

De la représentation machine des sentiments et des émotions

Patrice Bellot

Professeur en Informatique, Université Aix-Marseille
Chargé de mission « fouille de textes » à l'INS2I-CNRS

L'intelligence artificielle appliquée à l'analyse textuelle, dans une perspective de traitement automatique des langues (traduction, résumé, dialogue humain-machine) ou de recherche d'information, modélise les tâches à accomplir comme l'enchaînement de raisonnements selon la logique standard ou des extensions de cette dernière ou encore comme un ensemble de problèmes de classification pour lesquels les approches numériques sont nombreuses et efficaces. Si les raisonnements, approximatifs ou non, s'appliquent sur des représentations et exploitent des opérations symboliques, les approches les plus récentes et désormais les plus efficaces pour la majorité des tâches du traitement automatique des langues et de la recherche d'information, dont la traduction automatique, exploitent des représentations numériques continues. Ces dernières permettent l'estimation de similarités surfaciques, sur la forme des mots plutôt que leur sens, la recherche d'analogies, le calcul de probabilités de correspondance à partir de corpus d'exemples. De nombreux problèmes du traitement automatique des langues se ramènent ainsi à des problèmes d'apprentissage statistique sur des données linguistiques ; en déterminer une représentation machine adéquate est une question majeure. Si les informations factuelles peuvent être représentées par des relations entre couples de mots, si un domaine thématique peut l'être par le champ lexical qui reflète les usages, la question de la représentation des sentiments et des émotions reste largement ouverte.

Dans ce séminaire, je présenterai tout d'abord les principales approches récentes de l'intelligence artificielle appliquée au texte par des illustrations de modèles vectoriels et probabilistes et des approches d'apprentissage statistique mises en œuvre par exemple par des méthodes de réseaux neuronaux tels que l'apprentissage profond. Je me concentrerai ensuite sur la question de la représentation et de l'identification des sentiments et des émotions en étudiant leurs représentations selon des modèles sacs de mots et des ressources strictement lexicales. Je montrerai quelques exemples appliqués au français et à l'anglais tels qu'on les trouve dans les jeux de données des compétitions internationales SemEval (Semantic Evaluation) et DEFT (Défi Fouille de Textes). L'adéquation des représentations habituellement exploitées pour l'estimation de similarités sémantiques à l'analyse de sentiment, tout comme la possibilité d'exploiter des représentations apprises sur une langue et appliquées sur une autre, figurent parmi les questions ouvertes et pour lesquelles des regards pluridisciplinaires sont précieux.