Functional Brain Imaging, Multi-Objective Optimisation
and Spatio-Temporal Data Mining

Michèle Sebag

TAO, Université Paris-Sud

With Nicolas Tarrisson, Olivier Teytaud, Vojtech Krmicek, UPS
Julien Lefevre, Sylvain Baillet, La Pitié-Salpétrière

Motivations

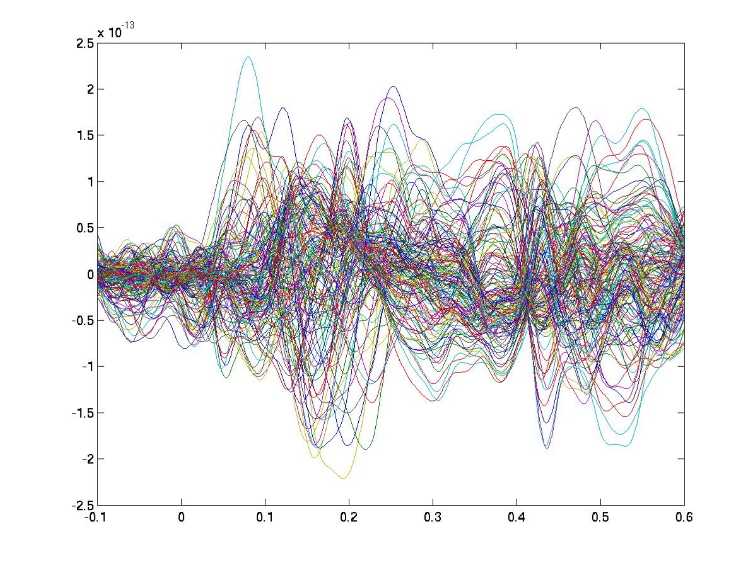
Functional Brain Imagery

- Patients, Experiments, Measures

- Magneto-Encephalography

1,000 measures *per* sensor *per* second.

The data



Spatio-temporal structure

- Sensors $i = 1..N$
- $i \quad \rightarrow \quad \begin{cases} M_i = (x_i, y_i, z_i) \in \mathbb{R}^3 \\ \{C_i[t], t = 1..T\} \in \mathbb{R}^T \end{cases}$

Overview

- Spatio-temporal Data Mining
- Multi-Objective Optimisation
- .. + Multi-modal Optimisation
- 4dMiner & Experimental Validation
- Discriminant Spatio-Temporal Patterns

Goal

Find spatio-temporal patterns

- Spatial region $A \subset \mathbb{R}^3$
- Temporal interval $I \subset \{1..T\}$

  defining

$$\mathcal{V}(A, I) = \{C_k[t],\ k\ \in\ A,\ t\ \in I\}$$

SUCH THAT

the variance of signals within $\mathcal{V}(A, I)$ is low
and $A \times I$ is a large spatio-temporal region

"active areas of the brain"

Position of the problem

In practice: done manually

- tedious
- non reproducible

Standard approach

1/ Extract a global spatio-temporal model

Independent Component Analysis — Hyvarinen et al., 2001

EM-based clustering of curves — Chudova et al., 2003

Markov Random Field — McCallum, 2004

or,

Inductive Database — Mannila et al., 1997

2/ Find specific spatio-temporal patterns from the global model

99.9 of the global model learned is useless

Data Mining

Goal

- From massive amounts of data *and knowledge*
- find novel, useful and valid knowledge

Vision

Ideally	pervasive knowledge
Actually	specialized expertise
The need	[human] knowledge management does not scale up
The opportunity	huge amounts of accessible data

Discussion

Goals

- subjective (novel & useful knowledge)
- multi-objective (valid = precise or general ?)

