

Éditorial : l'analyse des données symboliques aujourd'huiD. Desbois et E. Diday¹

Les participant(e)s à SDA 2014 (reproduction avec l'aimable autorisation du Professeur Chen)

L'analyse des données symboliques (ADS) introduit une nouvelle façon de penser en « *Data Science* », considérant que les classes d'individus peuvent être envisagées elles-mêmes comme des unités statistiques. Les classes y sont décrites par des données dites « symboliques » car, pour prendre en compte la variabilité des individus qu'elles rassemblent, elles ne sauraient être réduites à une description par des données classiques, quantitatives ou qualitatives. En raison de l'introduction de ce nouveau cadre paradigmatique, l'ADS a jusqu'ici privilégié la consolidation de ses fondements théoriques et le développement de méthodologies adaptées mais maintenant les publications consacrées aux applications apparaissent en nombre croissant dans de multiples domaines. Ces publications récentes proviennent de pays les plus divers dans leurs orientations spécifiques de Recherche et Développement : en Amérique (USA, Brésil, Costa Rica), Asie (Chine, Japon, Corée, Taiwan), Europe (France, Italie, Espagne, Portugal, Pologne, Slovaquie), Australie (*South Whale University*, Sydney). Ainsi, le domaine d'application de l'ADS couvre un champ en extension où elle se trouve mobilisée par des domaines confrontés à des données complexes. Ces publications montrent que l'ADS a eu des applications dans les domaines suivants : santé, génétique, cyber-sécurité, socio-démographie, épidémiologie, économie, écologie, finance, météorologie, données boursières, surveillance de la santé, ingénierie de travaux publics (ponts, voies ferrées, centrales nucléaires, matériaux de structure, béton, etc.), fouille de textes, agriculture, éducation.

Afin de diffuser des développements intervenus récemment dans ce champ de recherche, la Revue MODULAD propose un dossier dédié à l'Analyse des données

¹ Les éditeurs de ce numéro spécial sur l'Analyse des données symboliques remercient tous les relecteur.e.s ayant accepté de participer au processus de révision des articles, ainsi que la rédaction de la Revue MODULAD et la direction des éditions RNTI pour l'accueil réservé à ce projet éditorial.

symboliques, en partie issu de réflexions échangées lors du séminaire spécifique qui s'est tenu en 2014 à l'*Academia Sinica* (Taïpei), après les séminaires de Wienerwaldhof en 2009, Namur en 2011, et Madrid en 2012. Depuis, ces séminaires annuels se sont poursuivis témoignant du dynamisme des travaux dans le champ de l'Analyse des données symboliques : à l'Université d'Orléans en 2015, à l'Université de Ljubljana en 2017 et à l'Université de Porto en 2018 (<https://sda2018.wixsite.com/ipvc/>), enrichissant l'analyse des données symboliques par de nouvelles méthodes et étendant son champ d'application. En 2019, le séminaire international « *Advances in Data Science for Big and Complex Data* »² de l'Université Paris-Dauphine est consacré aux méthodes basées sur les classes en tant que nouvelles unités statistiques. Ces développements se traduisent également par la mise à disposition de nouveaux outils, comme ceux indiqués par (Afonso, Diday et Toque, 2018).

Sans pour autant refléter toute la diversité des travaux présentés lors du séminaire international de Taïpei (<http://www3.stat.sinica.edu.tw/sda2014/>), les contributions à ce dossier s'organisent selon trois sections : *Concepts et méthodes*, *Applications* et *Réflexions*.

La section « **Concepts et méthodes** » (I) propose deux contributions : la première signée par *Edwin Diday* nous introduit au pouvoir explicatif et discriminant des variables et tableaux symboliques, explicitant les liens entre explication et discrimination ; la seconde présente un estimateur contractant de régression linéaire pour des données à intervalles, proposé par *Oldemar Rodriguez Rojas*.

La section « **Applications** » (II) comporte trois textes : le premier, écrit par *Stéphanie Bougeard* et *Carole Toque*, applique l'analyse en composantes principales covariantes sur des données vétérinaires à intervalles ; le second, proposé par *Sun Makosso-Kallyth* présente deux extensions de l'ACP sur données en intervalles à l'étude des tableaux de distributions macroéconomiques en les comparant ; le troisième, de *Dominique Desbois*, explore la distribution des quantiles conditionnels grâce aux outils de l'analyse des données symboliques par intervalles.

La section « **Réflexions** » (III) est constituée par un article de *Frédéric Lebaron* qui se focalise sur les potentialités de l'analyse des données symboliques pour appréhender la complexité du social à partir d'indicateurs internationaux publiés par la Banque mondiale.

Enfin, ce numéro de la nouvelle série inaugure une section **Varia** (IV) permettant la publication de travaux hors-dossier : ainsi, *Daniel Defays* nous propose une comparaison de plusieurs algorithmes d'appariement basés sur la distance entre matrices de dissimilarités.

Référence

Afonso F., Diday E., Toque C. (2018). *Data science par analyse des données symboliques. Une nouvelle façon d'analyser les données classiques, complexes et massives à partir des classes - Applications avec Syr et R*, Éditions Technip, Paris, 442 p.

² <http://vldowiki.fmf.uni-lj.si/doku.php?id=sda:meet:pa19>.