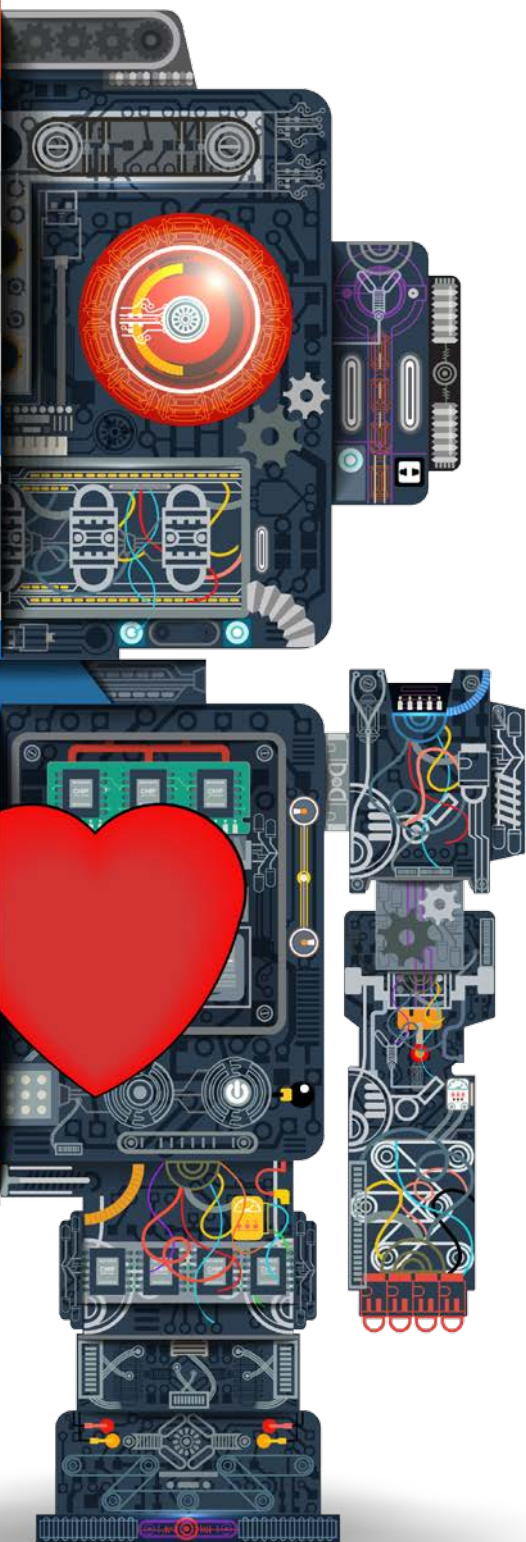




L'ESSENTIEL

■ Les gens interagissent de plus en plus souvent avec des machines commandées par la voix.

■ Ils attendent d'elles spontanément qu'elles saisissent leurs émotions, y compris les plus complexes, tels l'humour ou la moquerie.

■ C'est pourquoi des chercheurs travaillent à doter les intelligences artificielles de logiciels capables d'extraire des émotions de la voix et des mots.

■ Même si ces recherches débutent, de premiers robots empathiques existent et s'améliorent de jour en jour.

fabriqué par la société française Aldebaran Robotics afin d'accueillir les visiteurs dans les points de vente de l'opérateur téléphonique japonais Softbank Mobile. Autre cas, Jibo est un assistant personnel domestique de 2,7 kilogrammes. La « robotique empathique » n'en est encore qu'à sa naissance, mais les outils et les algorithmes qui amélioreront spectaculairement les machines en les dotant d'une « empathie artificielle » de qualité sont en cours de développement.