Requirements

- Scalability
- Flexibility

$\rightarrow$ tunable

$\rightarrow$ calibrated

$\rightarrow$ computational cost must be controllable

any-time algorithm

Zilberstein 98

MEG Mining: Multi-objective optimisation

Search space: Stable Spatio-Temporal Patterns

$$X = \begin{cases} I = [t_1, t_2] & \text{temporal interval} \\ i & \text{center of the spatial region} \\ r & \text{radius of the spatial region} \\ d_w = (a, b, c) & \text{distance weights} \quad \textit{ellipsoidal regions} \end{cases}$$

Objectives

- Temporal length $\ell(X) = t_2 - t_1$
- Spatial area $a(X) = |\mathcal{V}(X)| = |\{j \;/\; d_w(i, j) < r\}|$
- Spatio-temporal alignment

$$\sigma(X) = \frac{1}{\ell(X) \times a(X)} \sum_{j \in \mathcal{V}(X)} \sigma_I(i, j)$$

$\sigma_I(i,j)$: I-alignment of sensors i and j

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

with

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

$$\bar{C}_j^I = \text{Average}\,\{C_j[t],\; t \in I\}$$

Multi-objective Optimisation

$$\text{Find } ArgMax\{\mathcal{F}_i,\ i = 1, 2\ldots, \mathcal{F}_i : \Omega \to \mathbb{R}\}$$

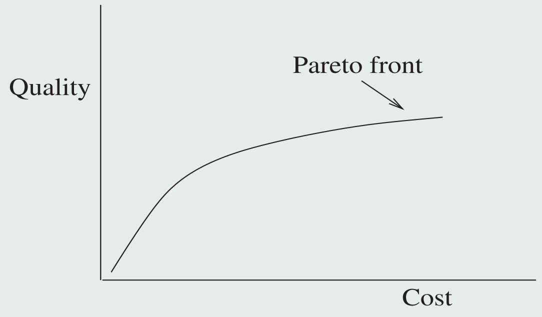
Pareto domination

- $x < y$ iff $\begin{cases} \forall i, & \mathcal{F}_i(x) \leq \mathcal{F}_i(y) \\ \exists i_0 & \mathcal{F}_{i_0}(x) < \mathcal{F}_{i_0}(y) \end{cases}$

Pareto Front

- Set of non dominated solutions.

objects with highest quality and smallest cost...



An Any-time Approach

Spatio-temporal data mining

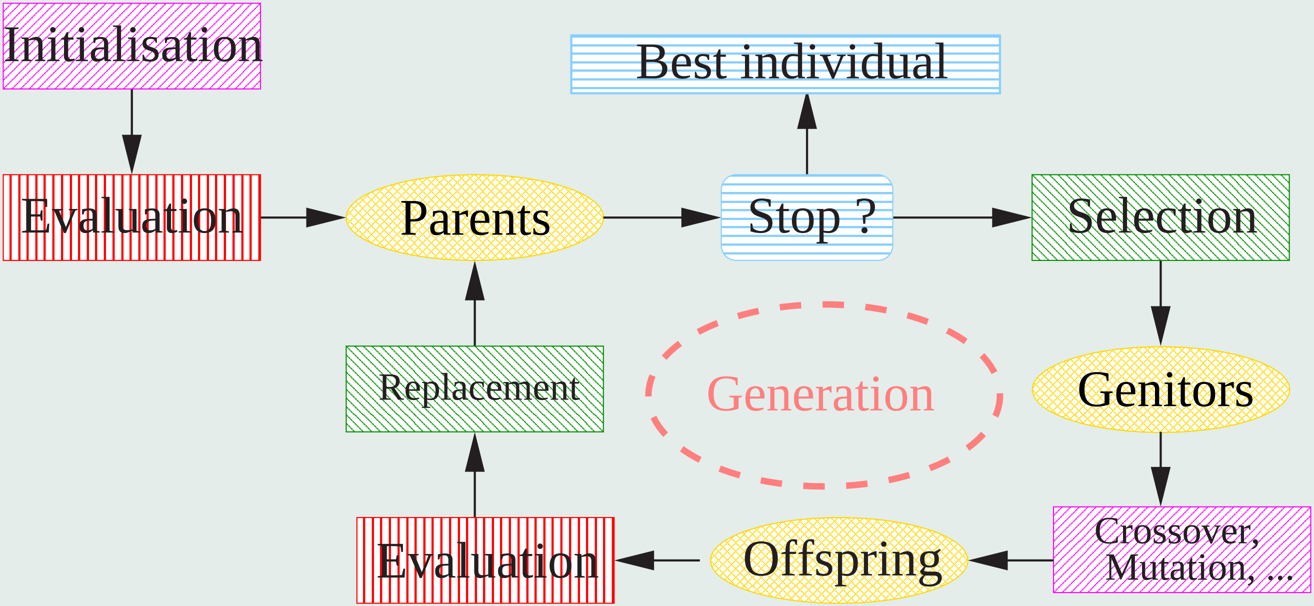
- Monotonous criteria (variance increases with I and A)
- Antagonistic criteria (decrease I or A to keep the variance low)

