L’ère des robots-journalistes

30/08/2010 à 12h15



Des chercheurs américains créent, dans le secret de leurs laboratoires, des techniques journalistiques révolutionnaires. Articles ou journaux télévisés sont conçus par des ordinateurs.

A première vue, rien de surprenant. Un compte rendu de sport d’une confondante banalité: «*Les efforts remarquables de Joe MauerMinnesota Twins contre les Texas Rangers lundi dernier au stade d’Arlington. Les Rangers l’ont emporté sur un score de 8 à 5 […] Quand il maniait la batte, Mauer a été excellent de bout en bout. Il a marqué une fois dans la première manche et deux fois dans la sixième. Du côté des Texans, l’artisan de la victoire est sans conteste Tommy Hunter, qui a remporté avec brio son cinquième match d’affilée…* »

Un article comme il en existe donc des milliers, publiés dans les pages sports de la presse américaine. Seule différence, mais de taille: il est signé «*The Machine*», préparé et rédigé par un programme d’intelligence artificielle, baptisé Stats Monkey.

> ***Depuis des décennies, dans le monde, des ouvriers découvrent un beau matin qu’ils vont être remplacés par un robot. Si les journalistes se croyaient à l’abri de ce genre de mésaventure, ils avaient tort. Depuis quelques mois, ils peuvent aller à Evanston (Illinois), près de Chicago, pour voir et tester le système qui va peut-être bientôt les suppléer. Il est tapi dans un réseau d’ordinateurs appartenant au laboratoire d’information intelligente***

(Infolab), installé sur le campus de l'Université du Northwestern.

Pour déclencher Stats Monkey, il suffit qu'un humain lui indique à quel match il doit s'intéresser.

Une fois lancé, il travaille automatiquement de A à Z. Il commence par télécharger les tableaux chiffrés publiés par les sites web des ligues de base-ball, et collecte les données brutes: score minute par minute, actions individuelles, stratégies collectives, incidents… Puis il classe cette masse d'informations et reconstruit le déroulement du match en langage informatique. Ensuite, il va puiser son vocabulaire dans une base de données contenant une liste de phrases, d'expressions toutes faites, de figures de style et de mots-clés revenant fréquemment dans la presse sportive. Il va alors rédiger un article, sans fautes de grammaire ni d'orthographe.

Il peut fournir plusieurs versions, rédigées dans un style plus ou moins imagé («L*es Minnesota Twins ont: perdu/reçu une sévère correction/esquinté leurs battes en pure perte…*») ou encore deux articles adoptant le point de vue de l'une ou l'autre équipe. Il ira même chercher sur Internet les photos des principaux joueurs. Le tout en deux secondes chrono, qui dit mieux? Le rêve de tout chef de service: un journaliste rapide, pas cher, sans états d'âme.

Stats Monkey a été imaginé par les professeurs Larry Birnbaum et Kris Hammond, spécialistes d'intelligence artificielle. Puis son développement a été confié à John Templon, 27 ans, diplômé de journalisme, et Nick Allen, 25 ans, informaticien. M. Allen estime que le but est quasiment atteint: «*Les articles écrits par The Machine sont très proches des dépêches sportives de l'agence Associated Press (AP), qui sont souvent reprises telles quelles par les journaux.*»

La première version de la liste de phrases clés a été réalisée

manuellement, mais, à l'avenir, Stats Monkey l'enrichira automatiquement, en décortiquant de gros volumes d'articles écrits par des humains. Il pourra même imiter le style d'écriture de tel ou tel journaliste connu.

Une version commerciale de Stats Monkey sera bientôt accessible en ligne. Kris Hammond vise en priorité les journaux locaux et les sites web de sport, qui n'ont pas les moyens de payer des pigistes pour écrire les comptes rendus de tous les matches de leur région: «*Il y a aux États-Unis 160 000 équipes scolaires de base-ball qui n'intéressent pas les journalistes, mais qui passionnent des millions de gens.*»

Infolab a l'intention d'adapter Stats Monkey à d'autres sports, notamment le football et le basket-ball. Il souhaite également se lancer dans le secteur de la finance et de la bourse – où, là aussi, les journalistes utilisent massivement un nombre assez limité d'expressions toutes faites. A nouveau, Kris Hammond parle chiffres: «***Cinquante-quatre mille sociétés américaines sont cotées en bourse, et chacune d'elles doit publier des données chiffrées, qui intéressent les actionnaires, les employés, les clients… Or, à peine 3000 d'entre elles sont suivies par la presse économique.***»

Reste une question épineuse: si l'on met en place un système efficace et bon marché pour couvrir les matches locaux et la vie des PME, pourquoi ne pas étendre peu à peu son usage aux rencontres importantes et aux grosses entreprises? M. Hammond a une réponse toute faite: «*Notre but est juste de fournir aux journalistes des outils qui les débarrasseront des tâches les plus répétitives et les moins intéressantes. Ils dégageront ainsi du temps pour accomplir leurs missions nobles: reportages de terrain, investigations, analyses…*»

Au même étage, trois chercheurs mettent au point un système expérimental baptisé News at Seven, qui fabrique des mini-journaux télévisés pour Internet, présentés par Zoe et George, deux personnages animés. Le spectateur se contente de choisir

trois thèmes d'actualité – par exemple politique intérieure, basket-ball et nouveau film – News at Seven se charge du reste. Il parcourt une série de sites d'informations pour trouver des textes pertinents, qu'il raccourcit. Puis il les envoie vers un logiciel de synthèse vocale, qui crée deux fichiers audio – une voix d'homme, et une de femme. Les textes sont aussitôt dits à l'écran par Zoe et George.

