

Le bon côté des **DONNÉES DE MASSE**

Christopher W. Surdak et Sara Agarwal

UN AGRICULTEUR se sert d'un téléphone portable pour consulter le cours des pommes de terre, puis trouver l'acheteur au meilleur prix : voilà un exemple courant de la façon dont la technologie peut stimuler le progrès social et économique.

Mais la technologie actuelle nous permet d'aller plus loin. Avec un téléphone intelligent et des applications simples, les acheteurs peuvent décompter le nombre d'agriculteurs qu'ils voient par jour, et les agriculteurs peuvent calculer combien ils gagnent par an. À partir des informations recueillies en un an, grâce à l'analyse prédictive, une application peut indiquer aux agriculteurs quels sont le meilleur jour et le meilleur endroit pour vendre leur marchandise. Des applications permettent de croiser les informations sur la demande et les conditions météo, et l'agriculteur sait ainsi quels sont les meilleurs jours pour semer et récolter. Les agriculteurs peuvent suivre la courbe des précipitations et connaître la condition du sol, comprendre comment cela affectera leurs cultures, être prévenus en cas d'alerte à la maladie des cultures et savoir comment s'en prémunir. Grâce à l'exploitation des données de masse, ils peuvent prendre des décisions mieux éclairées, avec à la clef des gains d'efficacité et de rentabilité.

L'expression «données de masse» renvoie à la collecte, l'analyse et l'exploitation d'une très grosse quantité d'informations. Il existe déjà quantité d'outils analytiques, mais les données de masse permettent de découvrir des liens jusque-là insoupçonnés à l'aide de vastes banques de données parfois sans rapport. Et on peut le faire via des infrastructures de partage en ligne, afin que n'importe qui puisse en faire usage.

Comme toutes les technologies, les données de masse ont leurs inconvénients. Elles peuvent améliorer la vie des gens, mais elles peuvent aussi servir à les contrôler, les manipuler et les exploiter. La plupart des organisations utilisent ces informations avec de bonnes intentions. D'autres pas. L'inquiétude grandit quant à la manière dont les données personnelles sont utilisées pour influencer

les comportements, et de nombreux gouvernements prennent conscience de la nécessité d'en réglementer l'usage.

Pour que cette technologie tienne ses promesses dans les pays en développement, il faut révolutionner à la fois son usage et son application. Des changements considérables sont nécessaires sur les deux fronts pour tirer tout le parti possible des données au service du développement. Les gouvernements, ainsi que les

L'analyse des données peut être mise au service de la croissance dans le monde en développement



Un agriculteur bangladais utilise son téléphone intelligent.

organisations non gouvernementales et les donateurs et fondations qui les soutiennent ont un rôle crucial à jouer.

L'étude des données de masse amène à poser les bonnes questions. Chaque fois que vous utilisez votre téléphone ou affichez un message sur un réseau social, vous produisez des données qui s'ajoutent au déluge d'activité numérique autour de vous. Les analystes peuvent les récupérer et, au lieu de simplement se demander *ce qui s'est passé*, ils peuvent se demander *pourquoi*. Pourquoi telle action? Pourquoi tel effet? Les réponses peuvent être très révélatrices.

Cherchant à comprendre pourquoi il y a plus de problèmes de santé mortels par habitant dans une région donnée, l'analyste peut découvrir qu'il y a moins de gens qui consultent un médecin pendant certaines périodes de l'année. Le croisement avec d'autres données fait apparaître une corrélation entre les visites médicales et la baisse du niveau des rivières pendant cette période. Lorsque les eaux baissent, les ennemis du pays voisin peuvent traverser la rivière pour atteindre l'hôpital. Craignant pour leur vie, les patients évitent l'hôpital, même s'ils sont malades. L'ouverture d'un dispensaire dans une autre zone serait une solution nouvelle à un problème ancien.

Des solutions pratiques

HP Earth Insights est un programme conjoint de Conservation International et Hewlett-Packard, qui améliore la collecte, l'analyse et la diffusion des données sur la végétation, les espèces, les stocks de carbone et le climat. Auparavant, les scientifiques passaient des semaines à étudier des données disparates et un nombre incalculable de photos d'animaux sauvages apparemment sans liens entre elles recueillies dans 17 forêts tropicales et 15 pays. Aujourd'hui, grâce à des logiciels d'analyse des données de masse, qui permettent l'intégration de différentes séries ainsi que des calculs et simulations, ils peuvent rapidement identifier des tendances et mieux comprendre comment la déforestation et le changement climatique causent la déperdition des espèces. On peut alors rapidement proposer des solutions adaptées et voir ce qui donne les meilleurs résultats.

