± 7	79 ± 16
1000	96 ± 6	96 ± 6	82 ± 14	67 ± 12
2000	96 ± 5	87 ± 12	72 ± 14	49 ± 15
4000	89 ± 10	81 ± 13	56 ± 14	32 ± 16

Table 1: Recall achieved by *4D-Miner* vs N and T (average percentage and standard deviation over 21 runs)

6. These results confirm the robustness of the proposed approach: a graceful degradation of the recall is observed as T and N increase. It must be noted that STPs occupy a negligible fraction of the spatio-temporal domain (circa 10^{-4} for $T = 8000, N = 4000$). The average precision is low, ranging from 12 to 20% (results omitted due to space limitations). However, post-pruning can be used to sort the wheat from the chaff in the final archive, and increase the precision up to

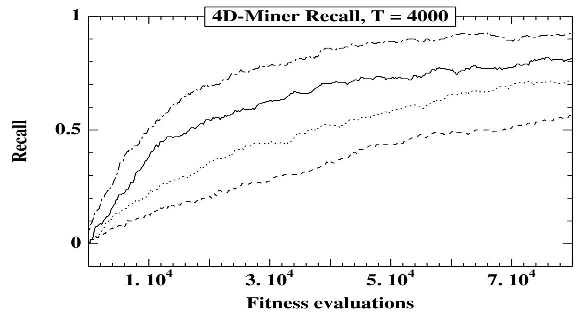

Figure 6: Recall achieved by *4D-Miner* vs the number of fitness evaluations. ($T = 4000$, $N = 500$, 1000, 2000, 4000, average over 21 runs).

50% without decreasing the recall; the post-pruning straightforwardly removes the STPs with small spatial or temporal amplitudes.
Quite different effects are obtained when the archive is pruned along the search (e.g. by increasing the thresholds on the minimal spatial and temporal amplitudes), which decreases the overall performances by an order of magnitude; interestingly, similar effects are observed in constrained evolutionary optimization when the fraction of admissible solutions is very low.

A final remark is that the performances heavily depend upon the user-supplied thresholds (section 3.1), controlling the diversity and effective size of the population. Indeed, a parametric model of the dataset would enable automatic adjustment of these parameters. It might also be viewed as a advantage of *4D-Miner* that it does not require any prior model of the data, and that inexpensive preliminary runs can be used to adjust the needed parameters.

6 Relevant work

This brief review does not claim exhaustiveness; the interested reader is referred to [Shekhar *et al.*, 2003; Roddick and Spiliopoulou, 2002] for comprehensive surveys. Spatio-temporal data mining applications (e.g., remote sensing, environmental studies, medical imaging) are known to be computationally heavy; as most standard statistical methods do not scale up properly, new techniques have been developed, including randomized variants of statistical algorithms. Many developments are targeted at efficient access primitives and/or complex data structures (see, e.g., [Wu *et al.*, 2004]); another line of research is based on visual and interactive data mining (see, e.g., [Keim *et al.*, 2004]), exploiting the unrivaled capacities of human eyes for spotting regularities in 2D-data. Spatio-temporal data-mining mostly focuses on clustering, outlier detection, denoising, and trend analysis. For instance, [Chudova *et al.*, 2003] used EM algorithms for non-parametric characterization of functional data (e.g. cyclone trajectories), with special care regarding the invariance of the models with respect to temporal translations. The main limitation of such non-parametric models, including Markov Random Fields, is their computational complexity, sidestepped by using randomized search for model estimates.

7 Discussion and Perspectives

This paper has proposed a stochastic approach for mining stable spatio-temporal patterns. Indeed, a very simple alternative would be to discretize the spatio-temporal domain and compute the correlation of the signals in each cell of the discretization grid. However, it is believed that the proposed approach presents several advantages compared to the brute force, discretization-based, alternative. First of all, *4D-Miner* is a fast and frugal algorithm; its good performances and scalability have been successfully demonstrated on medium sized artificial datasets. Second, data mining applications specifically involve two key steps, exemplified in this paper: i) understanding the expert's goals and requirements; ii) tuning the parameters involved in the specifications. With regard to both steps, the ability of working under bounded resources is a very significant advantage; any-time algorithms allow the user to check whether the process can deliver useful results, at a low cost. Further research is concerned with extending *4D-Miner* in a supervised learning perspective (finding the STPs that are complete − active in several persons undergoing the same experiment − and correct − not active in the control experiment). The challenge is to directly handle the additional constraints of completeness and correction in the multi-objective multi-modal optimization framework presented here.

