

ENQUÊTE Le « data mining » séduit un nombre croissant d'entreprises.

Des logiciels pour prédire l'avenir

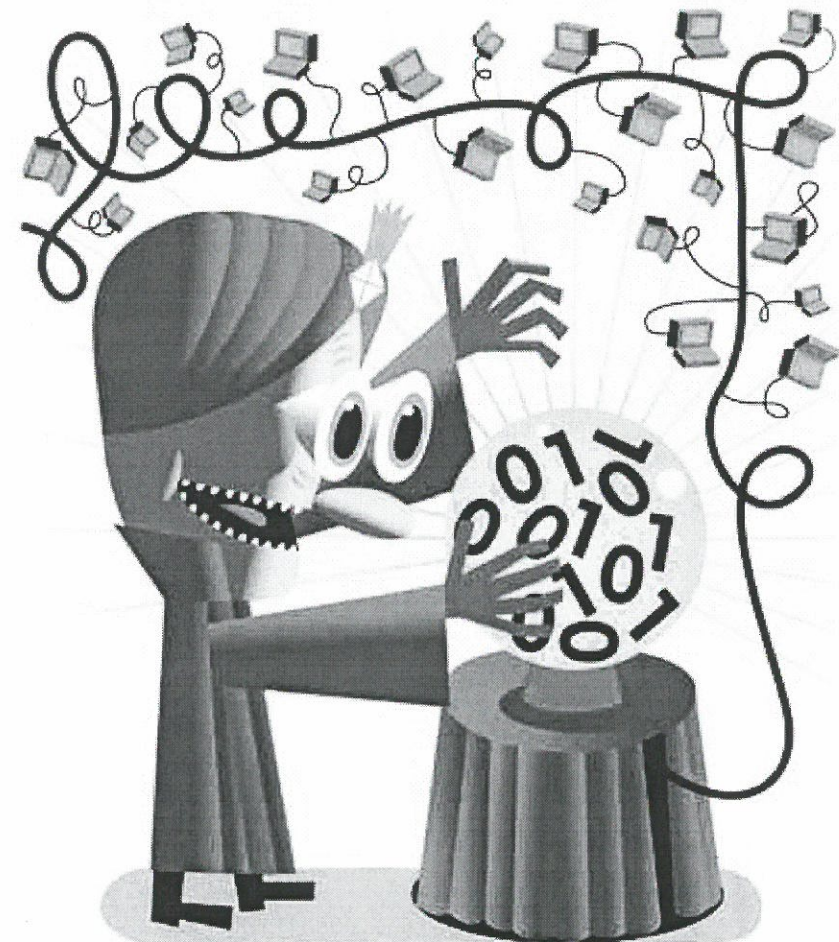
En permettant aux entreprises d'anticiper les attentes de leurs clients, les logiciels de « data mining » ont révolutionné le marketing. Pour l'instant, ces outils d'analyse de données s'intéressent surtout aux informations chiffrées, mais de nouvelles versions fouillent déjà les écrits des consommateurs. Prochain objectif : le Web.

La statistique et le traitement informatique des données (STID) ont le vent en poupe. Derrière cette expression un peu austère se cache un ensemble de techniques étonnantes, qui promettent de donner du sens aux montagnes de données qui s'accumulent dans les ordinateurs. D'ailleurs, les conséquences sont déjà observables sur le marché du travail. « Les jeunes qui ont suivi la formation de ce type sont très demandés cette année », se félicite Laurence Duval, responsable de la filière « systèmes d'information statistique » à l'Ensaï (Ecole nationale de la statistique et de l'analyse de l'information).

Jusqu'à présent, l'administration constituait le principal débouché, avec quelque 10.000 spécialistes travaillant pour l'Etat. Aujourd'hui, les entreprises sont en passe de prendre le relais. Elles embauchent pour trier les gigantesques masses de données qu'elles ont accumulées sur leurs clients. Mais plus question de photographier le passé en établissant de simples récapitulatifs des ventes. Désormais, les experts en STID doivent prédire... l'avenir !

Prévoir les pannes

Chez Bouygues, Orange ou SFR, on scrute le profil de chaque abonné (modèle de téléphone, carte ou abonnement, options, consommation...) pour savoir s'il risque de passer à la concurrence et il peut monter en gamme. La route et les 3 Suisses n'en voient plus leurs coûteux catalogues qu'à certains prospects - ceux dont il semble quasi certain qu'ils achèteront un article dans les prochains mois. Et des laboratoires pharmaceutiques connais-



sent les médicaments qui seront des best-sellers dans telle officine, à telle période, ce qui leur permet de personnaliser les prospectus envoyés aux pharmacies.

Les industriels s'y mettent aussi. Quelques entreprises analysent les comptes rendus de leurs agents de maintenance pour prévoir les pannes. D'autres ont pris le problème à l'envers et utilisent des outils de STID pour fabriquer des produits plus fiables (lire page suivante).

L'engouement pour ces techniques d'analyse de données prédictive ne touche pas seulement le secteur privé. L'Etat s'y intéresse de très près. Ainsi, la loi antiterroriste adoptée par le législateur le 22 décembre dernier prévoit d'identifier les voyageurs suspects à partir des informations contenues dans les serveurs des compagnies aériennes ou ferroviaires.