Pour les critiques de films, News at Seven apprend à faire le tri entre les articles élogieux et négatifs, grâce à un dictionnaire de mots-clés. Et il cherche sur Internet des vidéos pouvant illustrer les thèmes choisis, et les insère dans l'émission.

Dès le lancement des projets, les responsables d'Infolab avaient poussé les jeunes chercheurs à aller faire des stages de formation à l'école de journalisme Medill, rattachée à l'université. Nathan Nichols, diplômé d'informatique travaillant sur News at Seven, se souvient qu'au début la collaboration n'était pas idéale: «Des étudiants demandaient à leurs profs: faut-il vraiment aider ces gens à détruire nos futurs emplois? Et certains profs semblaient assez d'accord avec eux.» Pour combler ce fossé, Infolab et Medill ont créé en 2009 un organisme commun d'enseignement et de recherche, le Centre d'innovation en technologie, médias et journalisme, qui va accueillir des étudiants venus des deux écoles et leur apprendre à travailler ensemble.

Le rapprochement se fait aussi avec la presse. Bill Adee, directeur du département numérique du Chicago Tribune, est venu à Evanston pour étudier une éventuelle coopération avec Infolab sur des projets à venir, notamment des outils de veille pour repérer les sujets qui agitent Internet. Il a aussi invité plusieurs fois MM. Birnbaum et Hammond à la rédaction du Chicago Tribune: «*Je leur donne des conseils pratiques, en adoptant le point de vue du journaliste. Ça les aide à concevoir des outils qui nous seront réellement utiles.*»

M. Adee ne s'intéresse pas particulièrement à Stats Monkey, mais il sait que, face à ce rapprochement, certains de ses confrères pourraient se sentir menacés: «*Dans tous les journaux, il y a des gens qui passent leur temps à écrire des comptes rendus de matches. J'espère que, si on leur en offre la possibilité, ils seront capables à l'avenir de faire autre chose.*» De même, Larry Birnbaum est conscient de l'impact de ses inventions: «*Nous sommes en train de créer un paysage médiatique que nous ne comprenons pas encore, mais nous savons déjà que l'organisation économique des médias devra s'y adapter. Le défi sera d'intégrer les valeurs classiques du journalisme dans ces nouveaux outils.*»

En attendant, d'autres équipes travaillent sur une demi-douzaine de projets qui viendront compléter la panoplie d'Infolab. Le chercheur Francisco Iacobelli construit ainsi un système intelligent baptisé Tell Me More. Il commence par mémoriser un article politique publié sur CNN. com, puis il trouve d'autres articles traitant du même sujet, publiés par AP, Reuters, le Chicago Tribune. Si leur contenu est identique, il les rejette. En revanche, s'ils contiennent des informations supplémentaires (noms de personnes ou de lieux, chiffres, citations), il extrait les paragraphes concernés. Dans un second temps, Tell Me More va composer un nouvel article plus long et plus riche à partir du texte de CNN, auquel il aura incorporé, aux bons endroits, les phrases pertinentes tirées des autres articles. M. Iacobelli a testé ses méta-articles sur un panel de lecteurs: «*Ils ne voient pas la différence avec un texte écrit par un seul auteur. Il faut dire que, très souvent, les journalistes sautent d'un sujet à l'autre sans transition.*» Combiné à Stats Monkey, on imagine ce que pourrait donner Tell Me More…

Aucun aspect du journalisme n'échappe à Infolab. Patrick McNally, étudiant-chercheur, met au point un système de fabrication de bandes dessinées appelé Manatee Comics. Son but est de démonter, reproduire et automatiser le mécanisme des

plaisanteries basées sur une comparaison, une chute inattendue, un paradoxe…

A ce jour, les résultats sont assez déroutants, mais M. McNally semble sûr de lui: «*Je vais prouver qu'une machine peut générer du contenu humoristique de façon robuste et régulière.*» A terme, toute intervention humaine sera éliminée: Manatee Comics choisira le sujet de sa BD du jour en allant consulter Google pour connaître les événements les plus recherchés et les plus commentés par les internautes.

Le Temps

[youtube]http://www.youtube.com/watch?v=ZEPq0FvFm3g[/youtube]
(Merci à Gilles Bates pour la vidéo)