

Rapport sur les travaux présentés par monsieur Romain Tavenard en vue de l'obtention du diplôme d'Habilitation à diriger des recherches de l'Université de Rennes.

Florence d'Alché-Buc,
Professeur à Télécom Paris, Institut Polytechnique de Paris
Paris, 23 Novembre 2020.

Le rapport présenté par Romain Tavenard, maître de conférences à l'Université de Rennes est intitulé Apprentissage statistique et séries temporelles. Il comprend une présentation très synthétique des recherches qu'il a menées à la suite de son doctorat obtenu en 2011 à l'Université de Rennes et un curriculum vitae montrant ses activités relatives à l'enseignement, à l'encadrement, à la recherche contractuelle et à son rôle au sein de l'éco-système universitaire.

Travaux de recherche

Dans le cadre de l'apprentissage statistique à partir de données structurées, graphes ou séries temporelles, Romain Tavenard articule ses travaux autour de deux grands axes: la définition de métriques adaptées aux données structurées et l'apprentissage de représentation pour les séries temporelles. Ses travaux sont de nature méthodologique, avec un intérêt important pour la complexité en temps des algorithmes et le traitement d'applications réelles très variées. Il conclut par quelques perspectives.

Définition de métriques adaptées aux données structurées.

Sous ce chapeau sont réunis plusieurs travaux. Les premiers portent sur la définition d'un noyau produit dit temporel qui permet de coupler des informations sur la nature des objets qui sont observés dans le temps et les instants où ils sont observés. Romain Tavenard souligne qu'il s'agit ici de considérer les séries temporelles comme des distributions discrètes sur l'espace produit temps-variable explicative, une analogie qu'il reprend quand il s'intéresse aux distances de transport optimal entre graphes avec son doctorant Titou Vayer et ses autres co-encadrants. Assez naturellement il s'est aussi intéressé aux algorithmes de programmation dynamique (DTW) qu'il a notamment utilisé pour réaligner des séries temporelles multimodales.

Apprentissage de représentation pour les séries temporelles.

L'apprentissage de représentations latentes constitue un autre angle d'attaque pour traiter des données temporelles, souvent crucial pour obtenir de bonnes performances. Les outils que Romain Tavenard a mis en place sont fondés sur des modèles génératifs ou bien sur l'extraction de représentations apprises par réseaux de neurones.

Du côté des modèles génératifs, il a étendu lors de son postdoc un modèle hiérarchique latent de Dirichlet pour les motifs sémantique en vue de modéliser dans un cadre supervisé des actions dans des vidéos.

Dans le cadre d'un projet lié à la surveillance du trafic maritime, il s'est intéressé à l'identification de flux de navigation à partir d'un grand nombre de trajectoires observées par segmentation en modes de mouvements homogènes.

Les séries temporelles peuvent être aussi traitées par des réseaux convolutionnels et sur ce plan Romain Tavenard a proposé avec ses collègues d'intéressants développements. Parmi ceux-ci je distinguerai les travaux réalisés avec Lods et Malinowski qui imposent que la distance euclidienne entre représentations latentes coïncident avec une distance d'alignement au sens de DTW. Il s'est penché sur des représentations à base de shapelets ou sous-séries aléatoires pour lesquelles il définit une modification du Sparse-Group-Lasso afin de les sélectionner pendant l'apprentissage. Enfin les derniers travaux évoqués portent sur la classification précoce de séries temporelles où la contrainte de temps s'exprime dans la fonction de coût pour pénaliser une décision trop tardive.

Pour finir Romain Tavenard dresse quelques perspectives de travail qu'il situe notamment autour des séries temporelles d'objets arbitraires, l'adaptation temporelle de domaines et la prédiction structurée.

Le manuscrit présenté donne une petite idée des travaux de Romain Tavenard sans toutefois les unifier avec suffisamment de précision. C'est bien dommage qu'il ne dégage pas mieux la pertinence des travaux qu'on perçoit à la lecture des publications internationales de l'auteur dans différents journaux internationaux de très bon impact (10) et dans des conférences internationales variées et de très bon niveau (13). Il faut aussi citer notamment la collaboration avec Nicolas Courty, Lateitia Chapel et Rémi Flamary autour de la thèse de Titouan Vayer qui s'est en particulier avérée très fructueuse en terme de visibilité internationale (ICML, NeurIPS, *Algorithms*). Notons aussi qu'en parallèle de cette activité de recherche, Romain Tavenard est le créateur et le principal développeur d'une librairie tslearn Python dédiée à l'apprentissage à partir de séries temporelles.

Activités d'encadrement, projets et visibilité.

Romain Tavenard a participé au co-encadrement de 3 doctorants, deux ayant soutenus. Il a aussi formé de nombreux étudiants en Master. Il a développé ses recherches à travers de nombreux projets notamment ANR, pour lesquels il est souvent en première ligne. Ses travaux sont caractérisés par de nombreuses collaborations qu'il a nouées autour d'applications réelles ou avec des collègues ayant des centres d'intérêt proches. Actif dans sa communauté, il a été le co-organisateur d'un atelier international sur les données temporelles au sein d'ECML/PKDD de 2019.

Activités liées à la vie universitaire

Romain contribue en outre à la vie du département Statistiques et Informatique en assurant la fonction de directeur adjoint, ayant assuré aussi la fonction de directeur par intérim pendant près d'un an et demi.

Romain Tavenard mène donc de front des travaux de recherche de très bonne facture avec en ligne de mire la représentation et le traitement de séries temporelles de différentes natures et une analogie judicieuse avec les données structurées. L'ensemble de ses activités reflète une maturité certaine et a déjà permis à plusieurs chercheurs plus jeunes de se former. Pour toutes ces raisons je donne un avis favorable à la soutenance de l'habilitation à diriger des thèses.



Florence d'Alché-Buc

