

En conclusion, les résultats recueillis confirment la réputation du mouton, même si les résultats expérimentaux suggèrent que le processus d'imitation est plus complexe que ne le laisse présager Rabelais. Ils indiquent que manipuler un ou quelques individus permet finalement de contrôler l'ensemble du troupeau. Les humains en ont probablement profité empiriquement pour domestiquer les troupeaux de moutons sauvages. Remarquons que les règles comportementales simples, qui rendent les troupeaux de moutons faciles à contrôler, sont les mêmes qui permettent aux moutons de développer des comportements collectifs auto-organisés complexes, comme le comportement intermittent commenté ci-dessus qui permet aux moutons de s'auto-organiser (en utilisant l'imitation) pour maximiser les chances de manger sans être mangé. Au-delà de nos études sur les moutons, le message important pour les personnes intéressées par la modélisation mathématique est que l'étude du comportement animal s'est enrichie ces dernières années d'une approche quantitative. La modélisation mathématique peut aider à tester des hypothèses et dévoiler les règles de comportement des groupes d'animaux.

Bibliographie,

[1] Merriam-Webster on-line dictionary:

«<http://www.merriam-webster.com/dictionary/sheep>»

[2] F. Ginelli, F. Peruani, M-H. Pillot, H. Chate, G. Theraulaz, R. Bon, "Intermittent collective dynamics emerge from conflicting imperatives in sheep herds", *Proceedings of the National Academy of Sciences* 112, 12729-12734 (2015).

[3] S. Toulet, J. Gautrais, R. Bon, F. Peruani, "Imitation combined with a characteristic stimulus duration results in a robust collective decision-making in sheep", *PLoS ONE* 10(10): e0140188 (2015).

[4] CNN online, <http://edition.cnn.com/2011/WORLD/africa/02/05/protecting.somali.cave.art/>, part of the "World's Treasures" issue, February 5th 2011.

[5] God Khnum, <https://en.wikipedia.org/wiki/Khnum>.

[6] Bernhard Plockhorst's (1825-1907) "Good Shepherd", https://en.wikipedia.org/wiki/File:Bernhard_Plockhorst_-_Good_Shepherd.jpg

[7] *Comic designed* by Brian Kent, online available at: <http://www.mensocomic.com>



Laurent Oudre

Maître de conférences au L2TI (Université Paris 13) et chercheur associé au CMLA (ENS Cachan - CNRS) et à COGNAC G (Université Paris Descartes - CNRS - SSA). Il est diplômé de Supélec et de l'Imperial College London et a soutenu sa thèse en traitement du signal audio en 2010 à Telecom ParisTech. Ses thématiques de recherche portent sur le traitement du signal, la reconnaissance de formes, la segmentation, la classification et l'apprentissage, dans divers contextes applicatifs (son, signaux physiologiques, vidéo...).

Nicolas Vayatis

*Directeur du CMLA et professeur des universités au département de mathématiques de l'ENS Cachan. Il y anime un groupe de recherche dans le domaine de l'apprentissage statistique et du traitement de données massives qui s'implique dans des projets interdisciplinaires aux interfaces. Ses principaux thèmes de recherche sont la théorie de l'apprentissage, la modélisation prédictive et l'inférence à partir des données issues de graphes réels. Nicolas Vayatis est co-auteur d'environ 80 publications dans des revues et des conférences internationales et Action Editor pour la revue *Journal of Machine Learning Research*. Il est également responsable de la formation de master M2 MVA (Mathématiques-Vision-Apprentissage).*



Quantifier le comportement humain : Nouveaux challenges en mathématiques et en traitement du signal

Pendant des décennies, une partie importante de la recherche en bio-ingénierie s'est concentrée sur le développement de capteurs toujours plus performants, plus précis et plus robustes. La qualité toujours croissante des mesures réalisées a permis de nombreuses découvertes médicales, à la fois en recherche clinique et appliquée (notamment en biomécanique ou en posturographie). La mesure du comportement humain était alors réservée à un contexte clinique et à un public d'experts utilisant des systèmes de mesure hautement spécialisés, parfois encombrants et souvent onéreux.

Mais en parallèle, la société civile s'est emparée d'une nouvelle génération de capteurs, peu encombrants et produits à bas coût dans les pays émergents. En particulier, l'apparition des téléphones intelligents, puis des montres et autres objets connectés a donné à chacun la possibilité de mesurer chacune de ses activités physiques. Ce marché est en pleine expansion et on peut désormais pour quelques centaines d'euros mesurer son sommeil, sa course, sa dépense énergétique... Ce phénomène de « quantification du soi » a des répercussions sociologiques, mais également scientifiques.