Kreditech est une société privée qui se sert des données de masse pour évaluer les antécédents de crédit des emprunteurs potentiels sur la base d'informations à la disposition du public. C'est intéressant, parce que, dans les pays en développement, peu de gens ont un dossier auprès d'un bureau de crédit traditionnel, et il leur est donc difficile d'obtenir un prêt à des conditions raisonnables. Kreditech peut promouvoir l'octroi de prêts non garantis à des candidats précédemment rejetés. Si, par exemple, les amis du client en puissance sont solvables, ce serait une corrélation positive avec la qualité de sa signature. Jusqu'à présent, le taux de remboursement de Kreditech sur certains marchés est de 93 %, ce qui est élevé et suggère que ses critères d'appréciation sont non seulement corrects, mais plus efficaces que les indicateurs traditionnels. S'il est possible de se passer des bureaux de crédit traditionnels, cela signifie qu'il y a des millions de personnes qui pourraient bientôt avoir accès au crédit.

La compagnie française de télécommunications Orange présente dans plusieurs pays et la fondation suédoise sans but lucratif Flowminder sont d'autres exemples de l'influence des données de masse sur le développement; l'une et l'autre utilisent les communications vocales et écrites depuis des téléphones portables pour suivre les mouvements de population de manière à savoir où il serait préférable d'établir des centres de traitement contre le virus Ebola.

Les données de masse peuvent améliorer l'analyse des problèmes de développement, mais seulement si des esprits brillants se penchent sur ces questions et prennent la tête des efforts pour résoudre les problèmes économiques et sociaux. L'information contextuelle permet de démêler les découvertes simplement intéressantes de celles qui permettent une action de terrain. Il ne suffit pas de faire venir des spécialistes de l'étranger. Les gens du cru qui connaissent les problèmes et la culture sont mieux placés pour analyser les données. Les bailleurs de fonds et les gouvernements doivent donc soutenir et promouvoir la formation d'analystes sur place. En contrepartie, les décideurs doivent être prêts à tenter l'expérience des solutions créatives et audacieuses que permet l'exploitation des données de masse.

Piqûre de rappel technologique

Beaucoup de gens se demandent si les données de masse peuvent combler les lacunes des sources d'information traditionnelles. Nous pouvons déjà beaucoup apprendre des données disponibles dans les pays en développement grâce aux téléphones, aux compteurs électriques et aux satellites (par exemple sur les habitudes de voyage ou l'activité économique). Mais on pourrait faire beaucoup mieux avec des infrastructures mieux adaptées.

Il est possible de recueillir plus de données avec une version améliorée du téléphone cellulaire qu'ont déjà la plupart des citoyens des pays émergents. Tandis que de plus en plus de personnes remplacent leur cellulaire par un téléphone intelligent, la quantité de données pertinentes grossira. Les applications conçues pour ces téléphones encouragent l'interactivité, un pactole pour les analystes. La promotion de l'usage des téléphones intelligents pour résoudre les problèmes dans des secteurs (santé ou agriculture) ou des endroits précis (villes d'une certaine taille) pourrait se traduire par une augmentation géométrique de la quantité et de la qualité de l'information.

Les bailleurs de fonds et les pouvoirs publics rechignent à investir dans les téléphones intelligents à l'usage des gens qui n'en ont pas les moyens, parce qu'ils sont considérés comme des articles de luxe. Mais, compte tenu de la valeur pour la société des données produites, les gouvernements devraient envisager de subventionner soit le coût de ces appareils, soit les abonnements. Il est possible de récupérer ces coûts en revendant de manière responsable ces données anonymisées aux entités qui souhaitent les exploiter commercialement. Dans le monde de la collecte d'information, les téléphones intelligents ne sont pas un luxe, ils sont les derniers éléments d'une infrastructure publique essentielle.

Dès lors que les téléphones intelligents seront plus répandus, il faudra créer de nouvelles applications pour produire les informations locales pertinentes. Des applications qui localisent la source d'eau potable la plus proche ou signalent un début d'épidémie de choléra ou d'Ébola généreront d'autres données qui permettront d'affiner les analyses et de répondre à des questions auxquelles nous n'avons même pas songé. En fait, nous ne saurons pas jusqu'où il est possible d'aller tant que ces applications ne seront pas utilisées.

