

Contact :

Jean-Luc Lamotte et Bertrand Cuissart

E-mail : prenom.nom@unicaen.fr

Université de Caen Normandie

GREYC CNRS UMR 6072

6 Boulevard Maréchal Juin

F-14050 CAEN Cedex

Fouille de données sur des transformations électrochimiques : associer des conditions au rendement des réactions

Contexte Comme elle utilise un simple électron pour former ou rompre des liaisons au cours de processus complexes, l'électrosynthèse possède un très grand potentiel en termes de découverte de nouveaux processus et d'industrialisation à bas coût. Cependant, la maîtrise et l'optimisation de réactions électrocatalysées restent difficiles et l'apport des sciences du numérique à la résolution de problèmes chimiques représente une opportunité unique. Le projet AMPERE porté par les laboratoires LIMA (UMR 7042, INC), LBM (UMR 7203, INC), LHFA (UMR 5069, INC) et GREYC (UMR 6072, INS2I) s'inscrit dans cette dynamique et rassemble une communauté de chimistes et d'informaticiens souhaitant développer des processus d'aide à la décision facilitant la découverte et l'optimisation de transformations électrochimiques. AMPERE est financé dans le cadre du programme CNRS 80/PRIME, programme initié à l'occasion des 80 ans du CNRS, destiné à financer des projets interdisciplinaires originaux et en rupture (<https://www.cnrs.fr/fr/cnrsinfo/80-nouveaux-projets-pour-le-programme-80prime>). Au-delà de ce projet, ce dialogue interdisciplinaire entre intelligence artificielle et chimie enrichit chaque communauté en termes d'avancées fondamentales et appliquées.

Sujet Le travail de thèse proposé ici s'intègre entièrement au projet AMPERE, il s'agit donc d'un travail de recherche directement motivé par une application chimique. En s'appuyant sur un outil de criblage qui génère expérimentalement des données sur des réactions électrochimiques, il s'agit de concevoir et d'implémenter des méthodes informatiques de fouille de données. A partir des techniques actuelles utilisées dans les bases de données de réactions chimiques, nous commencerons par construire une base de données appropriée aux réactions électrochimiques. Cet outil jouant un rôle pivot dans le projet, il devra satisfaire les besoins de consultation des chimistes, et constituer une source pour les analyses informatiques. En conséquence, sa réalisation sera le fruit d'une discussion entre les experts des deux disciplines. En s'appuyant sur la base de données de réactions électrochimiques, la thèse comportera deux travaux de recherche informatique.

Premièrement, il s'agira de concevoir une méthode d'analyse adaptée aux données réactionnelles étudiées. En effet, le caractère dynamique et complexe des réactions nécessite de les représenter sous la forme de données structurées à partir desquelles seront extraits les descripteurs pertinents. L'analyse informatique associée à cette partie du travail calculera les associations statistiques remarquables entre conditions expérimentales et valeurs de rendements réactionnels. Pour une réaction donnée, le système pourra proposer des nouveaux paramètres afin d'augmenter le rendement, ainsi qu'une explication de ses choix, explication compréhensible par un chimiste.

Ensuite, un travail de recherche en fouille de données séquentielles est envisagé. Une réaction sera modélisée sous la forme d'une suite d'états, chaque état décrivant un point de la réaction. Les techniques de fouille de séquences adaptées à ce cas permettront d'ex-

traire les enchaînements de parties d'états réactionnels "remarquables". Par remarquable, on peut entendre "fréquent", "unique" ou très associé à un caractère externe comme un niveau de rendement. Pour modéliser d'éventuelles incertitudes, il peut être possible de remplacer les séquences par des arbres orientés. Pour réaliser l'analyse subséquente, il faudra concevoir un processus de fouille original; ce travail comporte un fort caractère innovant.

Apports de la thèse Il s'agit d'un travail informatique en Sciences des Données, une discipline très actuelle de l'intelligence artificielle. Le doctorant aura l'opportunité de construire une expertise dans le cadre spécifique de la recherche des associations au sein des données structurées, les données étant ici représentées sous la forme de graphes ou de séquences.

Le travail s'intégrant à un projet interdisciplinaire, le doctorant expérimentera concrètement le dialogue orientant les travaux de ce type de projets. De plus, il acquerra une compétence complémentaire dans le domaine de l'informatique appliquée à la chimie. L'innovation en analyse de données chimiques s'associant à des défis technologiques ou scientifiques, cette compétence offre l'opportunité de participer à des projets ambitieux, de nature variée et aux retombées importantes.

Profil du candidat Le candidat doit être inscrit en dernière année d'un Master ou d'un diplôme d'ingénieur dans un domaine lié à l'informatique ou aux mathématiques appliquées, et posséder de solides compétences en programmation; le candidat peut aussi être titulaire d'un tel diplôme. Une expérience en informatique pour la Science des Données sera un plus (fouille de données, apprentissage automatique, ...). Le candidat doit avoir des capacités à rédiger des rapports scientifiques et à communiquer des résultats de recherche lors de conférences en anglais.

Conditions de la thèse La thèse débutera à l'automne 2022 – début septembre ou début octobre. Le travail se déroulera principalement au GREYC, laboratoire académique normand situé à Caen. Le projet impliquant plusieurs laboratoires français, la thèse inclut plusieurs semaines annuelles de travail dans les laboratoires partenaires.

– Salaire brut mensuel : 2 135 €

Candidature Les candidatures doivent inclure les documents suivants au format électronique : une lettre de motivation, un CV détaillé décrivant vos études et votre expérience en recherche, les relevés de notes des diplômes obtenus, les coordonnées de personnes références préférentiellement issues du monde la recherche.

Veillez envoyer votre dossier de candidature à bertrand.cuissart@unicaen.fr et jean-luc.lamotte@unicaen.fr.

Date limite d'envoi des dossiers de candidature : 30 juin 2022