En effet, on est passé en quelques années d'une quantification experte nécessairement limitée à quelques sujets, à des possibilités presque illimitées de quantification à bas coût d'un grand nombre d'individus. Cette évolution rapide et spectaculaire ouvre la porte sur de nouveaux défis en mathématiques, en traitement du signal et en bio-ingénierie. Néanmoins, l'utilisation de ces nouvelles technologies dans un contexte clinique est encore décriée et contestée dans la communauté médicale. La première des objections concerne la qualité et la finition des capteurs utilisés dans les appareils connectés, qui ne satisfont pas les spécifications des systèmes actuellement utilisés en recherche clinique. Se pose alors non seulement la question de la qualité des données collectées, mais aussi de la fiabilité des appareils dans une utilisation clinique sécurisée et répétée. La deuxième crainte est liée au fait que, bien souvent, les données fournies à l'utilisateur ont été traitées par des algorithmes brevetés dont le contenu n'est pas disponible. Ces modules « boîte noire » ont tendance à inspirer la méfiance des médecins car ils ne bénéficient que rarement d'une validation dans des revues médicales. La troisième objection a pour objet les conditions de mesure dans un contexte clinique. Bien souvent, un médecin rencontre chaque jour de nombreux patients ayant tous des pathologies très variées. Et alors que les systèmes grand public sont adaptés pour une utilisation par des sujets sains, il est peu probable qu'ils soient adaptés à toutes les pathologies que rencontre un professionnel de santé. De la même façon, la robustesse et la précision attendues pour un système utilisé en clinique dépassent largement celles attendues par des utilisateurs amateurs.

Répondre à ces objections n'est pas chose facile : il s'agit à chaque fois d'apporter une réponse scientifique concrète à une interrogation justifiée, en introduisant à la fois de bonnes pratiques dans la collection des données, et en proposant des algorithmes robustes et documentés, accessibles et compréhensibles à la fois pour des mathématiciens ou traiteurs du signal, mais aussi pour les utilisateurs du système. Notre démarche s'inscrit dans le cadre d'un projet de rapprochement de deux unités mixtes de recherche du CNRS qui sont le *Cognition and Action Group* (Cognac G - CNRS INSB, Irba et université Paris-Descartes) d'une part et le Centre de mathématiques et de leurs applications (CMLA - CNRS Insmi et ENS Cachan) d'autre part. Nos deux unités travaillent conjointement avec des équipes de cliniciens du Service de santé des armées pour élaborer des méthodologies et des outils de quantification objective des aptitudes du sujet en contexte clinique. Les trois axes fondamentaux qui illustrent notre démarche sont décrits dans la suite.

Améliorer la qualité des données

Il est évident que la qualité des données obtenues avec des capteurs valant quelques euros ou avec des systèmes professionnels n'est pas comparable. Néanmoins, il existe plusieurs façons de limiter ce phénomène.

La première façon nous ramène aux débuts de la science expérimentale, par la définition de protocoles standards permettant de diminuer les effets parasites et de se focaliser sur le ou les phénomènes que l'on veut observer. Si par exemple, on décide de mesurer l'activité de course non pas sur une journée, mais sur un exercice bien déterminé, sur un tapis de course et durant quelques minutes, on s'affranchit de nombreux paramètres (le terrain, les conditions météo, les virages, etc.) qui permettent de se concentrer

sur l'activité motrice du sujet. La définition d'un protocole permet également d'introduire la comparabilité et la reproductibilité dans la prise de mesure. On peut par exemple compenser l'effet de capteurs de qualité médiocre, en demandant au sujet de réaliser un protocole plusieurs fois et en fusionnant les résultats obtenus. Le protocole permet également d'intégrer l'expertise du médecin dans la prise de mesure, en choisissant avec précision les mouvements ou les exercices que le sujet effectue en fonction des symptômes ou caractéristiques que l'on veut observer. Il est à noter que cette vision des choses est en opposition avec l'usage actuel des outils de quantification du soi, où l'idée est justement souvent de « tout enregistrer ». La qualité parfois médiocre des données obtenues par ce genre de systèmes est au moins autant due aux capteurs utilisés qu'au fait que l'on enregistre au cours de la journée différentes activités mal définies et qui ne sont pas toutes pertinentes pour l'analyse. En se recentrant sur une activité, dans des conditions maîtrisées, on augmente largement la pertinence des données récoltées.

La deuxième façon d'améliorer la qualité des données est de revenir à la source, c'est-à-dire aux données brutes fournies par les capteurs. Il est en effet illusoire d'utiliser des données déjà pré-traitées pour une analyse fine du comportement humain. En effet, les différents modules et algorithmes fournis avec les appareils « grand public » ont souvent pour but de résumer les données de la façon la plus concise possible. On n'a donc accès qu'à un nombre limité d'informations, qui ont bien souvent subi des filtrages et/ou des agrégations. Afin d'obtenir des signaux physiologiques de meilleure qualité, il convient d'accéder aux données avant pré-traitement, afin de contrôler toute la chaîne de mesure. En effet, un choix rigoureux d'algorithmes évolués de traitement du signal pour le ré-échantillonnage, l'interpolation d'échantillons manquants ou le débruitage de signaux peut parfois compenser les effets néfastes liés à l'utilisation de capteurs premier prix. De plus, le choix des paramètres de pré-traitement (fréquence d'échantillonnage, filtrage, ...) doit être guidé par le protocole et le phénomène à étudier. Ainsi, certains capteurs bas de gamme peuvent être inadéquats pour certaines applications, mais parfaitement suffisants pour d'autres, moyennant un pré-traitement adapté. Là encore, c'est la définition d'un protocole qui permet d'adapter le capteur à une tâche donnée, et donc d'améliorer la qualité de la mesure.