C'est ce sur quoi travaillent les équipes du laboratoire que je dirige à l'université des sciences et technologies de Hong Kong. Mon intérêt pour les robots empathiques est né il y a six ans, lorsque j'ai lancé mon équipe dans la mise au point de l'équivalent chinois de Siri. D'emblée, la façon dont les utilisateurs développent naturellement des réactions émotionnelles vis-à-vis des assistants personnels et leur frustration lorsque ces machines n'arrivent pas à les comprendre m'a fascinée. J'ai vite pris conscience que la clé de la conception de machines empathiques est la mise au point d'algorithmes efficaces de reconnaissance de la parole. Or je venais justement de passer vingt-cinq ans de ma carrière à en développer...

Le module d'empathie

Toute machine intelligente est, par essence, un ensemble de composants logiciels, tels des bases de données, des fichiers de configuration, des programmes... Chacune des composantes d'un tel « système logiciel » remplit une fonction unique. Ainsi, un robot intelligent pourra être doté d'un module de traitement de la parole, d'un programme de reconnaissance des objets présents dans les vidéos qu'il a enregistrées, etc. Toutefois, le « cœur » d'un robot empathique sera le composant logiciel conçu pour analyser les signaux faciaux et acoustiques des humains, puis le contenu du discours afin d'en dégager les émotions humaines... C'est ce « module d'empathie » qui traitera les informations émises par les humains et dictera au robot ses réactions.

Ces informations sont complexes. Deux humains qui communiquent envoient et reçoivent de nombreux signaux émotionnels. Qu'ils en aient conscience ou pas, chacun

© Doggygraph / Shutterstock.com

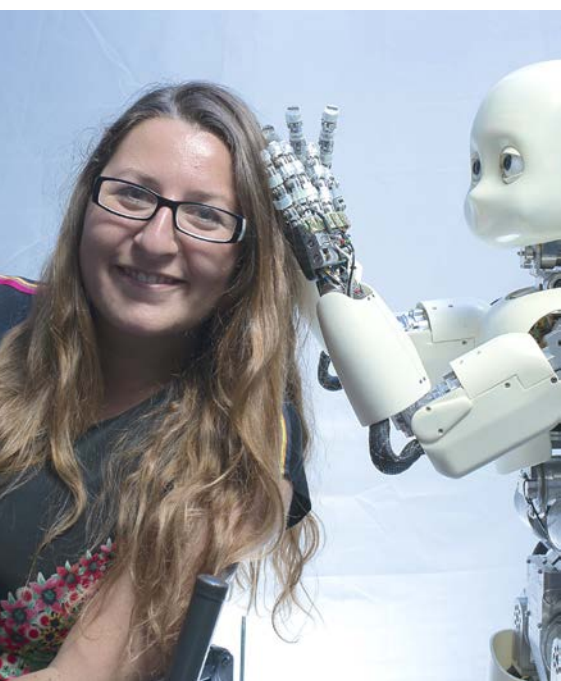
■ L'AUTEURE



Pascale FUNG est professeure d'électronique et de génie informatique à l'université des sciences et technologies de Hong Kong.

LES ROBOTS EMPATHIQUES

seront-ils un jour sympathiques ? La réponse dépendra de la qualité atteinte dans le développement de logiciels d'empathie artificielle et du sens du mot « amitié ». Pour certains, l'amitié avec une machine est par définition impossible. Pour d'autres, elle sera évidente. Comment vivrons-nous avec nos amis les gentils robots ?



les exploite pour saisir l'état émotionnel de l'autre et pour lui faire comprendre le sien. Ces signaux sont notamment les mimiques faciales, tous les éléments du langage corporel (attitude, gestes...), le ton de la voix, mais aussi, évidemment, le contenu du discours. La mise au point d'un module empathique implique donc d'identifier les traits de la communication humaine exploitables par les machines, puis d'« entraîner » des algorithmes d'apprentissage automatique.

