

Offre: Postdoctorat TAL, Deep Learning CDD 1 an, IMT Mines Alès

Dans le cadre du projet PIA3 (Programme d'investissements d'avenir) ESII (transfert technologique), entre la société ESII et l'équipe KID du laboratoire LGI2P d'IMT Mines Alès, nous souhaitons recruter un chercheur postdoctoral pour une durée de 1 an.

Contexte

ESII produit des bornes interactives permettant d'aiguillier des usagers sur des guichets adaptés à leur demande, par exemple dans les administrations (préfecture, mairie) ou autre services (par exemple hôpitaux). Les usagers arrivent au service en question et passent devant la borne sur laquelle ils sélectionnent la nature de leur demande. La borne assigne alors l'usager à la file d'attente du guichet le plus apte à répondre à sa requête et lui donne un ticket.

ESII souhaite maintenant développer une borne actionnée vocalement, c'est à dire qu'au lieu de choisir une procédure sur l'écran, l'usager peut maintenant énoncer naturellement le motif de sa venue.

La parole de l'usager est transcrite par une brique logicielle existant et notre travail vise à proposer une solution générique qui à partir de cette transcription (imparfaite) et de la base de données domaine doit être en mesure d'automatiquement suggérer le bon guichet (en temps contraint).

Missions attendues

Le travail de la personne recrutée consistera à travailler sur l'implémentation d'un système de machine learning à base de modèles de langue profonds contextualisés (deep contextualized language models, par exemple OpenAI ELMo ou Google BERT) pour permettre cette recommandation :

1. La première étape consistera à construire un système non-supervisé sur la base des plongements de mots extraits des couches supérieures d'un modèle de langue profond contextualisé et de mesures de similarité sémantiques hybrides (combinant plongements et mesures hiérarchiques).
2. Dans un deuxième temps, des données d'entraînement seront disponibles. Il faudra partir du même modèle de langue contextualisé profond et proposer une architecture permettant de répondre à cette tâche. Cela qui permettra de spécialiser le modèle de langue pré-calculé en le raffinant au travers de l'entraînement par les données annotées.
3. Les données d'entraînement initiales (construites synthétiquement) seront complétées au fur et à mesure de l'utilisation de la solution par un feedback immédiat sur le résultat de l'affectation à un guichet (correct on non). Il faudra intégrer cette composante d'entraînement actif dans le système pour permettre de déterminer le moment opportun pour faire la transition entre le modèle non-supervisé et le modèle supervisé et pour raffiner l'entraînement du modèle supervisé.

Profil recherché

Formation Doctorat en Informatique (TAL) avec une première expérience de mise en opération de systèmes de machine learning et/ou une expérience en traitement automatique des langues par des approches en deep learning.

Compétences recherche

- Apprentissage Automatique et Profond
- Traitement Automatique des Langues (word embeddings, modèles de langue profonds)
- Des compétences en Web Sémantique seraient un plus (RDF, SPARQL, etc.)

Compétences techniques

- Python 3
- Scikit-learn
- PyTorch ou TensorFlow

Compétences transversales

- Maîtrise native du Français ou assimilé (Niveau C1+)
- Anglais technique

Savoir être

- Implication et sérieux
- Être force de proposition
- Savoir travailler en équipe
- Savoir communiquer sur l'avancement du travail à la fois avec des chercheurs spécialistes et avec des décideurs.

Rémunération et bénéfices

- Statut : Contractuel de la fonction publique
- Durée du contrat : 12 mois
- Lieu du contrat : IMT Mines Alès, Campus de Crouillac, Alès
- Rémunération brute annuelle 44 357€, soit 2052€ nets mensuels.

Pour postuler, envoyez un CV et une lettre de motivation à Andon Tchechmedjiev (<mailto:andon.tchechmedjiev@mines-ales.fr>) et à Jacky Montmain (<mailto:jacky.montmain@mines-ales.fr>)