

# Apprentissage non supervisé de motifs temporels, multidimensionnels et hétérogènes

## Application à la télésurveillance médicale

Florence Duchêne<sup>1</sup>, Catherine Garbay<sup>1</sup> et Vincent Rialle<sup>1,2</sup>

<sup>1</sup> Laboratoire TIMC-IMAG, Faculté de médecine de Grenoble, France

<sup>2</sup> Département d'Informatique Médicale de l'hôpital Michallon, Grenoble, France

Florence.Duchene@sophia.inria.fr,

Catherine.Garbay@imag.fr, Vincent.Rialle@imag.fr

**Résumé** : On propose une méthode générique d'extraction non supervisée de motifs temporels dans des séries multidimensionnelles et hétérogènes, appliquée à l'apprentissage des comportements récurrents d'une personne à domicile.

**Mots-clés** : Fouille de séries temporelles, multidimensionnelles et hétérogènes, Motifs temporels, Apprentissage non supervisé, Télésurveillance médicale.

Pour la détection des évolutions critiques à long terme de personnes à domicile, on propose un système d'apprentissage d'un profil comportemental dans la vie quotidienne. Un écart de comportement par rapport à ce profil peut être inquiétant car significatif d'une dégradation de l'état de santé. Il s'agit d'extraire des *motifs* "haut niveau" de séquences "bas niveau" collectées de capteurs installés au domicile. Un *motif* est le représentant d'une classe de sous-séquences récurrentes, et correspond à un comportement type de la personne. Ce problème d'apprentissage a les caractéristiques suivantes :

1. **Séquences analysées** – Les séquences analysées sont multidimensionnelles, hétérogènes, et *mixtes* : elles contiennent à la fois des sous-séquences représentatives de *motifs* et des "*non motifs*".
2. **Méthode utilisée** – L'extraction de motifs est *non supervisée* pour s'adapter aux spécificités individuelles et au manque de connaissances *a priori*.
3. **Séquences temporelles** – Les séquences analysées sont multidimensionnelles, hétérogènes, et *mixtes* – i.e. contenant à la fois des *motifs* et des "*non motifs*".
4. **Motifs extraits** – On recherche des *motifs multidimensionnels* afin d'éviter une sur-simplification du système observé, et la non détection de certaines évolutions critiques. Les instances d'un motif ont les caractéristiques suivantes :
  - **Variabilité dans les valeurs**, due à celle des comportements humains.
  - **Présence d'interruptions** dans la réalisation d'une activité (toilettes, etc.).
  - **Déformations et translation dans le temps**, car une même activité se répète à des instants et sur des durées variables.

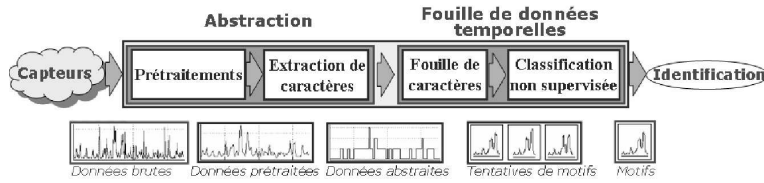


FIG. 1 – Méthode d’identification non supervisée de motifs temporels. Les signaux représentés illustrent sur une dimension le type des données disponibles après chaque étape.

Une originalité de ce travail est la considération de séquences multidimensionnelles et hétérogènes, dans l’objectif d’extraction de motifs multidimensionnels. Il est en particulier nécessaire de définir une mesure de similarité adaptée à la comparaison de séquences de ce type. La mesure proposée est basée sur la plus longue sous-séquence commune – *Longest Common Subsequence (LCSS)* (Duchêne *et al.*, 2004).

La première étape de la méthode proposée (voir figure 1) consiste en l’**abstraction** des données brutes issues des capteurs, pour leur donner un sens au regard de l’objectif de l’analyse. Il s’agit de mettre en évidence les tendances à plus ou moins long terme, en résumant les situations “stationnaires” observées. Une séquence est ainsi représentée par une succession de symboles estampillés. La **fouille de données** comprend d’abord une étape de **fouille de caractères**, basée sur la méthode des projections aléatoires (Buhler & Tompa, 2002; Chiu *et al.*, 2003), pour l’identification des sous-séquences récurrentes. Les sous-séquences significatives sont sélectionnées selon deux critères maximum de distance et minimum de collisions. Une méthode de synthèse basée sur une classification divisive des sous-séquences récurrentes permet ensuite de générer un ensemble de sous-séquences disjointes – les *tentatives de motifs*. Leur **classification** ascendante hiérarchique en *motifs* est enfin réalisée sur la base d’une mesure de distance.

L’approche proposée est expérimentée dans le cadre de la télésurveillance médicale à domicile, à partir des déplacements, postures, niveau d’activité, fréquence cardiaque moyenne d’une personne. De manière générale, on constate de bonnes performances des étapes d’extraction et de classification des motifs. À partir de séquences simulées pour une personne dans des conditions habituelles de vie, on vérifie qu’il est possible d’extraire des sous-séquences représentatives de comportements qu’on sait interpréter *a posteriori* en terme de la réalisation de certaines activités de la vie quotidienne.

## Références

- BUHLER J. & TOMPA M. (2002). Finding motifs using random projections. *Journal of Computational Biology*, 9(2), 225–242.
- CHIU B., KEOGH E. & LONARDI S. (2003). Probabilistic discovery of time series motifs. In *Proceedings of the 9<sup>th</sup> ACM International Conference on Knowledge Discovery and Data Mining (KDD’03)*, Washington DC., USA, p. 493–498.
- DUCHÊNE F., GARBAY C. & RIALLE V. (2004). Similarity measure for heterogeneous multivariate time-series. In *Proceedings of the 12<sup>th</sup> European Signal Processing Conference (EU-SIPCO)*, Vienna, Austria.