

● **DES MÉTHODES PROBABILISTES ÉPROUVÉES** permettent, si nécessaire, de reconstituer des données manquantes. Corollaire : il n'est pas toujours nécessaire de collecter autant de données qu'on le croit pour parvenir à un résultat voulu.

PAR BERNARD BEAUZAMY

PDG DE LA SOCIÉTÉ DE
CALCUL MATHÉMATIQUE SA

● Polytechnicien,
docteur ès sciences
en mathématiques,
58 ans, il est le fondateur
de la Société de calcul
mathématique SA.



ET OLGA ZEYDINA

INGÉNIEUR
DE RECHERCHE

● Diplômée en mathématiques
de l'université de Donetsk
(Ukraine),
23 ans, est ingénieur de recherche
à la Société de calcul
mathématique SA.

DONNÉES MANQUANTES, PERDUES, INUTILES...

Toutes les entreprises, un jour ou l'autre, sont confrontées à ce problème irritant : on ne dispose pas des données qui seraient nécessaires pour prendre une décision. Tantôt il s'agit des chiffres de vente : ils sont connus ici, mais absents là ; tantôt il s'agit de l'état des stocks, pour lequel des éléments font défaut ; tantôt il s'agit d'évaluer les rebuts de production, pour lesquels on ne dispose pas de statistiques précises, etc. Dans certains cas, on a enregistré les données, et les enregistrements ont été perdus, ou bien on n'y avait pas pensé, ou bien on n'avait pas d'argent : toutes les raisons habituelles peuvent être invoquées.

À l'inverse, certaines entreprises collectent absolument tout : le nombre de pièces sorties de chaque usine, par jour, par taille, par couleur ; le nombre de boîtes vendues, le nombre d'incidents de paiement, les débits mesurés par des capteurs fonctionnant toutes les secondes, etc. Au bout de la journée, elles croulent sous des téraoctets de données dont, en définitive, elles ne savent absolument pas quoi faire.

On pourrait penser que la récolte des données doit être en relation avec l'usage que l'on veut en faire : c'est du bon sens. Par exemple, le niveau d'un stock doit être utilisé à des fins comptables, à des fins statistiques et à des fins de réapprovisionnement : il suffit généralement de données journalières. Mais l'expé-

rience prouve que lorsqu'on commence à mesurer quelque chose, cela donne une idée à un petit malin qui réclame aussitôt des compléments. Les ventes d'un magasin, si elles sont relevées toutes les heures, permettent de connaître les habitudes des clients, d'adapter les ressources en personnel et les incitations commerciales.

Des méthodes robustes et grossières

On ne peut donc pas dire qu'il existe un format unique, précis, optimal. Le mieux est de savoir s'adapter et de disposer d'outils flexibles, permettant à la fois la reconstitution des données manquantes, la mise en évidence des données aberrantes et la validation croisée (comparaison) des données recueillies.

Des méthodes probabilistes permettent d'at-

teindre ces objectifs. Ce sont, par définition, des méthodes robustes et grossières, qui reposent sur des corrélations, des comparaisons entre indicateurs. On comparera deux centres de production A et B, et on établira des tables du type : sachant que A vend telle quantité de tel produit, voici la répartition des ventes de B en ce cas. C'est ce qu'on appelle des tables de "probabilité conditionnelle". À partir de ces tables, on voit immédiatement si les liens entre A et B sont forts ou non. S'ils sont forts, on peut reconstituer une donnée manquante pour B, pourvu que A existe, en se servant de ces tables : on prendra la moyenne des valeurs observées pour B, lorsque A a indiqué une situation similaire.

Si on constate que la connaissance de A se traduit par une bonne connaissance de B (forts liens), on peut de la même manière cesser d'enregistrer B et le reconstituer à partir de A : on réalise ainsi une économie.

La philosophie générale qui se dégage de toute réflexion à propos des données est qu'il s'agit d'un problème irritant, parce que quotidien. Il est utile d'y consacrer un peu de temps : on s'aperçoit alors que la bonne approche doit être flexible et robuste, parce qu'on ne sait jamais exactement de quoi l'avenir sera fait. ●

**« IL N'EXISTE PAS
UN FORMAT UNIQUE,
PRÉCIS, OPTIMAL.
LE MIEUX EST
DE SAVOIR S'ADAPTER
ET DE DISPOSER
D'OUTILS FLEXIBLES. »**

Bernard Beauzamy et Olga Zeydina :
"Méthodes probabilistes pour la reconstruction
de données manquantes - Manuel théorique
et pratique", éditions de la Société de calcul
mathématique SA (www.scmsa.com)