Complete vs Stochastic Search

- For experts to look at
- Agregate the results of several runs

Evolutionary Computation, the skeletton

$$\mathcal{F} : \Omega \mapsto \mathbb{R}$$



Multi-objective evolutionary optimisation

Standard EC — Find $ArgOpt(\mathcal{F})$

- Initialisation
- Selection
- Variation (crossover, mutation)

Multi-objective EC — Find $ArgOpt\{\ell(X), a(X), \sigma(X)\}$

Differences

Goal: sample the Pareto front — Archive

Selection after $\mathcal{F}'(X)$, measuring:

The Pareto rank of X in the current population

The percentage of the archive dominated by X

...

4d Miner: Multi-Objective Evolutionary Algorithm

Components

- Search space $\Omega \subseteq \mathbb{N}^3 \times \mathbb{R}^4$

$$\{X = (I, i, r, w),\ I \subset [1, T],\ i \in [1, N],\ r \in \mathbb{R},\ w \in \mathbb{R}^3\}$$

- Objectives a, ℓ, σ
- Operators
  - Initialisation sampling mechanism
    average interval length
    minimal acceptable alignement
    minimal pattern size
  - Variation operators

Sampling mechanism $X = (i, w, I, r)$

Daida 1999

- i : uniformly drawn in $[1, N]$;
- $w = (1, 1, 1)$ *initial = Euclidean*
- $I =$
  - t_1 : uniformly drawn in $[1, T]$
  - $\ell(I)$ drawn $\sim \mathcal{N}(min_\ell, min_\ell/10)$ min_ℓ *user supplied*
  - reject if $t_1 + \ell(I) > T$
- r : such that the ball contains all neighbors with bounded I-alignment: min_σ *user-supplied*

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

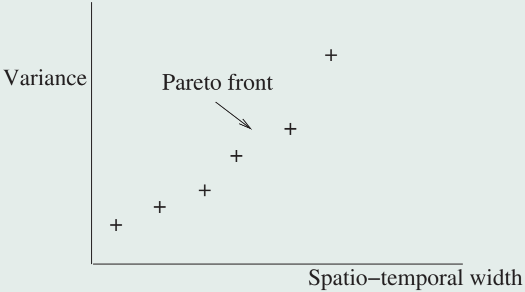
- reject if $a(X) = |\mathcal{B}(i, r)| < min_a$ min_a *user supplied*

Complexity: $\mathcal{O}(N \log N \times min_\ell)$

First results: Failure!

Diversity of stable spatio temporal patterns

- seems OK...



...but all patterns represent the same spatio-temporal region...

Failure analysis

- Experts are not interested in the Pareto front only.
- ... but in ALL active areas of the brain...

Overview

- Spatio-temporal Data Mining
- Multi-Objective Optimisation
- .. + Multi-modal Optimisation
- Experimental Validation
- Discussion

Multi-modal optimisation

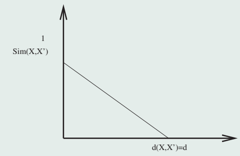
Goal

FIND ALL global (and local) optima.

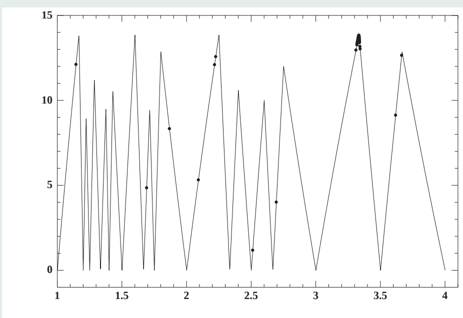
Multi-modal heuristics in EC

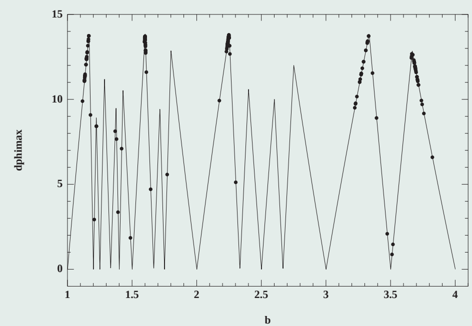
since Mahfoud, 95

Niching or **Fitness sharing**



Selection based on $\mathcal{F}'(X) = \dfrac{\mathcal{F}(X)}{\Sigma_{X'} sim(X,X')}$





