

Transformer en jeu un problème scientifique réel et difficile est une belle idée mise en œuvre depuis peu, mais une idée proche l'a anticipée. Un groupe de chercheurs peut demander de l'aide à l'extérieur pour réaliser un travail qu'il n'arrive pas à effectuer seul, car cela exige trop de temps ou d'énergie. C'est ce qui s'est passé avec le projet *Galaxy Zoo*.

Des millions de galaxies

Pour comprendre la formation et l'évolution des galaxies, il est nécessaire de les classer selon leur forme et de disposer de nombreux exemples. Une classification sommaire distingue les galaxies spirales et les galaxies elliptiques. Un spécialiste disposant d'une image de galaxie indiquera assez rapidement son type et en classera quelques centaines sans se donner beaucoup de mal. La difficulté provient du nombre de galaxies que les chercheurs veulent prendre en compte (plusieurs millions), car il est important pour eux de

disposer de classifications aussi exhaustives que possible. Une première version du projet *Galaxy Zoo* (<http://www.galaxyzoo.org>) a permis, grâce aux internautes volontaires, de classer un million de galaxies. Les images provenaient d'une caméra numérique installée à l'Observatoire d'Apache-Point, au Nouveau-Mexique (États-Unis). Sans la mobilisation des 80 000 bénévoles, la caractérisation des 60 millions de galaxies repérées par le télescope spatial *Hubble* aurait été impossible.

Par ailleurs, lors de ces travaux d'observation et de classification, un objet étrange a été identifié par une internaute, Hanny van Arkel. L'objet, nommé *Hanny's Voorwerp*, est mystérieux et a suscité discussions et conjectures chez les spécialistes, prouvant à nouveau le bien-fondé d'un examen aussi exhaustif que possible, par des humains, des données astronomiques aujourd'hui disponibles en quantité... astronomique.

Si l'intelligence humaine réussit souvent mieux à calculer ou à classer que la machine, c'est que le problème a été bien traduit et mis en images pour que la recherche de sa solution devienne amusante ou très simple (dans le cas de *Galaxy Zoo*) grâce à un logiciel interactif mis en ligne et conçu dans ce but. Cela est possible depuis que les performances graphiques des ordinateurs de bureaux ont atteint les extraordinaires niveaux actuels.

Une autre raison du succès de ces méthodes est l'existence des réseaux informatiques et de la communication instantanée qu'ils permettent entre les organisateurs du travail et une multitude de joueurs potentiels dispersés dans le monde. Grâce à ces progrès technologiques, une masse importante de joueurs tente sa chance et fournit un travail irremplaçable. L'idée est de confier à une « foule » une tâche de calcul qu'on ne sait pas obtenir des machines et qu'un humain seul a peu de chance de mener à bien. L'intelligence humaine bat l'ordinateur dans ces exemples parce qu'elle s'est massivement mobilisée et coordonnée.

Cette idée de s'adresser au grand nombre pour obtenir ce qu'on ne réussit pas seul ou avec ses machines n'est pas

4. Karpov et Kasparov battent la foule

L'idée de confier à la foule une tâche trop difficile pour un être humain, et dont les exemples présentés dans l'article montrent qu'elle fonctionne souvent à merveille, a pu donner l'illusion que le *crowdsourcing* était apte à tout faire.

Il n'en est rien : dans un match sur Internet en 1996, Anatoli Karpov écrasa son adversaire, « Le reste du monde ». Toutefois, les coups étaient décidés à la majorité et certains joueurs

faibles votaient pour les mauvais coups.

Le match qui a été organisé en 1999 entre Gary Kasparov et « 50 000 bons joueurs » était plus difficile et a prouvé à nouveau que la foule, dans un tel cas, se fait battre par un individu seul. Alors que 50 000 personnes provenant de 75 pays différents et conseillées par quatre joueurs professionnels, dont Étienne Bacrot, collaboraient pour jouer et battre G. Kasparov en une fantastique

partie d'échecs, ce fut lui qui, en définitive, l'emporta après des échanges étendus sur 62 coups (voir www.chessgames.com/perl/chessgame?gid=1252350). Puisque l'ordinateur *Deep Blue* abattu G. Kasparov aux échecs, et que G. Kasparov a battu la foule, doit-on en conclure que, nécessairement, *Deep Blue* ou l'un de ses successeurs battrait lui aussi la foule ?

A. Karpov rejoua en juillet 2012 une partie majoritaire et gagna en 37 coups (www.europe-echecs.com/art/parte-majoritaire-contre-karpov-fin-de-la-partie-majoritaire-de-karpov-4421.html).

Ni les échecs ni la science ne sont démocratiques : ce n'est pas la majorité qui gagne. En science, le novateur a raison contre les habitudes de pensée de tous.



akora.com

5. Le service reCAPTCHA

Pour certaines opérations, on vous demande de retranscrire quelques symboles mal écrits formant un mot. Le but est d'empêcher un programme de se substituer à vous et, par exemple, de créer massivement des comptes de courrier électronique. On suppose que seul un être humain peut traduire le dessin présenté en symboles. Comme ce sont des tests pour distinguer homme

et machine, tel celui qu'Alan Turing proposa en 1950, on les nomme CAPTCHA, de l'anglais *Completely Automated Public Turing Test to Tell Computers and Humans Apart*.