Acknowledgments

The authors are partially supported by the PASCAL Network of Excellence (IST-2002-506778); this study was supported by a grant from the French Ministry of Research & Technology "ACI, New Applications of Mathematics".

References

[Bäck, 1995] T. Bäck. *Evolutionary Algorithms in theory and practice*. New-York:Oxford University Press,1995.

[Bhattacharyya, 2000] S. Bhattacharyya, Evolutionary algorithms in data mining: multi-objective performance modeling for direct marketing. In *KDD'00*, pp 465-473. ACM Press, 2000.

[Chudova *et al.*, 2003] D. Chudova, S. Gaffney, E. Mjolsness, and P. Smyth. Translation-invariant mixture models for curve clustering. In *KDD'03*, pages 79–88. ACM Press, 2003.

[Corne *et al.*, 2000] D. Corne, J. D. Knowles, and M. J. Oates. The Pareto envelope-based selection algorithm for multi-objective optimisation. In *PPSN VI*, pages 839–848. Springer Verlag, 2000.

[Daida, 1999] J.M. Daida. Challenges with verification, repeatability, and meaningful comparison in Genetic Programming. In *GECCO'99*, pages 1069–1076. Morgan Kaufmann, 1999.

[Deb, 2001] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley, 2001.

[Francisci *et al.*, 2003] D. Francisci, M. Collard, Multi-Criteria Evaluation of Interesting Dependencies according to a Data Mining Approach, In *CEC'03*, Vol. 3, pp. 1568-1574, IEEE Press, 2003.

[Ghosh and Nath, 2004] A. Ghosh and B. Nath. Multi-objective rule mining using genetic algorithms. *Inf. Sci.*, 163(1-3):123–133, 2004.

[Goldberg, 2002] D. Goldberg. *The Design of Innovation: Lessons from and for Genetic and Evolutionary Algorithms*. MIT Press, 2002.

[Hyvarinen *et al.*, 2001] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley New York, 2001.

[Hämäläinen *et al.*, 1993] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography: theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys*, 65:413–497, 1993.

[de la Iglesia *et al.*, 2003] B. de la Iglesia, M.S. Philpott, A.J. Bagnall, V.J. Rayward-Smith, Data Mining Using Multi-Objective Evolutionary Algorithms, In *CEC'03*, Vol. 3, pp. 1552-1559, IEEE Press, 2003.

[Keim *et al.*, 2004] D. A. Keim, J. Schneidewind, and M. Sips. Circleview: a new approach for visualizing time-related multidimensional data sets. In *Proc. of Advanced Visual Interfaces*, pages 179–182. ACM Press, 2004.

[Laumanns *et al.*, 2002] M. Laumanns, L. Thiele, K. Deb, and E. Zitsler. Combining convergence and diversity in evolutionary multi-objective optimization. *Evolutionary Computation*, 10(3):263–282, 2002.

[Li *et al.*, 2002] J.-P. Li, M. E. Balazs, G. T. Parks, and P. J. Clarkson. A species conserving genetic algorithm for multimodal function optimization. *Evolutionary Computation*, 10(3):207–234, 2002.

[Pantazis *et al.*, 2005] D. Pantazis, T. E. Nichols, S. Baillet, and R.M. Leahy. A comparison of random field theory and permutation methods for the statistical analysis of MEG data. *Neuroimage*, to appear in 2005.

[Roddick and Spiliopoulou, 2002] J.F. Roddick and M. Spiliopoulou. A survey of temporal knowledge discovery paradigms and methods. *IEEE Trans. on Knowledge and Data Engineering*, 14(4):750–767, 2002.

[Saunders *et al.*, 2004] C. Saunders, D. R. Hardoon, J. Shawe-Taylor, and G. Widmer. Using string kernels to identify famous performers from their playing style. In *ECML04*, pages 384–395. Springer Verlag, 2004.

[Shekhar *et al.*, 2003] S. Shekhar, P. Zhang, Y. Huang, and R. R. Vatsavai. Spatial data mining. In H. Kargupta and A. Joshi, eds, *Data Mining: Next Generation Challenges and Future Directions*. AAAI/MIT Press, 2003.

[Wu *et al.*, 2004] K.L. Wu, S.K. Chen, and P.S. Yu. Interval query indexing for efficient stream processing. In *13^{th} ACM Conf. on Information and Knowledge Management*, pages 88–97, 2004.

[Zilberstein, 1996] S. Zilberstein. Resource-bounded reasoning in intelligent systems. *Computing Surveys*, 28(4), 1996.