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

courbe de lift. On obtient alors une sélection plus fine qu'avec le taux de bonne prédiction habituellement utilisé.

Nous souhaitons minimiser le nombre de variables coûteuses dans la sélection. Nous réalisons alors une série de sélections en ajoutant au fur et à mesure les variables les plus coûteuses. Deux critères de performance sont utilisés: le KI (Kxen Information Indicator) correspondant au rapport entre l'aire sous la courbe de lift et l'aire sous la courbe idéale, et la mesure F1 (2*précision*rappel/ (précision+rappel)).

Les variables V1 de poids 1 apportent de la valeur prédictive aux variables V0 de poids 0, les courbes du KI et de la mesure F1 affichent un premier pic de valeur pour la sélection effectuée sur l'ensemble {V0,V1}. Les courbes stagnent pour les sélections basées sur les sous ensembles {V0,V1,V2}, {V0,V1,V2,V3} mais connaissent une évolution croissante jusqu'à la sélection effectuée sur la totalité des variables. Un modèle performant utilisant des variables coûteuses est-il préférable à un modèle un peu mois performant mais calculable en temps réel ? C'est l'application visée qui doit permettre de répondre.

La variation d'énergie dans la voix, la présence de commentaire dans le tour de parole ou de répétition à l'identique de la réponse de l'utilisateur à une même question du système sont autant de variables qui se révèlent pertinentes. Or, la détection de commentaire est une variable coûteuse qui nécessite un module de reconnaissance de parole extrêmement complexe. La mise en place d'outils de mesure automatique de ces variables apparaissant comme significatives dans la caractérisation d'émotion dans la parole constitue un axe de travail pertinent proposé aux équipes travaillant dans le domaine.

L'étude est poursuivie par la prise en compte de données globales au dialogue (accroissement de l'intensité au cours du dialogue, seuil de tolérance des erreurs,…).

Références

Batliner, A. et Fischer, K. (2003) *"How to find trouble in communication."* Speech Communication 40: 117-143.

Boullé, M. (2005). *A Bayes Optimal Discretization Method for Continuous attributes* Machine Learning, à paraître.

Boullé, M. (2006). *An Enhanced Selective Naive Bayes Method with Optimal Discretization*, Feature extraction, foundations and Applications (Guyon et al, à paraître).

Langley, P. et Sage, S. (1994) I*nduction of selective bayesian classifiers*. In Proceedings 10thConference on Uncertainty in Artificial Intelligence. Morgan Kaufman

Scherer K.R. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, vol. 40, p. 227-256, 2003.

Summary

The study concerns an exploratory process of the selection of variables based on an extension of a selective naïve bayes within the modelling of expressions of emotions in a human-machine oral dialogs context.