

NRC Publications Archive Archives des publications du CNRC

L'âge des données

Tapping, Ken

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.4224/23000789>

L'astronomie au gré des saisons, 2016-09-20

NRC Publications Archive Record / Notice des Archives des publications du CNRC :

<https://nrc-publications.canada.ca/eng/view/object/?id=70e37e6b-507a-4975-aa40-b4a8614eb289>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=70e37e6b-507a-4975-aa40-b4a8614eb289>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

L'ÂGE DES DONNÉES

Ken Tapping, le 20 septembre 2016

Il y a quelques années, je discutais avec un collègue du CNRC à Ottawa des difficultés qu'il y avait à extraire de l'information utile des grands ensembles de données. On peut en effet comparer les données brutes au minerai de fer : en soi, il a peu d'utilité, contrairement au fer que l'on en extrait.

Mon collègue et moi discutons en fait de la difficulté de trouver dans une meule de foin des éléments qui ne sont pas du foin, mais qui y ressemblent étrangement. Ces objets ressortent du lot, mais l'on ignore au départ en quoi ils se distinguent. Pour les trouver, il faut recourir à des techniques informatiques telles que « l'exploration de données », les « réseaux neuronaux » et « l'apprentissage machine », qui sont des techniques d'analyse de données assistée par ordinateur (ADAO).

À l'époque, mon collègue m'avait fait part d'une expérience qu'il avait menée sur les données d'observation amassées sur quelque 22 000 galaxies. Il avait demandé à l'ordinateur de les classer en catégorie, sans donner aucune précision sur les galaxies ni sur ce qu'il recherchait. Le tri a pris du temps, mais l'ordinateur a produit une mine de nouveaux renseignements au sujet de ces galaxies.

L'un des projets sur lesquels nous travaillons consiste à trouver des techniques pour analyser des carottes de glace, des anneaux de croissance d'arbres et des données sur les taches solaires afin de monter une base de données sur l'activité solaire couvrant une très longue période, sans préciser au départ à l'ordinateur le genre d'information que nous recherchons. En sciences, comme dans bien d'autres domaines d'activité humaine, on a souvent raté des découvertes importantes en se cantonnant trop « à ce que l'on recherchait ».

Le problème d'analyse est devenu épineux dans le domaine de l'astronomie. Autrefois, les observations astronomiques se soldaient par des

notes jetées dans un cahier, quelques images sur pellicule et parfois, des courbes spectrales. Ces données étaient enregistrées sur bandes magnétiques pour analyse ultérieure. Depuis l'avènement des télescopes optiques et des radiotélescopes modernes, cette époque est complètement révolue. Par exemple, le radiotélescope CHIME (Canadian Hydrogen Intensity Mapping Experiment ou expérience canadienne de cartographie de l'hydrogène), dont la construction est bien avancée ici à l'Observatoire, pourra capturer d'un seul coup une grande région du ciel visible et les milliers d'objets qui s'y trouvent. Dès qu'il se mettra en marche, cet instrument produira un torrent de données. En comparaison, le Réseau d'un kilomètre carré (Square Kilometre Array), projet international auquel participe le Canada et qui vise à construire le radiotélescope le plus grand et le plus puissant jamais réalisé, produira un raz de marée de données. Le volume de données sera tellement grand que nous ne pourrons toutes les stocker. Il faudra d'une manière ou d'une autre procéder à un traitement préliminaire pour en réduire la quantité sans perdre aucune information scientifique valable, et ce, bien souvent sans savoir à l'avance ce qui sera utile. C'est là qu'interviennent les techniques d'ADAO.

Le problème ne se limite pas à l'astronomie ni même à la science. En raison de l'augmentation exponentielle de nos capacités sur tous les plans – pour acquérir, stocker, traiter et déplacer d'immenses volumes de données facilement et rapidement à l'échelle du globe –, nous produisons des masses de données qui dépassent l'entendement. Tout ce que nous souhaitons savoir se trouve probablement « quelque part » dans ces données. Les navigateurs et les moteurs de recherche que nous utilisons couramment sont utiles pour trouver de l'information dans des données qui ont déjà été extraites et traitées. Il est toutefois beaucoup plus difficile de fouiller des données brutes, surtout lorsque l'on n'a qu'une idée floue de ce que l'on recherche. Les nouvelles techniques d'ADAO permettent de faire de l'ordinateur un assistant plutôt qu'un simple

supercalculateur. Si les hommes préhistoriques vivaient à l'âge de pierre parce qu'ils créaient des outils en pierre, nous qui vivons ensevelis sous des masses de données de toutes sortes, y compris sur nous-mêmes, vivons donc à l'âge des données.

Le 22 septembre, en descendant vers l'hémisphère sud, le Soleil croisera l'équateur céleste, ce qui marquera l'équinoxe d'automne et la fin officielle de l'été. Mars et Saturne apparaissent très bas à l'horizon au sud-ouest, après la tombée de la nuit. Mars est à gauche, Saturne à droite. La Lune entrera dans son dernier quartier le 23.

Ken Tapping est astronome à l'Observatoire de radioastrophysique du Conseil national de recherches du Canada, à Penticton (C.-B.) V2A 6J9.

Tél. : 250-497-2300, téléc. : 250-497-2355

Courriel : ken.tapping@nrc-cnrc.gc.ca