

Sélection de variables et modélisation d'expression d'émotion dans les dialogues Homme-Machine

Barbara Poulain

France Télécom R&D, 2 avenue Pierre Marzin 22300 Lannion
barbara.poulain@francetelecom.com

L'émotion s'exprime à travers plusieurs canaux de communication : la parole, la gestuelle, les mimiques faciales et l'haptique (Scherer, 2003). Elle est un axe de recherche important dans les travaux sur la communication et, plus spécifiquement, pour l'amélioration des dialogues homme-machine. Différents types de variables sont susceptibles de détecter ces expressions d'émotions mais toutes ne sont pas construites automatiquement. L'objectif de l'étude est d'aider les experts dans la sélection des variables à prendre en compte pour la détection d'émotion afin d'orienter le travail de construction automatique des variables détectées. Le protocole d'étude répond à ce besoin de sélection de variables tout en prenant en compte leur poids. La sélection des attributs explicatifs est effectuée grâce au sélecteur bayésien naïf ESNB (Boullé, 2006).

Le contexte d'étude est celui de la détection et de la caractérisation des expressions émotionnelles dans la parole d'utilisateurs de services vocaux interactifs. L'intérêt de pouvoir détecter les émotions exprimées relève en partie de la possibilité de réguler le dialogue entre la personne et la machine en cas de problème : par exemple technique (le système ne comprend pas la demande de l'utilisateur) ou relevant de la pertinence de l'information délivrée (le système n'apporte pas de réponse satisfaisante à la demande de l'utilisateur). La présence de problèmes dans le dialogue induit un état émotionnel négatif. Le détecter permet, d'une part, de diminuer les difficultés de compréhension du système et, d'autre part, d'apporter une réponse plus adaptée à la demande et/ou de lui proposer un déroulement différent du dialogue (reprise point par point de la demande ou basculement vers un téléconseiller (Batliner, 2003)). A long terme, on espère pouvoir donner automatiquement le ton adéquat à la machine en réponse à cet état émotionnel, sans intervention humaine.

Le corpus d'étude est constitué de 5406 tours de parole issus de dialogues boursiers entre des testeurs humains et la machine. Un dialogue est une séquence alternative de tours de parole système/utilisateur. Les variables de l'étude sont de natures différentes. Celles liées à l'acoustique sont calculables de façon automatique avec des temps de calculs inégaux. Les experts métiers proposent l'intégration de nouvelles variables d'ordre linguistique, prosodique ou dialogique. Elles sont créées manuellement puis leur pertinence est testée avant d'envisager une création automatique. La difficulté d'obtention des variables (notée entre 0 et 5) est prise en compte pour leur sélection (difficulté de calcul pour les variables définies automatiques et difficulté de calcul estimée pour les variables manuelles). Le meilleur compromis entre performance du modèle et coût de calcul des variables est recherché. On choisit le sélecteur ESNB qui améliore l'approche sélective du bayésien naïf (Langley et Sage, 1994) en exploitant une méthode optimale de discrétisation de Bayes (MODL (Boullé, 2005)) et en évaluant les prédicteurs par maximisation de la surface sous la