Multi-objective multi-modal optimisation

Sharing-Pareto dominance

$X = (I, i, r, w)$ sp-dominates $Y = (I', i', r', w')$ iff

- X Pareto dominates Y wrt a, ℓ, σ
- X and Y overlap $\mathcal{V}(X) \cap \mathcal{V}(Y) \neq \emptyset$

$I \cap I' \neq \emptyset$

Experimental validation

Goals of experiment

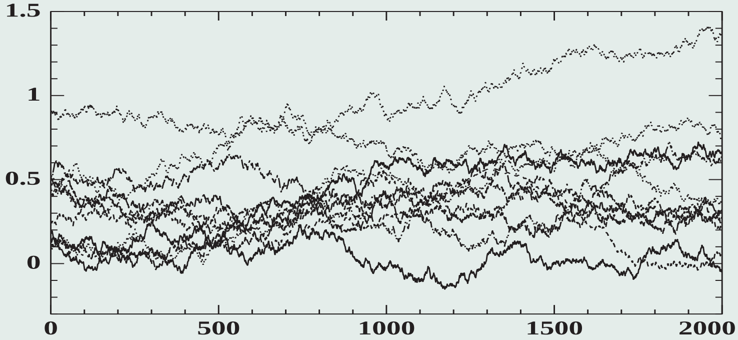
- Usability
- Scalability
- Performance / Recall

Datasets

- La Pitié-Salpétrière
- Artificial datasets

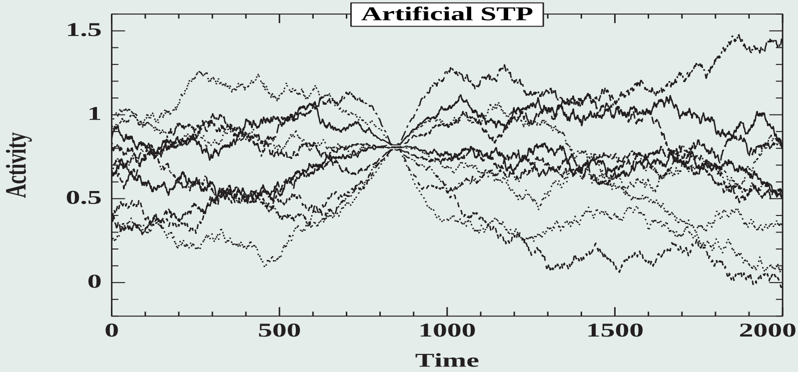
Artificial Datasets

N: number of sensors: $\{500 \mathinner{..} 4000\}$
T: number of time steps: $\{1000 \mathinner{..} 8000\}$



Artificial Datasets, 2

Draw 10 spatio-temporal patterns
Bias signals accordingly

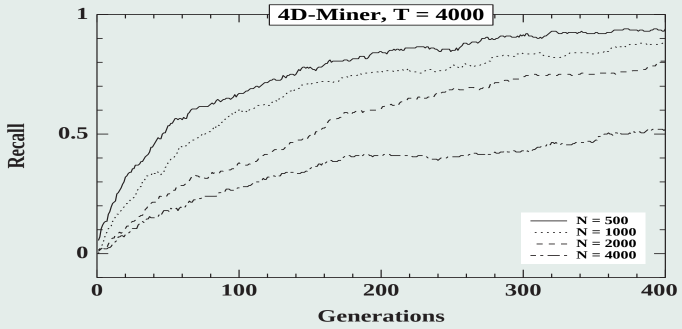


Performances

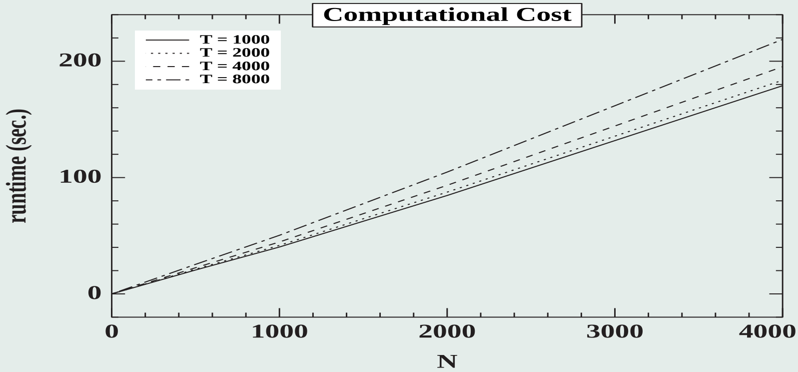
Recall : percentage of target patterns with representants in the archive.

