9 novembre 2004

Fouille de Données Spatio Temporelles et Optimisation multi-critères

Nicolas Tarrisson & Michèle Sebag

IA – TAO, CNRS – INRIA

Université Paris-Sud Orsay

[Wiki:StagesTao]

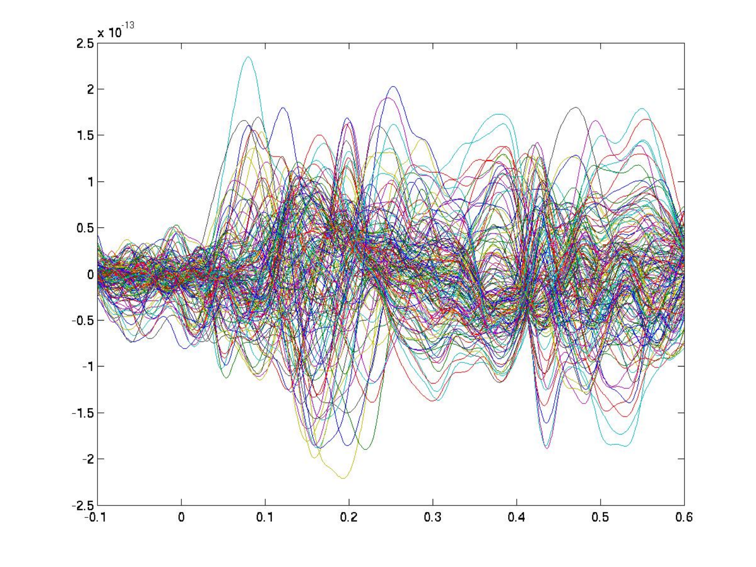
Contexte

Neuro-imagerie

- Des sujets, une expérience, des mesures

- Une nouvelle technologie : électro-magnéto-encéphalographie.
  pas de temps: .001 seconde

Les données



Structure (3D - 2D)

- Courbes $i = 1..N$
- $i \longrightarrow M_i = (x_i, y_i, z_i), X_i[t], t = 1..T$

Le but

Trouver des motifs spatio-temporels

- Une aire A : Boule $\mathcal{B}(M_A = (x_A, y_A, z_A), d_A, \varepsilon_A)$
- Un intervalle temporel $I \subset \{1..T\}$

caractérisant

$$\mathcal{V}(A,T) = \{X_k[t], t \in I, k \; tq \; d_A(M_k, M_A) < \varepsilon_A\}$$

tel que

deviation standard $(\mathcal{V}(A,T))$ faible

Procédure courante

- à la main $\longrightarrow$ i) ennuyeux; ii) subjectif
- peu de volontaires.

Grandes lignes

Propriétés voulues

- Passage à l'échelle
- Flexible $\Rightarrow$ Paramétrable $\Rightarrow$ Doit être calibré

$\Rightarrow$ Contrôle des ressources possible - Algorithme anytime

Discussion

- critères monotones (en ε_A, en I)
- critères antagonistes ($I \nearrow$, $\varepsilon_A \searrow$)
- exhaustivité ? Non : le résultat doit être vu par un humain.

Optimisation multi-critères

Optimisation classique

Trouver $ArgMax\{\mathcal{F}(x), \mathcal{F} : \Omega \rightarrow \mathbb{R}\}$

Optimisation multi-critères

Trouver $ArgMax\{\mathcal{F}_i,\ i = 1, 2..., \mathcal{F}_i : \Omega \rightarrow \mathbb{R}\}$

Evidemment, $\mathcal{F}_i$ antagonistes.

De qualité maximale, de prix minimal...