Des algorithmes pour les médecins, avec les médecins et l'enjeu de la reproductibilité

Le morcellement des compétences résultant de l'émergence de technologies et d'interfaces numériques dans le domaine médical a des conséquences néfastes sur la validité de l'observation empirique et de là sur la validation de protocoles médicaux et de voies thérapeutiques. D'une part, l'expert-clinicien n'est pas à même d'évaluer la justesse des mesures et des indicateurs proposés dans les logiciels qu'ils utilisent, d'autre part, la taille réduite des cohortes impliquées dans les essais cliniques ne permet pas de garantir suffisamment de validité statistique pour les corrélations observées. Par ailleurs, la dimension « boîte noire » des logiciels intégrés dans les appareils médicaux traduit une opacité de systèmes d'aide à la décision qui serait inacceptable dans la plupart des contextes industriels présentant des enjeux de sécurité et de monitoring de systèmes. La démarche promue par notre groupe de recherche est de contribuer à l'éveil des utilisateurs finaux quant au traitement et la valorisation des données en co-développant avec

eux les protocoles de collecte, de quantification des examens cliniques, ainsi que le prototypage d'algorithmes. Notre méthodologie implique une documentation exhaustive des développements proposés, la réalisation de plates-formes et d'outils en ligne permettant de réaliser des bancs d'essais à la fois des capteurs et des algorithmes assurant la reproductibilité des traitements de données. La démarche proposée repose largement sur le savoir-faire et les réalisations existantes dans le cadre du projet IPOL (*Image Processing OnLine* <<http://www.ipol.im>>) dont le leader scientifique est Jean-Michel Morel (Grand prix INRIA-Académie des sciences 2013, Médaille de l'innovation du CNRS 2015). IPOL est une plate-forme offrant la possibilité d'expérimentation numérique en-ligne d'algorithmes de traitement d'images. Elle recense actuellement plus d'une centaine d'algorithmes et a donné lieu à plus de 18 000 expériences. Le principe est à présent décliné pour l'analyse de signaux physiologiques.

L'enrichissement des bases de données et les retours d'expériences sur les algorithmes : collecter l'expertise du médecin

Après s'être assuré de disposer de données de qualité et d'outils de traitement transparents, l'enjeu suivant est celui de la valorisation effective des données. Il existe aujourd'hui pléthore de concepts, d'outils, d'algorithmes et d'implémentations issus des domaines de la statistique en grande dimension et de l'apprentissage (machine learning) qui rendent possibles la prise en compte de données présentant un certain degré de complexité et d'hétérogénéité. Or, rappelons-le les données ne savent pas ce qu'elles sont ! et le mathématicien, quel que soit son talent ou son niveau d'expertise, ne pourra pas à lui seul proposer des modélisations et des interprétations pertinentes. Tout traitement statistique doit d'emblée considérer des a priori qui vont conditionner la façon de

résumer et d'agréger des données, de déterminer les échelles d'analyse appropriées aux phénomènes qu'on souhaite appréhender. Un point clé est le recueil d'annotations par le médecin au cours de l'acquisition de données par les capteurs (ces annotations constituent des 'métadonnées'). En effet, l'étude empirique de la physiopathologie révèle une diversité des troubles et des atteintes possibles si bien que de nombreux cas révèlent en réalité une forme de pluripathologies, notamment chez le sujet âgé qui cache de nombreux pièges en cas de traitement statistique superficiel. A l'autre bout de la chaîne, le mathématicien peut largement bénéficier de l'expérience et l'expertise du médecin au moment de la confrontation des résultats des analyses statistiques ou prédictives. Cette étape est également clé puisqu'elle structure les échanges au travers de la validation de la chaîne de traitement des données depuis le capteur jusqu'à la « décision » et elle constitue parfois le moment de la découverte scientifique lorsqu'elle révèle des effets inattendus ou propose des vues nouvelles sur la pratique clinique.

Voici en résumé, les grandes lignes de notre action aux interfaces des mathématiques et de la neurophysiologie à travers le projet COGNAC G-CMLA. Par les problématiques d'ingénierie, de plates-formes logicielles, de médecine translationnelle, elle peut paraître éloignée du cœur scientifique de la recherche en mathématique. Or, elle se veut poser les bases pour rendre accessible à la communauté de recherche en traitement du signal et mathématiques appliquées des données de qualité qui stimuleront les questionnements scientifiques futurs. Les outils méthodologiques et informatiques développés dans le cadre du projet constituent des vecteurs importants pour la pérennisation et la consolidation des contributions scientifiques en rendant possible leur validation via la reproductibilité et la confrontation au réel. N'est-ce pas ce que tout scientifique se doit de rechercher ?