

Amélioration des modèles pour une tâche d'association de mots

Sujet de stage au master de sciences cognitives, Grenoble, 2011-2012

Gilles Bisson (LIG – AMA) – Gilles.Bisson@imag.fr

Benoit Lemaire (LPNC) – Benoit.Lemaire@upmf-grenoble.fr

Contexte

Ces 20 dernières années, beaucoup de recherches ont tenté de simuler sur ordinateur la manière dont nous traitons les informations langagières. En effet, un modèle formel capable de rendre compte d'un grand nombre de données expérimentales serait un bon candidat pour décrire la manière dont nous traitons les concepts et les mots. Certains de ces travaux se sont intéressés à simuler le comportement humain dans la tâche d'associations de mots. Cette tâche consiste à demander aux participants de produire des mots associés à un mot cible. Par exemple, si le mot présenté est neige, le participant peut produire blanc ou montagne.

Un bon modèle formel doit être capable de classer les mots associés dans le même ordre que les humains. Plusieurs modèles ont été proposés en sciences cognitives pour résoudre cette tâche. Ils sont basés, soit sur une représentation hiérarchique (Wordnet, ..), soit sur une représentation vectorielle (analyse de la sémantique latente (LSA), modèles géométriques (ACOM), ...). Cependant, dans le domaine de l'apprentissage automatique, des techniques de co-clustering ont aussi été développées pour d'autres objectifs, mais leur plausibilité cognitive n'a jamais été étudiée, bien qu'ils semblent théoriquement aptes à résoudre la tâche d'association sémantique.

Travail demandé

Pendant le stage, le candidat devra comparer plusieurs modèles de la littérature et développer une méthode pour comparer les modèles classiquement utilisés en sciences cognitives et ceux développés en apprentissage automatique, en se focalisant plus particulièrement sur l'algorithme X-SIM développé dans l'équipe AMA du LIG.

En fonction des résultats, le candidat devra proposer et implémenter des améliorations de ces modèles dans le but d'augmenter la précision des résultats.

Pré-requis:

Le candidat devra avoir des compétences en programmation et montrer un réel intérêt pour les sciences cognitives et l'apprentissage automatique.