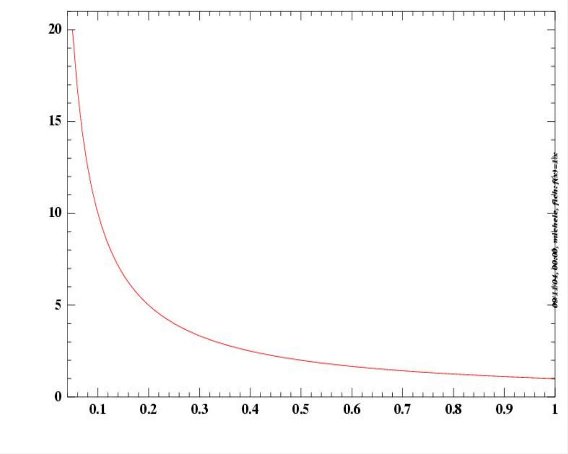
Front de Pareto

Domination de Pareto

- $x < y$ ssi $\mathcal{F}_i(x) \leq \mathcal{F}_i(y)$ et inégalité stricte pour au moins un i.

Front de Pareto

- Ensemble des solutions non dominées.



Stable Spatio-Temporal Patterns

Espace de recherche

$X =$

I_X intervalle temporel

k M_k centre de la boule spatiale

r rayon de la boule spatiale

$d_w = (a, b, c)$ distance pondérée

avec

$$d_w(M_k, M_j) = a.(x_k - x_j)^2 + b.(y_k - y_j)^2 + c(z_k - z_j)^2$$

Futur

$d_w \longrightarrow$: matrice W, $||M_k - M_j||_W = (M_k - M_j)^t.A.(M_k - M_j)$

Critères de dominance

$X = (I = [deb, fin], k$ **(**M_k **centre)**$, d_w = (a, b, c), r)$

- Longueur temporelle $l(X) = fin - deb$
- Le voisinage spatial $\mathcal{V}(X) = \{j \;/\; d_w(M_k, M_j) < r\}$
- La taille $|X| = l(X).|\mathcal{V}(X)|$
- La cohérence spatiale

$$a(X) = \sum_{j \in \mathcal{V}(X)} e^{-d_w(M_k, M_j)}$$

- La cohérence spatio-temporelle

$$\rho(X) = sum_{j \in \mathcal{V}(X)} \rho_I(X_k, X_j)$$

Où $\rho_I(X_k, X_j) = covariance(X_k[t], X_j[t])$ pour t dans l'intervalle I.

EC Multi-critères

EC classique

Trouver $ArgMax(\mathcal{F})$

- Initialisation
- Variations (croisement, mutation)
- Sélection

Problème multi-critères

Trouver $X = ArgMax\{l(X), a(X), \rho(X)\}$

Modifications essentielles

But : couvrir le front de Pareto

Archive

Sélection : d'après $\mathcal{F}'(x)$, où $\mathcal{F}'$ mesure :

Le rang de Pareto de x dans la population courante

Le pourcentage de l'archive dominé par x

...

Initialisation

Paramètres

Tirage des intervalles temporels — Nombre n_T

Tirage i uniforme in $[1, ..T]$

Tirage $\ell \sim \mathcal{N}(L, \sigma)$ — L, σ

$I = [i, i + \ell]$

Commentaires

Intervalles diversifiés de longueur faible, mais admissible

Initialisation de $d_w = (1, 1, 1)$ — pour commencer

Futur : connaissances du domaine

Initialisation, suite

Initialisation des motifs

Pour tout I — exhaustif sur les intervalles tirés

Pour tout $k = 1..N$ — exhaustif sur les capteurs

1. Ordonner M_j par $d_w(M_k, M_j)$ croissant

2. Soit $\mathcal{V}(I,k) = min\{j \ / \ d_w(M_k, M_j) < \ \textit{Seuil}_d\}$

3. Si $|\mathcal{V}(I,k)| > \ \textit{Seuil}_c$, alors
   - $r = max\{d_w(M_k, M_j) \ / \ j \in \mathcal{V}(I,k)\}$
   - Population $\bigcup = \{X = (I, k, w, r)\}$

