

Les répercussions des Big Data sur les procédés de recherche en psychologie

Emilie Joly-Burra et Paolo Ghisletta, Faculté de psychologie et des sciences de l'éducation, Université de Genève

44

S'il est aujourd'hui indéniable que les Big Data ont créé une véritable révolution dans le domaine des computer-sciences et du marketing, c'est au tour de la psychologie de modifier ses pratiques en fonction de cette nouvelle réalité de génération, récolte, stockage et analyse des données.

Depuis plusieurs décennies déjà, les chercheurs en psychologie ont compris l'intérêt de récolter de très larges volumes de données pour mieux comprendre les processus sous-tendant le fonctionnement psychologique tel qu'en témoignent, par exemple, les nombreuses études longitudinales sur de larges échantillons ou cohortes de population (ex. Berlin Aging Study, LIVES). Cependant, ces dix dernières années ont vu une nette augmentation de la quantité d'indicateurs récoltés pour comprendre le fonctionnement humain (ex. données socio-économiques, cognitives, affectives, de personnalité, génétiques, neuro-fonctionnelles). Les processus mêmes de création et de récolte des données ont profondément évolué avec l'avènement de l'ère des médias sociaux et du «quantified self». Des terabytes de posts Facebook, de tweets ou de données physiologiques issues de trackers d'activité physique ou de smartphones sont créés chaque jour, et il ne suffit que de quelques lignes de codes pour les extraire.

La nécessité de développer des infrastructures de stockage et d'analyse de cette masse de données en constante augmentation apparaît donc plus évidente que jamais. Pour répondre à ce besoin, la discipline vit actuellement un véritable essor dans le développement de techniques et d'outils statistiques permettant de tester les théories psychologiques sur de gros volumes de données^[1]. On voit

ainsi apparaître de nouveaux outils statistiques mêlant, par exemple, apprentissage automatique («machine learning») et modèles à équations structurales (ex. Structural Equation Model Forests^[2]).

Applications actuelles et défis futurs

Les modèles cités ci-dessus ont ainsi permis d'identifier les prédicteurs les plus influents de différentes dimensions du bien-être ainsi que leurs effets d'interactions non linéaires^[3]; de montrer que deux des 65 prédicteurs les plus puissants de survie dans un échantillon de 6203 personnes sont de nature psychologique et pas uniquement médicale ou sociodémographique^[4]; ou encore, dans le domaine de la psychologie de la personnalité, les patterns de «likes» Facebook ont fourni une meilleure prédiction des traits de personnalité que l'évaluation classique par questionnaires^[5].

Si les Big Data semblent ouvrir un vaste champ des possibles pour la recherche en psychologie, cela n'est pas sans soulever un ensemble de défis. D'une part, plusieurs psychologues se retrouvent confrontés à des défis techniques tels qu'un manque de connaissances approfondies en programmation, souvent nécessaires au traitement des Big Data. D'autre part, la facilité d'accès à des données personnelles potentiellement sensibles pose des questions éthiques majeures, comme l'a rappelé le récent scandale du Facebook - Cambridge analytica. Sans même parler d'utilisation commerciale des données, les chercheurs se doivent de garder le respect de l'humain et de la vie privée au centre de leur démarche, et d'adopter des pratiques empêchant les retombées négatives du traitement de données personnelles. Finalement, se pose la question de la véracité des données et du risque de surinterpréta-

tion des modèles obtenus sur de très larges échantillons, ce qui peut conduire à une surestimation de l'importance des effets mis en lumière (d'où l'importance de considérer non seulement la significativité statistique d'un résultat, mais également la taille de son effet, un aspect mis en avant en psychologie).

De par leurs connaissances des théories psychologiques de l'humain, de la psychométrie et des statistiques, les psychologues se trouvent dans une position privilé-

giée pour donner du sens à ces masses colossales de données^[6]. Dans une discipline où la recherche est résolument guidée par le test strict d'hypothèses théoriques, l'ère des Big Data permet d'entrevoir un enrichissement mutuel entre recherche guidée par la théorie (Theory driven) et recherche guidée par les données (Data driven) afin d'aboutir à une compréhension plus fine du fonctionnement humain dans toute sa complexité.

Références

- ^[1] Cheung, M. W. L., & Jak, S. (2016). Analyzing big data in psychology: A split/analyze/meta-analyze approach. *Frontiers in Psychology, 7*, 738.
- ^[2] Brandmaier, A. M., Prindle, J. J., McArdle, J. J., & Lindenberger, U. (2016). Theory-guided exploration with structural equation model forests. *Psychological Methods, 21*(4), 566.
- ^[3] Miller, P. J., Lubke, G. H., McArtor, D. B., & Bergeman, C. S. (2016). Finding structure in data using multivariate tree boosting. *Psychological Methods, 21*(4), 583.
- ^[4] Aichele, S., Rabbitt, P., & Ghisletta, P. (2016). Think fast, feel fine, live long. A 29-year study of cognition, health, and survival in middle-aged and older adults. *Psychological Science, 27*(4), 518–529. <https://doi.org/10.1177/0956797615626906>
- ^[5] Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences, 112*(4), 1036–1040.
- ^[6] Cheung, M. W. L., & Jak, S. (2016). Analyzing big data in psychology: A split/analyze/meta-analyze approach. *Frontiers in Psychology, 7*, 738.

Les auteurs

Emilie Joly-Burra



Emilie Joly-Burra a obtenu son Master en Psychologie à l'Université de Genève en 2013 et poursuit actuellement son travail de doctorat sur l'utilisation des méthodes mixtes (quantitatives et qualitatives) pour étudier la relation entre les buts et le bien-être chez les personnes âgées. Elle enseigne également des travaux dirigés en statistiques appliquées à la psychologie en Bachelor et Master.

Paolo Ghisletta



Paolo Ghisletta est professeur de méthodologie et d'analyse de données à la Section de psychologie de la Faculté de psychologie et des sciences de l'éducation de l'Université de Genève, et doyen de la filière Bachelor of Science in psychology francophone à la Formation universitaire à distance, Suisse (Brigue). Il s'intéresse aux changements psychologiques au cours de la vie et aux méthodologies pour les étudier. (Pôle de recherche national LIVES: Surmonter la vulnérabilité: Perspective du parcours de vie. Universités de Lausanne et de Genève)