AVIS DE SOUTENANCE DE THESE

Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que

M^me^ **: EL-ALLALY ED-DRISSIYA**

Soutiendra **: le 22/07/2022** à **10H**
Lieu **: Centre Polyvalent des Etudes Doctorales (Amphi 1)**

Une thèse intitulée :

Contributions to improving Adverse Drug Events Extraction from Biomedical Documents

En vue d'obtenir le Doctorat
FD : Sciences et Technologies de l'Information et de Communication (STIC)
Spécialité : Informatique

Devant le jury composé comme suit :

Président	Pr. LAMRINI Mohamed	PES	Faculté des Sciences Dhar El Mahraz – Fès
Directeur de thèse	Pr. EN NAHNAHI Noureddine	PH	Faculté des Sciences Dhar El Mahraz - Fès
Co-directeur de thèse	Pr. OUATIK EL ALAOUI Said	PES	Ecole Nationale des Sciences Appliquées-Kenitra
Rapporteurs	Pr. CHOUGDALI Khalid	PH	Ecole Nationale des Sciences Appliquées- Kenitra
	Pr. AOURAGH Si Lhoussain	PH	Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes - ENSIAS-Rabat
	Pr. NFAOUI El Habib	PES	Faculté des Sciences Dhar El Mahraz - Fès
Membres	Pr. BEN ABBOU Rachid	PES	Faculté des Sciences et Techniques - Fès
	Pr. NAJAH Said	PH	Faculté des Sciences et Techniques – Fès
	Pr. ALAOUI ZIDANI Khalid	PH	Faculté des Sciences Dhar El Mahraz - Fès

Résumé :

Les Evénements Indésirables des Médicaments (EIM) regroupent tout symptôme nocif survenant chez une personne durant un traitement. Ils représentent un véritable problème de santé publique susceptible de provoquer des conséquences négatives, voire mortelles. Ainsi, leur détection précoce est primordiale pour la recherche pharmaceutique. Bien que les EIM courants puissent être identifiés par les essais cliniques, la découverte des événements rares nécessite des systèmes de notification spontanée au cours de la surveillance après commercialisation. Cependant, la sous déclaration empêche la détection exhaustive de tous les EIM. Les techniques de traitement automatique de la langue, destinées à extraire les connaissances sur EIM à partir de documents biomédicaux, constituent un complément remarquable pour accélérer la pharmacovigilance. Le processus d'extraction automatique de ces événements à partir de documents biomédicaux comprend principalement trois tâches : (1) l'extraction de mentions, (2) l'extraction de relations et (3) l'extraction de bout en bout. L'objectif de cette thèse est de concevoir des systèmes robustes pour chacune de ces tâches. Actuellement, les méthodes basées sur l'apprentissage profond représentent les solutions de l'état de l'art pour chaque tâche. Cependant, elles ne parviennent pas à extraire efficacement les mentions complexes (imbriquées, discontinues et chevauchantes) ainsi que leurs relations potentielles. Nous proposons donc quatre contributions pour pallier ces lacunes. Dans la première contribution, nous proposons une méthode basée sur la machine d'apprentissage extrême récurrente en ligne et pondérée pour améliorer la performance de la tâche d'extraction de mentions simples. La méthode est divisée en deux étapes. La première étape vise à trouver les régions des mentions via un étiquetage séquentiel, tandis que la deuxième étape sert à classifier les fragments de texte détectés selon leurs types appropriés. Dans la deuxième contribution, nous proposons un modèle neuronal profond pour extraire les mentions complexes, appelé DeepCADRME. Il transforme la tâche en un étiquetage de séquences à N niveaux pour couvrir différents types de mentions. Ensuite, il exploite une représentation profonde contextualisée pour traiter les séquences obtenues dans un processus en pipeline. Dans la troisième contribution, nous développons ADERel, un modèle conjoint attentif pour traiter la tâche d'extraction de relations.