Opérateurs de Variation

$$X = (I = [deb, fin], k \text{ } (M_k \text{ centre}), d_w = (a,b,c), r)$$

Mutation

- Muter w ou r mutation auto-adaptative
- Incrémenter/Décrémenter deb
- Incrémenter/Décrémenter fin
- Muter k (en l'un des capteurs voisins)

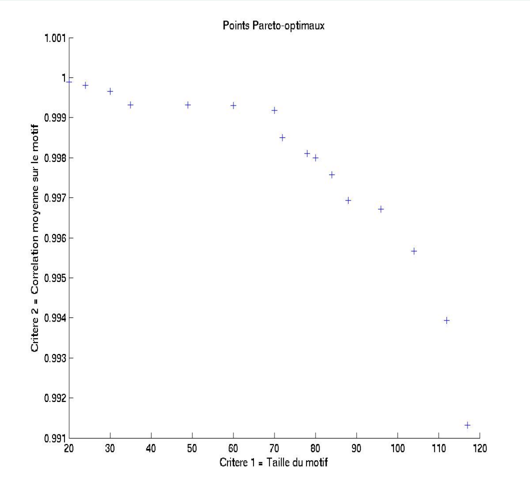
Premiers essais

Problème

$$\text{Trouver } X = ArgMax\{l(X), a(X), \rho(X)\}$$

Echec détecté par visualisation

Pas de variété : les Pareto dominants sont des variantes de la même zone.



Relaxation des critères

Domination ensembliste

$$A \subset_p B =_{def} |A \bigcap B| > p|A|$$

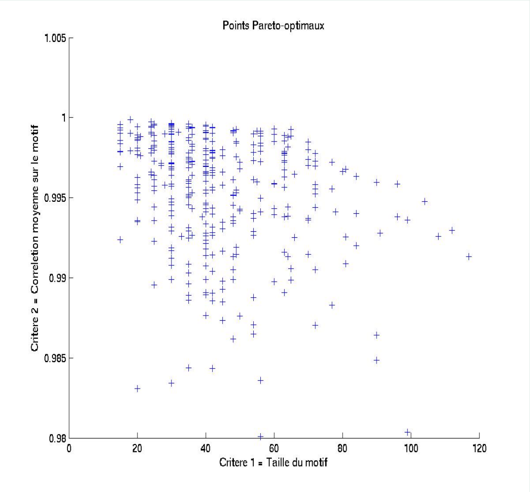
Critère agrégé de dominance $X = (I, k, w, r) \prec X' = (I', k', w', r')$ ssi

- $I \subset_p I'$
- $\mathcal{V}(X) \subset_p \mathcal{V}(X')$
- $a(X') \geq a(X)$ AND $l(X')^\alpha \rho(X') \geq l(X) l(X)^\alpha \rho(X)$
- $a(X') > a(X)$ OR $l(X')^\alpha \rho(X') > l(X) l(X)^\alpha \rho(X)$

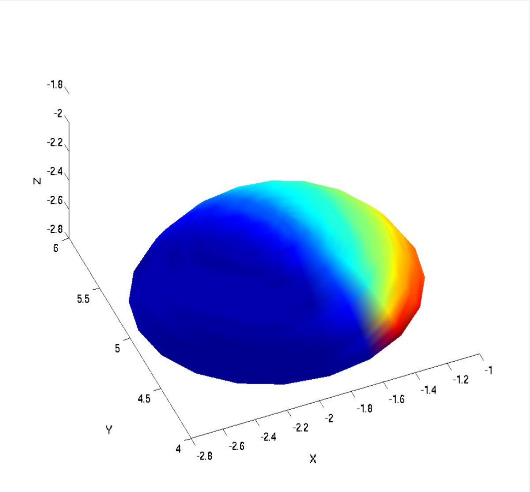
Futur

Ajuster α et p.