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

Online Performance



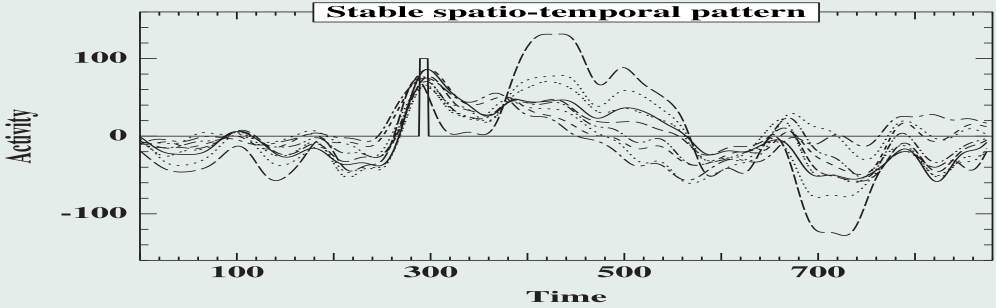
Results, cont'd

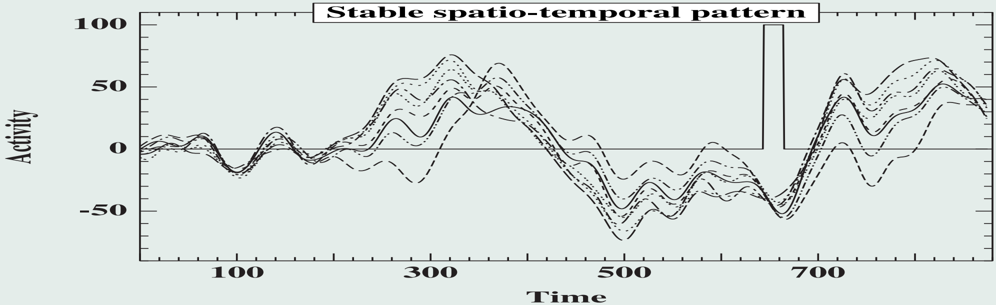


Limitations

Very sensitive to the initialisation parameters

Functional Brain Imagery





Discriminant Spatio-Temporal Patterns

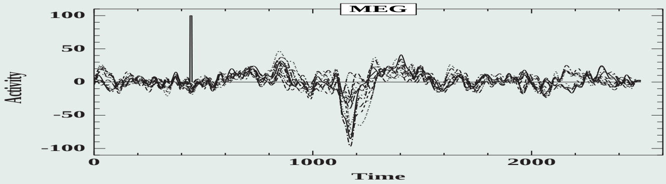
Experimental setting

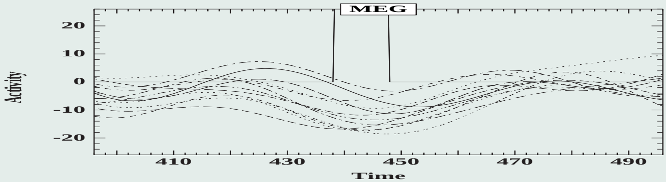
- A single person
- Setting 1 : sees a ball and let it go
- Setting 2 : sees a ball and catches it

Goal

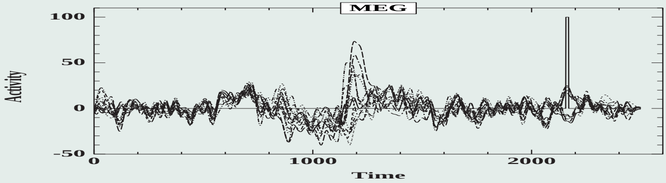
- Find STPs with different activities in Setting 1 and Setting 2 (should be related to motor skills)

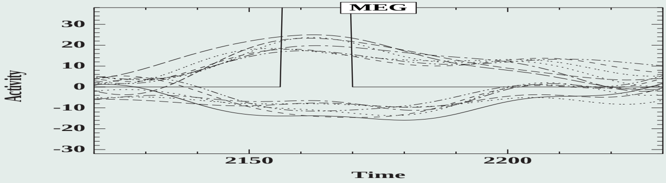
Discriminant Spatio-Temporal Patterns ?





