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 $i = 1..N$
- $i \rightarrow \begin{cases} M_i = (x_i, y_i, z_i) & \in \mathbb{R}^3 \\ \{C_i[t], t = 1..T\} & \in \mathbb{R}^T \end{cases}$
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
   EM-based clustering of curves
   Markov Random Field
   or,
   Inductive Database
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
  → tunable
  → calibrated
  → computational cost must be controllable
any-time algorithm

Zilberstein 98
MEG Mining: Multi-objective optimisation

Search space: Stable Spatio-Temporal Patterns

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

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 \( j \)

\[
\sigma_I(i, j) = \langle i, j \rangle_I \times \left(1 - \frac{|\bar{C}^I_i - \bar{C}^I_j|}{|\bar{C}^I_i|}\right)
\]

with

\[
\langle i, j \rangle_I = \frac{\sum_{t=t_1}^{t_2} C_i(t) \cdot C_j(t)}{\sqrt{\sum_{t=t_1}^{t_2} C_i(t)^2} \times \sum_{t=t_1}^{t_2} C_j(t)^2}
\]

\( \bar{C}^I_j = \text{Average } \{ C_j[t], t \in I \} \)
Multi-objective Optimisation

Find \( \text{Arg Max}\{\mathcal{F}_i, \ i = 1, 2\ldots, \mathcal{F}_i : \Omega \rightarrow \mathbb{R}\} \)

**Pareto domination**

- \( x < y \) iff \( \exists i_0 \ F_{i_0}(x) < F_{i_0}(y) \)

**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
- Aggregate the results of several runs

Multi-Objective Evolutionary Computation

Kalyan Deb 2001
Evolutionary Computation, the skeleton

\[ F : \Omega \rightarrow \mathbb{R} \]
Multi-objective evolutionary optimisation

Standard EC
• Initialisation
• Selection
• Variation (crossover, mutation)

Multi-objective EC

Differences
• Goal: sample the Pareto front
• Selection after $\mathcal{F}'(X)$, measuring:
  The Pareto rank of $X$ in the current population
  The percentage of the archive dominated by $X$

Find $\operatorname{ArgOpt}(\mathcal{F})$

Find $\operatorname{ArgOpt}\{\ell(X), a(X), \sigma(X)\}$

Archive
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 $a, \ell, \sigma$

- Operators
  - Initialisation sampling mechanism
    - average interval length
    - minimal acceptable alignment
    - minimal pattern size
  - Variation operators
Sampling mechanism $X = (i, w, I, r)$

- $i$: uniformly drawn in $[1, N]$;

- $w = (1, 1, 1)$ \hspace{1cm} \text{initial} = \text{Euclidean}$

- $I =$
  - $t_1$: uniformly drawn in $[1, T]$
  - $\ell(I)$ drawn $\sim \mathcal{N}(\text{min}_\ell, \text{min}_\ell/10)$ \hspace{1cm} \text{min}_\ell \text{ user supplied} \hspace{1cm}$$
  - \text{reject if } t_1 + \ell(I) > T$

- $r$: such that the ball contains all neighbors with bounded $I$-alignment:
  \hspace{1cm} \text{min}_\sigma \text{ user-supplied}$

  $$r = \text{min}_k \{d_w(i, k) \text{ s.t. } \sigma^I_{i,k} > \text{min}_\sigma\}$$

- \text{reject if } a(X) = |\mathcal{B}(i, r)| < \text{min}_a \hspace{1cm} \text{min}_a \text{ user supplied}$

Complexity: $O(N \log N \times \text{min}_\ell)$
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

Niching or Fitness sharing

since Mahfoud, 95

Selection based on

$$F'(X) = \frac{F(X)}{\sum_{X'} \text{sim}(X,X')}$$
Multi-objective multi-modal optimisation

Sharing-Pareto dominance

\[ X = (I, i, r, w) \text{ sp-dominates } Y = (I', i', r', w') \text{ iff} \]

- \( X \) Pareto dominates \( Y \) wrt \( a, \ell, \sigma \)
- \( X \) 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-Salpêtrière
- Artificial datasets
Artificial Datasets

$N$: number of sensors: $\{500 .. 4000\}$

$T$: number of time steps: $\{1000 .. 8000\}$
Artificial Datasets, 2

Draw 10 spatio-temporal patterns
Bias signals accordingly
Performances

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

<table>
<thead>
<tr>
<th></th>
<th>1,000</th>
<th>2,000</th>
<th>4,000</th>
<th>8,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>98 ± 5</td>
<td>93 ± 9</td>
<td>92 ± 7</td>
<td>79 ± 16</td>
</tr>
<tr>
<td>1000</td>
<td>96 ± 6</td>
<td>96 ± 6</td>
<td>82 ± 14</td>
<td>67 ± 12</td>
</tr>
<tr>
<td>2000</td>
<td>96 ± 5</td>
<td>87 ± 12</td>
<td>72 ± 14</td>
<td>49 ± 15</td>
</tr>
<tr>
<td>4000</td>
<td>89 ± 10</td>
<td>81 ± 13</td>
<td>56 ± 14</td>
<td>32 ± 16</td>
</tr>
</tbody>
</table>

Online Performance
Limitations

Very sensitive to the initialisation parameters
Functional Brain Imagery

Stable spatio-temporal pattern

Stable spatio-temporal pattern
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

![Graph 1: Time vs. Activity](image1)

![Graph 2: Time vs. Activity](image2)
Discriminant Spatio-Temporal Patterns
Discussion

**Discriminative learning**
- Given $H$, hypothesis space
- Find $h$, discriminating positive and negative examples.

**Generative learning**
- build $D_+$, $D_-$ distribution of positive / negative examples
- Example $X$ is positive iff $P_{D_+}(X) > P_{D_-}(X)$
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 \([1, N]\);  
- \( w = (1, 1, 1) \) \( \text{initial} = \text{Euclidean} \)  
- \( I = \)
  - \( t_1 \): uniformly drawn in \([1, T]\)  
  - \( \ell(I) \) drawn \( \sim \mathcal{N}(\min_{\ell}, \min_{\ell}/10) \) \( \min_{\ell} \text{ user supplied} \)  
  - reject if \( t_1 + \ell(I) > T \)
- \( r \): such that the ball contains all neighbors with bounded \( I \)-alignment:
  \[
  r = \min_k \{ d_w(i, k) \text{ s.t. } \sigma^I_{i,k} > \min_\sigma \}
  \]
  - reject if \( a(X) = |\mathcal{B}(i, r)| < \min_a \) \( \min_a \text{ user supplied} \)

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”  
  \textit{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
$N = 500, \ldots, 4,000$
$T = 1,000, \ldots, 8,000$
For $i = 1 \ldots N$
  For $t = 1 \ldots T$
    $C_i(t) = C_i(t - 1) + \epsilon \ast \pm 1$

10 Target patterns
$P = (i \text{ in } 1 \ldots N; \ I \subset [1, T]; \ w \in \mathbb{R}^3; \ r \in \mathbb{R}).$
$C_P = \text{average of } \{C_j(t), t \in I, d_w(i, j) < r\}$

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