

Sujet de la thèse	Les données fonctionnelles et le transport optimal. Application à la détection des émotions par analyse la parole.	
Spécialité de la thèse	Statistique Mathématique, Analyse Numérique et IA, Technologies de l'information de la communication	
Mots-clefs	Données fonctionnelles, Transport optimal, Opérateur de covariance, Traitement de la parole, Intelligence artificielle	
Type de thèse	cotutelle	
Pays	France	Tunisie
Laboratoire d'accueil	L@bISEN Yncréa Ouest, Equipe VISION-AD	Laboratoire COSIM, équipe traitement des signaux physiologiques
Directeurs de thèse	Maher Jridi	Sofia Ben Jebara
Encadrants	Jérémy Rouot, Matthieu Saumard	
Contexte Les données fonctionnelles sont un ensemble de méthodes pouvant traiter des applications de nature a priori très différentes comme l'écriture, le mouvement ou la parole. Dans ce contexte où les données sont des éléments d'un espace de dimension infinie et où plusieurs populations sont à discriminer, une méthode consiste à calculer la distance entre des opérateurs de covariance pour analyser les variations entre ces populations, Pigoli et al. (2014). Cette distance, appelée distance procustéenne, a permis d'étudier des applications intéressantes en imagerie médicale Zhou et al. (2016), en biologie Cabassi et al. (2017) ou encore en linguistique Pigoli et al. (2018).		
Projet Récemment, Masarotto et al. (2019) ont revu les propriétés mathématiques de cette distance dans le contexte du transport optimal capable de s'adapter aux données fonctionnelles. Ce point de vue permet d'utiliser les méthodes géométriques, analytiques et numériques propres à la théorie du transport optimal qui ont été développées au cours de la dernière décennie. Nous proposons dans cette thèse de prolonger leurs travaux et de développer des applications liées à l'analyse de la parole. Un des buts de la thèse est de trouver des formes universelles des différentes émotions contenues dans le discours oral et de détecter les différences fondamentales entre les émotions par analyse de la voix. En effet, en plus du contenu linguistique, la parole représente l'un des miroirs de l'état émotionnel de la personne. On voudrait le détecter au cours de cette thèse en utilisant des techniques d'intelligence artificielle combinées aux données fonctionnelles. Ce cadre applicatif constitue une continuité de travaux antérieurs menés dans le laboratoire COSIM dans le cadre d'un projet de recherche fédéré de lutte contre le terrorisme où l'une des composantes est la détection du comportement suspect par analyse de la voix. Cette application supporte les travaux réalisés par l'équipe VISION-AD du laboratoire L@ISEN pour contrôler l'accès à des zones restreintes à partir de la reconnaissance faciale à travers le contrôle d'accès par reconnaissance vocale.		

Problématiques

- Compréhension des différences intrinsèques et découverte de motifs communs entre les différentes émotions à travers la parole
- Étude théorique des relations entre transport optimal et données fonctionnelles
- Développement de méthodes numériques
- Application concrète avec démonstrateur

Approches méthodologique et technique envisagées

- Données fonctionnelles : estimation de l'opérateur de covariance, mise en pratique numérique de la distance procustéenne
- Outils mathématiques de l'analyse fonctionnelle et du transport optimal
- Traitements du signal et de la parole
- Méthodes d'intelligence artificielle propres au traitement naturel du langage
- Apprentissage machine (approches basées sur l'extraction de descripteurs et classification, approche d'apprentissage profond ou deep learning).

Références succinctes

- [Cabassi et al. 2017] Cabassi, A., Pigoli, D., Secchi, P., Carter, P. A. and others Permutation tests for the equality of covariance operators of functional data with applications to evolutionary biology. *Electronic Journal of Statistics* 11(2) (2017) 3815–3840.
- [Masarotto et al. 2019] Masarotto, V. and Panaretos, V. M and Zemel, Y. Procrustes metrics on covariance operators and optimal transportation of Gaussian processes. *Sankhya A* 81(1) (2019) 172–213.
- [Pigoli et al. (2014)] Pigoli, D., Aston, J. A. D., Dryden, I. L. and Secchi, P. Distances and inference for covariance operators. *Biometrika*, 101(2) (2014) 409–422.
- [Pigoli et al. 2018] Pigoli, D. and Hadjipantelis, P. Z. and Coleman, J. S. and Aston, J. AD The statistical analysis of acoustic phonetic data : exploring differences between spoken Romance languages. *Journal of the Royal Statistical Society : Series C (Applied Statistics)* 67(5) (2018) 1103–1145.
- [Zhou et al. 2016] Masarotto, V. and Panaretos, V. M and Zemel, Y. Regularisation, interpolation and visualisation of diffusion tensor images using non-Euclidean statistics. *Journal of Applied Statistics*, 43(5) (2016) 943–978.

Profil recherché

- La ou le candidat(e) doit avoir un diplôme de Master et/ou Ingénieur dans des domaines liés au Mathématiques appliquées (statistique, analyse numérique), télécommunications, technologies de l'information et de la communication, signaux et systèmes.
- Avoir une aptitude au développement de méthodes mathématiques avec un focus spécifique sur leurs implantations.
- Avoir un vif intérêt pour la recherche scientifique et être familier au moins avec l'un des outils/langages suivants : R, et/ou Python, et/ou Matlab.

Financement de la thèse en cotutelle

Le sujet de thèse est financé par les fonds propres du laboratoire L@bISEN de l'ISEN Yncréa Ouest. Le doctorant, au cours de sa période de présence en France (période de 4 à 6 mois par année), est financé avec une bourse de thèse à hauteur de 1000€/mois. Il est aussi possible de demander une bourse d'alternance pour un séjour prolongé en France.

Au cours de son séjour en Tunisie, il (elle) bénéficiera de la bourse gouvernementale.

Modalités de candidature

Le dossier de candidature doit comprendre votre CV, lettre de motivation, Relevés de notes de L3, M1, M2 (ou années équivalentes) et une éventuelle lettre de recommandation. L'ensemble du dossier doit être adressé avant le 20/08/2020 à

- Maher.jridi@isen-ouest.yncrea.fr
- Jeremy.rouot@isen-ouest.yncrea.fr
- Matthieu.saumard@isen-ouest.yncrea.fr
- sofia.benjebara@supcom.tn

en précisant le moyen grâce auquel vous avez eu accès au sujet.

Les dossiers seront traités par ordre d'arrivée. Seuls les dossiers complets seront considérés et seuls les candidats retenus seront contactés et convoqués pour un entretien en visio-conférence.