Discriminant Spatio-Temporal Patterns ?





Discussion

Discriminative learning

- Given H, hypothesis space
- Find h, discriminating positive and negative examples.

Generative learning

- build $\mathcal{D}_+$, $\mathcal{D}_-$ distribution of positive / negative examples
- Example X is positive iff $P_{\mathcal{D}_+}(X) > P_{\mathcal{D}_-}(X)$

Discussion, 2

Remark
- Generative learning more demanding
- But often more efficient

Why ?
- Easier to incorporate prior knowledge...

Discriminant STPs : a generative approach
- Find relevant hypotheses (STPs)
- Sort the discriminant ones

Contributions: Spatio-temporal data mining

Based on multi-objective optimization

as opposed to, constraints Mannila Toivonen 97

An any-time algorithm

controllable cost $\rightarrow$ effective flexibility

Perspectives

Convergence

Type I and Type II errors

Pruning

a posteriori, increases the precision
(a priori, kills the recall...)

Functional brain imagery and variability

among patients; among trials

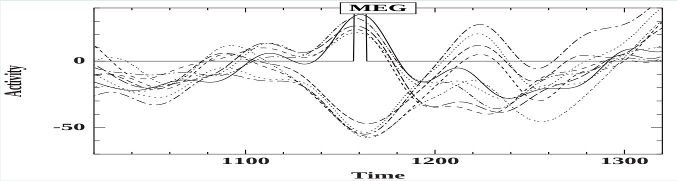
Activation scenarios

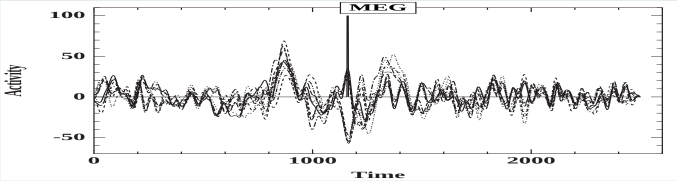
The “grammar” of cell assemblies activity

Learn the user’s criteria

Interactive optimization

Discriminant Spatio-Temporal Patterns





Sampling mechanism $X = (i, w, I, r)$

- i : uniformly drawn in $[1, N]$;
- $w = (1, 1, 1)$ *initial = Euclidean*
- $I =$
  - t_1 : uniformly drawn in $[1, T]$
  - $\ell(I)$ drawn $\sim \mathcal{N}(min_\ell, min_\ell/10)$ min_ℓ *user supplied*
  - reject if $t_1 + \ell(I) > T$
- r : such that the ball contains all neighbors with bounded I-alignment:

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

- reject if $a(X) = |\mathcal{B}(i, r)| < min_a$ min_a *user supplied*

Complexity: $\mathcal{O}(N \log N \times min_\ell)$

Variation operators

Mutation $X = (i, w, I, r)$

- Self adaptive mutation of w, r
- Specific mutation operators for i and I.
- Random (initialisation operator)

Crossover $X = (i, w, I, r) \times Y = (i', w', I', r')$

- Restricted mating:

if the spatio-temporal areas are “close enough” *user-supplied*

Selection : Pareto Archive

Steady state

- In each step, select an individual (tournament wrt Archive)
- Apply crossover or mutation
- Evaluate
- If non dominated in the population, store :
- Replace an individual (anti-selection)

Artificial datasets

Curves

$N = 500, ..4\,000$ nb sensors

$T = 1\,000, ..8\,000$ nb time steps

For $i = 1..N$

For $t = 1..T$

$C_i(t) = C_i(t-1) + \epsilon * \pm 1$

10 Target patterns

P = (i in $1..N$; $I \subset [1, T]$; $w \in \mathbb{R}^3$; $r \in \mathbb{R}$).

C_P = average of $\{C_j(t), t \in I, d_w(i,j) < r\}$

Action

$$C_j(t) = (1-\alpha)C_j(t) + \alpha C_P \times exp(-d(t, I) - d(i, j))$$