Les applications créées et gérées sur place par des gens qui connaissent les besoins de leurs voisins seront cruciales. Les concepteurs locaux joueront un rôle essentiel, et il faut encourager leurs initiatives. La multiplication des téléphones intelligents créera un marché pour la conception d'autres applications, et les pouvoirs publics, les bailleurs de fonds et les ONG peuvent fournir des formations, des locaux et d'autres formes de soutien. Il faut considérer le travail

des concepteurs de logiciels comme le prolongement de celui des instituts officiels des statistiques : ils jouent un rôle aussi vital dans la collecte de l'information.

Certains craignent que les données de masse soient utilisées dans les pays émergents et en développement pour consolider les régimes autoritaires. Il peut certes y avoir des violations de la vie privée si les données recueillies sont employées à mauvais escient. Mais c'est l'emploi de la technologie, et non la technologie elle-même, qui détermine si c'est une bonne ou une mauvaise chose.

Les règles de confidentialité et autres garde-fous sont nécessaires pour donner confiance dans la collecte des informations et garantir leur bon usage. Les données recueillies sont actuellement aux mains des opérateurs de téléphonie mobile, des pouvoirs publics, des concepteurs d'applications et des gérants des réseaux sociaux. Il faut ouvrir l'accès à ces données après les avoir agrégées et anonymisées, et l'État doit définir des règles pour que cela puisse se faire (même pour les fournisseurs privés) lorsqu'il y a un besoin légitime.

Optimisation des moustiquaires

Certains diront que les données de masse et les investissements qu'elles nécessitent ne sont pas aussi importants pour les pays en développement que les moustiquaires, par exemple. Et si l'exploitation de ces données permettait d'améliorer la distribution des moustiquaires? C'est déjà le cas. Le système d'aide à la décision aWhere's Mosquito Abatement Decision Information System exploite une combinaison d'images-satellite, de modèles météorologiques et de recherches sur la biologie des moustiques pour déceler les risques épidémiques potentiels avant même l'éclosion

des œufs de moustiques, ce qui aide les fournisseurs et utilisateurs de moustiquaires à les placer là où ils seront le plus utiles.

Cet exemple montre comment l'utilisation plus efficace des données peut soutenir les efforts de développement actuels en optimisant l'emploi des ressources rares. Les données de masse ne sont pas seulement un outil de sensibilisation ou de prévision plus précis; elles servent à faire plus avec moins.

Pour mettre l'analyse au service de la croissance des pays en développement, il faut passer par les étapes suivantes : investir dans les bonnes technologies et la formation, soutenir les entrepreneurs et faire évoluer les mentalités, et fixer des règles concernant le respect de la vie privée et les données libres. Il est préférable de mener ces initiatives de front, faute de quoi les lacunes dans un domaine empêcheront la réalisation de l'ensemble.

Les données de masse peuvent et vont trouver leur utilité dans les pays en développement, mais la rapidité et l'efficacité avec laquelle cela se produira dépendent de ce que nous faisons aujourd'hui. Pour en réaliser pleinement le potentiel, nous devons établir de nouvelles sources d'information et de nouveaux modes de pensée sur la manière de les utiliser. C'est seulement à ce prix que nous pourrons utiliser ce puissant nouvel outil pour rehausser la croissance économique du monde en développement. ■

Christopher W. Surdak est l'auteur de Data Crush: How the Information Tidal Wave Is Driving New Business Opportunities et travaille dans le domaine de la découverte et de l'analyse électronique; Sara Agarwal est Directrice de Hewlett-Packard Company pour les organisations de financement international.



PROGRAM IN ECONOMIC POLICY MANAGEMENT (PEPM)

Confront global economic challenges with the world's leading economists, policymakers, and expert practitioners, including Jagdish Bhagwati, Guillermo Calvo, Robert Mundell, Arvind Panagariya, and many others.

A 14-month mid-career Master of Public Administration focusing on:

- rigorous graduate training in micro- and macroeconomics
- emphasis on the policy issues faced by developing economies
- option to focus on Economic Policy Management or International Energy Management
- tailored seminar series on inflation targeting, international finance, and financial crises
- three-month capstone internship at the World Bank, IMF, or other public or private sector institution

The 2015–2016 program begins in July of 2015. Applications are due by January 5, 2015.

pepm@columbia.edu | 212-854-6982; 212-854-5935 (fax) | www.sipa.columbia.edu/academics/degree_programs/pepm
To learn more about SIPA, please visit: www.sipa.columbia.edu