A l'origine de cette révolution, l'automatisation de l'analyse des données et la baisse du prix des équipements informatiques. Longtemps, la statistique est demeurée une science réservée aux mords de variance ou de tests non paramétriques. Pour comprendre les variations des données, et donc les comportements des consommateurs, il fallait émettre des hypothèses, puis tester ces dernières, une à une. Fastidieux. A la fin des années 1980 sont apparus les premiers logiciels de « data mining » (littéralement « fouille de données »), qui tournaient sur de gros ordinateurs.

« Avec ces programmes, vous

ne vous contentez plus de vérifier des hypothèses issues des a priori ou de l'expérience du passé : la machine vous aide à formuler intelligemment les hypothèses colant le mieux aux données, explique Stéphane Tufféry, en charge de ces techniques dans un grand groupe bancaire français, enseignant à Dauphine et auteur de « Data mining et statistique décisionnelle » (Editions Technip). Conduite par votre savoir-faire, elle va trouver ce que vous recherchez sur de grands volumes de données. »

Domination américaine

Depuis, les capacités de stockage et de traitement des ordinateurs ont été, à prix constant, décuplées. « Aujourd'hui, un ordinateur portable permet de traiter rapidement une base de données contenant des centaines de variables sur un million de clients », poursuit Stéphane Tufféry. Résultat, l'analyse prédictive est à la portée de n'importe qui, ou presque. Dans l'Etat du Queensland, en Australie, les pompiers peuvent savoir, en quelques secondes, si l'incident (feu de voiture, blessure par balle...) pour lequel ils ont été appelés risque de dégénérer : 400.000 rapports d'intervention ont été analysés par un logiciel de data mining.

Ces innovations tombent à point nommé, car les entreprises croulent sous l'information. « On estime qu'en 2001, plus de données ont été produites que dans toute l'histoire de l'humanité. Et,

chaque année, cette masse augmente encore : par exemple, le moindre échange entre une firme et ses clients laisse désormais une trace », rappelle Thierry Vallaud, responsable du département « analyse de données » chez Arvato Information Services, qui vient d'adapter pour la France un livre de Daniel Larose, un des grands spécialistes américains du data mining.

Tous les grands noms de l'informatique (IBM, Microsoft, Oracle...) proposent des solutions. Mais sur ce marché évalué, dans le monde, à quelque 3 milliards de dollars par an, trois éditeurs américains ont réussi à se faire un nom : Insightful (19 millions de dollars de chiffre d'affaires en 2004), SAS (1,53 milliard de dollars) et SPSS (250 millions). Coût d'un logiciel ? 80.000 euros en moyenne. « L'application la plus classique de nos outils est la CRM, la gestion de la relation client », décrit Christine Le Bras, en charge du développement commercial de ces produits chez SAS France. « Nos logiciels se branchent sur les bases de données commerciales des entreprises et permettent, pour chaque client, d'évaluer son score d'appétence - la possibilité qu'il achète un produit -, son score d'attrition - le risque qu'il change de fournisseur -, ainsi que le segment de consommation auquel il appartient, détaille Hervé Mignot, directeur technique chez SPSS. Il est alors possible de bâtir des programmes marketing cohé-

rents qui permettent de fidéliser les acheteurs. »

Les banques, suivies par les entreprises de vente par correspondance, puis par les assurances, ont été les premières à adopter cette approche. Les compagnies aériennes, les constructeurs automobiles, les opérateurs de téléphonie et les stations-service leur ont emboîté le pas. Renault a ainsi construit un modèle statistique identifiant les clients susceptibles de changer de véhicule dans les six mois.

Les textes après les chiffres

Aujourd'hui, c'est au tour des fabricants d'appareils électroniques, de la grande distribution (Carrefour, Super U...) et des laboratoires pharmaceutiques. Et les banques, dans le cadre de la réforme Bâle II, sont en train de mettre en place de nouveaux outils d'évaluation des risques (lire page suivante).

L'analyse prédictive n'en est cependant qu'à ses débuts. Les prochaines étapes devraient être la généralisation du « text mining » et du Web sémantique (lire page suivante). « 80 % des données des entreprises ne sont pas des données chiffrées, mais des données textuelles : les lettres des clients, les courriers électroniques internes, les verbatim des centres d'appels, etc. », estime Christine Le Bras. Déjà, Hewlett-Packard utilise le text mining pour améliorer la qualité de service dans ses call centers.

Plus inattendu, NRJ et Skyrock s'en servent pour cerner les attentes de leurs auditeurs. L'analyse porte en fait sur le contenu des blogs mis en ligne par les jeunes sur les sites communautaires de ces radios. « Vous y trouvez deux types de données. Les informations explicites sont communiquées par les jeunes lorsqu'ils s'inscrivent, répondent à des enquêtes ou publient leurs textes. L'implicite, c'est ce qu'ils font sur le site : les pages qu'ils visitent, les pubs sur lesquelles ils cliquent », révèle Prashant Nair, cofondateur d'Avedya, une start-up spécialiste des communautés interactives et prestataire d'AOL, NRJ, Orange...

Pour l'instant, la CNIL (Commission nationale de l'information et des libertés) ne s'oppose pas à ces techniques. « Le tri des clients doit respecter la loi Informatique et Libertés, c'est-à-dire être déclaré et prévoir un droit d'information, d'accès, d'opposition et de rectification pour les personnes concernées », rappelle Sophie Nerbonne, de la direction des affaires juridiques de la CNIL.

JACQUES HENNO

SUITE DE L'ENQUÊTE

- Le Web, une mine d'informations encore mal exploitée
- L'analyse prédictive au secours des banquiers
- L'industrie fait parler ses lignes de production