

Conférence spéciale

Méthodes de classification sur séries temporelles. Applications à la prévision et à la désaisonnalisation

Dominique LADIRAY¹

L'analyse des données "à la française" et l'analyse des séries temporelles ont chacune une longue histoire mais curieusement leurs chemins ne se sont que rarement croisés, au moins jusqu'à un passé récent. Dans les vingt dernières années, avec la mise à disposition d'énormes bases de données temporelles, il y a eu une explosion d'intérêt pour l'exploration de ces données. Des centaines d'articles ont alors proposé des méthodes et algorithmes pour classer, indexer, segmenter et discriminer les séries temporelles.

De nombreuses méthodes de classification, mesures de similarité et algorithmes ont été développés au cours des ans, essentiellement pour des données d'enquêtes. Malheureusement, la plupart de ces mesures de similarité ne peuvent être directement utilisées sur des données temporelles. De nouvelles distances et de nouvelles stratégies ont été définies, certaines d'entre elles étant basées sur des outils ou résultats récents de l'analyse des séries temporelles : coefficients cepstrum, transformée par ondelettes, modèles markoviens cachés, etc.

La première partie de l'exposé sera consacrée à la présentation des méthodes les plus utilisées en classification de séries temporelles et d'un outil SAS les mettant en œuvre.

Deux applications de ces méthodes seront ensuite proposées :

- une méthode de construction d'un modèle de prévision basée sur la classification de séries qui sera aussi comparée aux méthodes habituellement utilisées (approche « General to specific » ou modèles factoriels dynamiques) ;
- une méthode d'évaluation des performances de logiciels de désaisonnalisation à partir d'une base de « séries tests ».

¹Département Statistiques de Court-terme, DSE, Insee - dominique.ladiray@insee.fr