ADERel formule la tâche comme un étiquetage de séquence par la modélisation des relations à différents niveaux en vue de les apprendre conjointement. Il intègre un réseau convolutif de graphe pondéré basé sur le transformateur pour tirer profit des caractéristiques contextuelles et syntaxiques. Le mécanisme d'attention à multiples têtes est appliqué pour échanger des connaissances sur les limites entre les niveaux. Dans la dernière contribution, nous proposons un système d'extraction de bout en bout, appelé MTTLADE. Ce dernier convertit le problème en un étiquetage de séquence à double tâche : la tâche d'extraction des mentions sources et la tâche d'extraction des attributs et des relations. Le système combine l'apprentissage par transfert et multi-tâche pour apprendre les deux tâches simultanément. Diverses expérimentations ont été menées pour prouver l'efficacité de l'ensemble des solutions proposées.

Mots clés :

Evènement Indésirable des Médicaments, Pharmacovigilance, Traitement Automatique de la Langue, Reconnaissance d'Entités Nommées, Extraction de Relations, Apprentissage Profond, Apprentissage par Transfert, Apprentissage Multi-Tâches, Apprentissage Conjoint.

CONTRIBUTIONS TO IMPROVING ADVERSE DRUG EVENTS EXTRACTION FROM BIOMEDICAL DOCUMENTS

Abstract

Adverse Drug Events (ADE) are injuries that occur during a patient's drug therapy. They constitute a serious public health issue that can result in poor outcomes or even death. Therefore, their early detection is critical for pharmaceutical discovery as it improves patient safety and reduces their seriousness. While common ADE may be recognized by clinical trials during pre-approval surveillance, uncovering rare ones requires spontaneous reporting systems during post-marketing surveillance. However, under-reporting is one of the main factors that inhibit exhaustive detection of all ADE. Natural language processing techniques towards extracting ADE knowledge from
biomedical documents provide a remarkable complement to accelerate pharmacovigilance and drug-safety monitoring. The process of automated extraction of ADE from unstructured biomedical documents mainly includes three tasks: (1) mention extraction, (2) relation extraction and (3) end-to-end ADE extraction. The purpose of this thesis is to design robust and accurate systems so as to address the challenging issues related to each task. To date, deep learning-based methods have been the state-of-the-art solutions in each task due to their ability to automatically learn relevant features to be used without requiring an expert domain knowledge. However, they still face several challenges that should be addressed. Among these, the existing methods fail
to effectively extract complex mentions (nested, discontinuous and overlapping) as well as the potential relationships between them. Thus, we propose four contributions to solve these issues by following the novel line of work on deep neural networks. In the first contribution, we propose a weighted online recurrent extreme learning machine to improve the performance of simple mention extraction task. The method is divided into two stages: span detection and mentions classification. The first stage aims at finding the boundary of mentions via sequence labelling, whereas the second stage aims at classifying the detected text fragments into an according mention types. In the second contribution, we propose a deep neural model for extracting complex mentions, called DeepCADRME. Specifically, it transforms the task as N-level sequence labelling to cover various kinds of mentions. Then, it exploits a deep contextualized representation to learn the obtained sequences in a pipelined process. In the third contribution, we develop ADERel, an attentive joint model for dealing with relation extraction task. ADERel formulates the task as a sequence labelling problem by modeling relations in different levels to learn them jointly. It integrates a transformed-based weighted graph convolutional network to leverage richer contextual and syntactical features. A multi-head attention is applied for exchanging boundary knowledge across levels. In the last contribution, we propose an end-to-end ADE extraction system, called MTTLADE. This latter converts the problem to a dual-task sequence labelling which includes source mention extraction task and attribute-relation extraction task. It combines multi-task and transfer learning to process the two proposed tasks simultaneously. Several experiments were carried out on well-known datasets from various biomedical documents in order to demonstrate the effectiveness of our contributions.

Keywords: Adverse Drug Event, Pharmacovigilance, Natural Language Processing, Biomedical Named Entity Recognition, Biomedical Relation Extraction, Deep Learning, Transfer Learning, Multi-Task Learning, Joint Learning.