Du moins est-ce ainsi que s'y est prise mon équipe pour développer des machines capables non seulement d'identifier les mots et leur sens, mais aussi d'extraire plusieurs caractéristiques fondamentales du signal acoustique qu'est la voix. Ainsi, même si nous la considérons peu sous cet angle, la communication entre humains est aussi une forme de traitement du signal. Le cerveau est entraîné à détecter les émotions convoyées par la voix à partir des caractéristiques des signaux acoustiques qui leur sont associés, qu'il s'agisse de stress, de joie, de peur, de colère, de dégoût... Une personne joyeuse parlera plus vite et sur un ton aigu ; stressée, elle adoptera au contraire une voix plate et sèche. Doté d'un algorithme d'apprentissage automatique convenablement entraîné, un ordinateur peut, comme un cerveau, détecter ce genre de caractéristiques, un peu comme un détecteur de mensonges mesure la tension artérielle, les pulsations cardiaques et la conductivité de la peau.

Pour obtenir une machine ayant, par exemple, la capacité de détecter le stress, nous avons supervisé l'entraînement d'un algorithme en lui faisant analyser des enregistrements de la voix de personnes stressées.

Déceler la détresse dans la voix

C'est ainsi que nous avons commencé à apprendre aux machines à déceler la détresse dans des extraits vocaux d'étudiants de notre université, lesquels s'y connaissent puisque, entre eux, ils surnomment notre établissement l'« université du stress et de la tension de Hong Kong ». Pour obtenir ces extraits, nous leur avons posé une série de 12 questions de plus en plus stressantes et avons obtenu ainsi un premier corpus de signaux vocaux stressés en anglais, mandarin et cantonais. Dix heures d'enregistrement ont suffi pour que nos algorithmes reconnaissent le stress dans 70 % des cas où il est présent

dans la voix. Ce niveau de performance est comparable à celui des auditeurs humains.

Pendant que nous effectuions ces travaux, une autre équipe du groupe apprenait aux machines à reconnaître l'humeur exprimée par la musique d'un morceau de musique de variété, indépendamment des paroles. Contrairement à l'émotion, l'humeur est une ambiance qui reste généralement pendant toute la durée d'un morceau. Nos chercheurs ont commencé par rassembler quelque 5000 morceaux en tous genres dans les principales langues européennes et asiatiques. Les musicologues avaient d'ailleurs déjà classé plusieurs centaines de ces morceaux en 14 catégories d'humeur.

À l'aide d'ordinateurs, nous avons extrait de ces chansons 1000 caractéristiques fondamentales – par exemple l'énergie ou des propriétés acoustiques telles la fréquence fondamentale, les harmoniques... –, puis nous avons utilisé ce corpus pour entraîner 14 programmes « classificateurs », chacun spécialisé dans une humeur particulière. L'un d'eux décèle par exemple l'humeur mélancolique d'une chanson, un autre l'humeur joyeuse, etc. Comme ces 14 classificateurs peuvent être employés en même temps, leurs résultats peuvent être combinés, mais il arrive qu'ils se contredisent, par exemple que le détecteur d'humeur joyeuse trouve joyeuse une chanson mélancolique. Dans un tel cas, nous le réentraînons au cycle d'apprentissage suivant. D'anomalie en anomalie, nous avons ainsi fini par rendre performant le système formé des 14 détecteurs d'humeur. Nous avons ainsi obtenu une machine capable d'extraire l'humeur de n'importe quel morceau de musique juste en « l'écouter ».

Depuis, nous nous sommes appuyés sur ces recherches pour fonder avec d'anciens étudiants la société Ivo Technologies de développement de machines empathiques utilisables à domicile. *Moodbox*, son premier produit, sera un centre de contrôle d'ambiance, qui réglera la musique et la lumière de chaque pièce en fonction des émotions de ses utilisateurs.

Toutefois, pour saisir l'humour, le sarcasme, l'ironie et d'autres caractéristiques complexes de la communication verbale interhumaine, une machine aura besoin de plus que juste un logiciel de reconnaissance des émotions véhiculées dans le son. Il lui faudra aussi comprendre la signification sous-jacente d'un discours et en comparer le contenu avec l'émotion qui le baigne.