Une utilisation des CAPTCHA pour mettre en place un « crowd-computing » massif est pratiquée depuis plusieurs années. Le principe, nommé reCAPTCHA, consiste à demander à la personne qui s'au-

thentifie non pas une seule lecture de lettres brouillées, mais deux. Le premier mot sert à tester si l'on a bien affaire à un humain ; la seconde est une image tirée d'un document que les gestionnaires du reCAPTCHA veulent numériser et que les algorithmes de reconnaissance automatique de caractères ont repéré comme délicat ou totalement illisible. La seconde réponse de l'utilisateur permet de numériser le morceau incompréhensible, travail contrôlable par un autre utilisateur pour que la sûreté soit accrue. Ainsi sont traités des documents trop difficiles pour les numérisations automatiques. Le système est utilisé pour les archives du *New York Times*.



nouvelle. En 2006, Jeff Howe, rédacteur au magazine *Wired*, l'a nommée *crowdsourcing* (« externalisation vers la foule » ou « distribution du calcul à grande échelle »). Des exemples en dehors des sciences ont été mis en œuvre, dont un programme particulièrement étonnant, nommé reCAPTCHA, où les internautes participent à la validation des numérisations d'ouvrages et travaillent sans même en avoir conscience.

Quand la foule calcule

En réfléchissant à ce *crowdsourcing*, certains ont noté des utilisations anciennes du principe datant d'une époque où les ordinateurs n'existaient pas. Ainsi, le *Oxford English Dictionary* a été élaboré au milieu du XIX^e siècle avec l'aide de 800 volontaires qui copiaient des passages de livres, ce qui mit à la disposition des éditeurs un grand nombre de citations précises et une information fiable sur l'usage des mots et leur sens en langue anglaise. *Wikipedia* est bien sûr une forme de *crowdsourcing* et, bien qu'écrite par ceux qui le souhaitent, sa fiabilité n'est pas loin d'égaler celle des autres encyclopédies, tout en proposant une quantité d'articles bien supérieure à

la plus grosse des encyclopédies fondées sur le travail d'experts.

Notons toutefois que le fait de confier une tâche à un très grand nombre de personnes ne produit pas nécessairement des résultats meilleurs que ceux qu'on tire d'un individu isolé et compétent. L'exemple le plus spectaculaire d'un échec du *crowdsourcing* est donné par le match organisé en 1999 entre G. Kasparov et le reste du monde, constitué de 50 000 personnes provenant de 75 pays différents. Durant toute une partie qui s'étendit sur 62 coups, à raison d'un coup tous les deux jours, G. Kasparov affronta la foule des joueurs d'échecs du monde entier qui tentait de le battre. Malgré une organisation permettant aux bons joueurs de peser suffisamment dans les décisions collectives des coups à jouer, G. Kasparov gagna. La partie a été jugée l'une des plus importantes et des plus riches de l'histoire du jeu d'échecs, et un commentaire de 200 pages a été publié pour en analyser les nombreuses nouveautés et subtilités. Reste que la foule a perdu ! N'oublions pas non plus que la « sagesse des masses » a conduit aux plus grandes bulles spéculatives, à de multiples catastrophes économiques, et à l'avènement des pires dictateurs. ■

L'AUTEUR



Jean-Paul DELAHAYE est professeur à l'Université de Lille et chercheur au Laboratoire d'informatique fondamentale de Lille (LIFL).

BIBLIOGRAPHIE

A. Kawrykow *et al.*, *Phylo : A citizen science approach for improving multiple sequence alignment*, *PLoS ONE*, vol. 7(3), e31362, 2012.

Kasparov versus the world, *Wikipedia*, 2012 : http://en.wikipedia.org/wiki/Kasparov_versus_the_World

S. Thomas, *9 examples of crowdsourcing, before crowdsourcing existed*, 2012 : <http://memeburn.com/2011/09/9-examples-of-crowdsourcing-before-crowdsourcing-existed/print/>

F. Khatib *et al.*, *Algorithm discovery by protein folding game players*, *PNAS*, vol. 108(47), pp. 18949-18953, 2011.

F. Khatib *et al.*, *Crystal structure of monomeric retroviral protease solved by protein folding game players*, *Nat. Struct. Mol. Biol.*, vol. 18, pp. 1175-1177, 2011.

C. Lintott *et al.*, *Galaxy Zoo : Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey*, *Monthly Notices of the Royal Astronomical Society*, vol. 389(3), pp. 1179-1189, 2008.

J. Howe, *The rise of crowdsourcing*, *Wired Magazine*, juin 2006.

J. Surowiecki, *The Wisdom of Crowds*, Anchor Books, 2004.