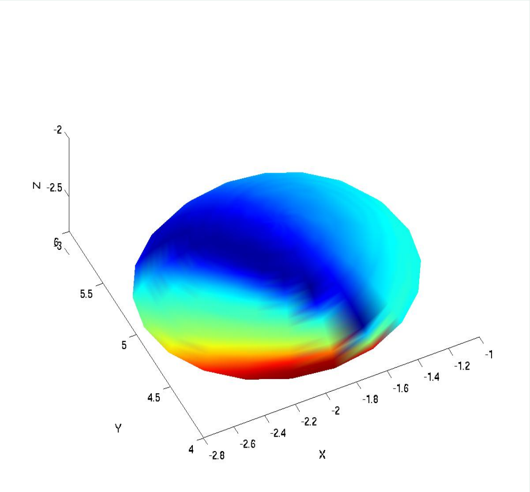
Avec Pareto dominance ajustée



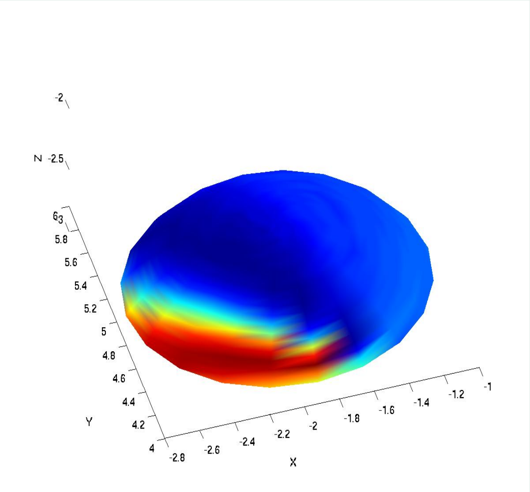
Motif 1



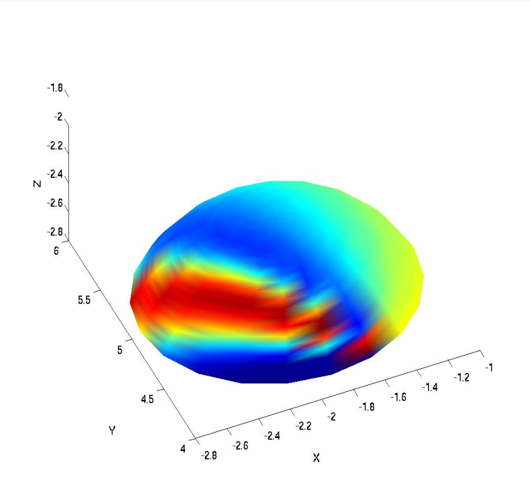
Motif 2



Motif 3



Motif 4

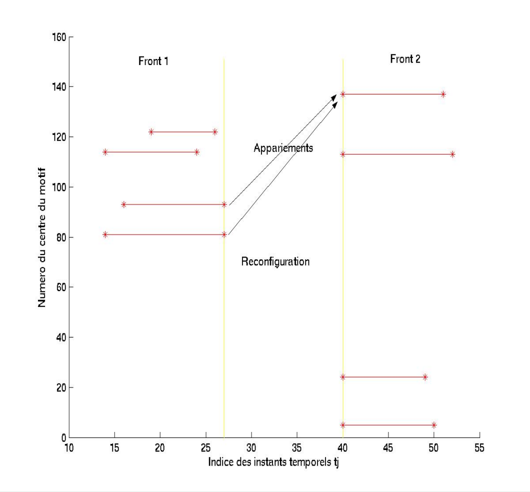


Du point de vue de l’expert

Construire des scénarios

telle zone “passe le témoin” à telle autre zone.

Appariement temporel simple



Alternatives

Independent Component Analysis

- Les aires spatiales se recombinent.
- Disposer des intervalles temporels ?

FD Spatio-temporelle

- Essentiellement spatio ou temporelle.
- Spatiale :
  - segmentation
  - analyse de dépendances
  - deviations et outliers
  - tendances
  - generalisation, caractérisation