

Détection automatique de spammeurs pour l'annotation de données

Encadrant : Caio Corro (ISIR, MLIA)

L'annotation des données est une étape cruciale pour créer des jeux de données pour l'apprentissage supervisé. Il est aujourd'hui courant de faire appel à des systèmes de crowdsourcing où les annotateurs sont payés à la tâche. Certains annotateurs automatisent ce processus pour gagner plus d'argent, mais cela conduit à des données annotées de mauvaise qualité.

L'objectif de ce projet est de ré-implémenter le modèle présenté dans "Learning Whom to Trust with MACE" (Hovy et al.) qui permet de détecter automatiquement les annotateurs n'annotent pas correctement les données.

1. Implémentation du modèle de "base" via EM. Cela demandera de bien comprendre le modèle probabiliste, la dérivation de l'ELBO et l'algorithme d'EM ;
2. Évaluation de l'implémentation sur les données ;
3. Extension en ajoutant des priors sur des paramètres. Ceci nécessite de faire des approximations dans l'algorithme d'EM (mean field theory) ;
4. Évaluation du modèle avec priors sur les données.

Références :

- Learning Whom to Trust with MACE (Hovy et al.) <https://aclanthology.org/N13-1132.pdf>
- The Variational Approximation for Bayesian Inference (Tzikas et al.) <https://www.cs.uoi.gr/~arly/papers/SPM08.pdf>