



HAL
open science

De l'extraction des interactions médicament-médicament vers les interactions aliment-médicament à partir de textes biomédicaux: Adaptation de domaine

Tsanta Randriatsitohaina, Thierry Hamon

► To cite this version:

Tsanta Randriatsitohaina, Thierry Hamon. De l'extraction des interactions médicament-médicament vers les interactions aliment-médicament à partir de textes biomédicaux: Adaptation de domaine. 26e Conférence sur le Traitement Automatique des Langues Naturelles, Jul 2019, Toulouse, France. pp.223-232. hal-02567771

HAL Id: hal-02567771

<https://hal.science/hal-02567771>

Submitted on 30 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

De l'extraction des interactions médicament-médicament vers les interactions aliment-médicament à partir de textes biomédicaux: Adaptation de domaine

Tsanta Randriatsitohaina¹ Thierry Hamon^{1,2}

(1) LIMSI, CNRS, Université Paris-Saclay, Campus universitaire d'Orsay, 91405 Orsay cedex, France

(2) Université Paris 13, 99 avenue Jean-Baptiste Clément, 93430 Villetaneuse, France

tsanta@limsi.fr, hamon@limsi.fr

RÉSUMÉ

Les interactions aliments-médicaments (FDI) se produisent lorsque des aliments et des médicaments sont pris simultanément et provoquent un effet inattendu. Nous considérons l'extraction de ces interactions dans les textes comme une tâche d'extraction de relation pouvant être résolue par des méthodes de classification. Toutefois, étant donné que ces interactions sont décrites de manière très fine, nous sommes confrontés au manque de données et au manque d'exemples par type de relation. Pour résoudre ce problème, nous proposons d'appliquer une adaptation de domaine à partir des interactions médicament-médicament (DDI) qui est une tâche similaire, afin d'établir une correspondance entre les types de relations et d'étiqueter les instances FDI selon les types DDI. Notre approche confirme une cohérence entre les 2 domaines et fournit une base pour la spécification des relations et la pré-annotation de nouvelles données. Les performances des modèles de classification appuie également l'efficacité de l'adaptation de domaine sur notre tâche.

ABSTRACT

From the extraction of drug-drug interactions to the food-drug interactions in biomedical texts : domain adaptation.

Food-drug interactions (FDI) occur when food and drug are taken simultaneously and cause unexpected effect. We consider the extraction of these interactions in texts as a relation extraction task that can be resolved with classification methods. Nevertheless, since these interactions are described in a very fine way, we are confronted with a lack of data and a lack of examples per type of relation. To solve this problem, we propose to apply a domain adaptation from drug-drug interactions (DDI) which is a similar task, in order to match relation types and label FDI instances according to DDI types. Our approach confirms consistency between the 2 domains and provides a basis for specifying relations and pre-annotating new data. The performance of classification models also supports the effectiveness of domain adaptation on our task.

MOTS-CLÉS : Adaptation de domaine, Classification, Extraction de relation, Textes biomédicaux.

KEYWORDS: Domain adaptation, Classification, Relation Extraction, Biomedical texts.

1 Introduction

Bien qu'il existe des bases ou des terminologies recensant les connaissances d'un domaine de spécialité, disposer d'informations à jour nécessite souvent le recours à la consultation d'articles scientifiques. Ce constat est d'autant plus vrai lorsque les connaissances à recenser ne sont pas déjà présentes dans une base. Ainsi, si les interactions entre médicaments (Aagaard & Hansen, 2013) ou les effets indésirables d'un médicament (Aronson & Ferner, 2005) sont répertoriés dans des bases telles que DrugBank¹ ou Theriaque², d'autres informations comme les interactions entre un médicament et un aliment y sont très peu présentes et souvent fragmentées et dispersées dans des sources hétérogènes, principalement sous forme textuelle. Afin de répondre à ces problématiques de mises à jour ou de recensement de ces informations, des méthodes de fouille de textes sont généralement mises en œuvre (Cohen & Hunter, 2008; Rzhetsky *et al.*, 2009; Chowdhury *et al.*, 2011).

Dans cet article, nous nous intéressons à l'identification automatique de mentions d'interaction entre un médicament et un aliment (Food-Drug Interaction - FDI) dans des résumés d'articles scientifiques issus de la base Medline. A l'instar des interactions entre médicaments, une interaction entre un médicament et un aliment correspond à l'apparition d'un effet non attendu quand leur prise est combinée. Par exemple, le pamplemousse est connu pour avoir un effet inhibiteur sur une enzyme impliquée dans le métabolisme de plusieurs médicaments (Hanley *et al.*, 2011). D'autres aliments peuvent avoir des effets sur l'absorption d'un médicament ou sur sa distribution dans l'organisme (Doogue & Polasek, 2013). Pour extraire ces informations des résumés, nous faisons face à plusieurs difficultés : (1) les médicaments et les aliments sont mentionnés de manière très variable dans les résumés. Il peut s'agir des dénominations communes internationales ou des substances actives de médicaments tandis que pour les aliments, il peut aussi être fait mention d'un nutriment, d'un composant particulier ou d'une famille d'aliment ; (2) les interactions sont décrites de manière assez fine dans le corpus annoté à notre disposition ce qui conduit à un nombre d'exemples peu conséquent ; (3) bien que nous disposons de résumés annotés avec des interactions aliment/médicament, l'ensemble des annotations ne couvrent pas de manière homogène les différents types d'interaction et l'ensemble d'apprentissage est bien souvent déséquilibré.

Nous considérons l'extraction de ces interactions comme une tâche d'acquisition de relation étant donné les mentions d'un aliment et d'un médicament. Afin de répondre à ces difficultés, nous proposons d'utiliser une méthode à base d'apprentissage par transfert en tirant partie des données existantes sur les interactions médicaments-médicaments.

Après avoir présenté un état de l'art des méthodes d'acquisition de relations en corpus de spécialité (section 2), nous décrivons à la section 3, le corpus annoté que nous avons utilisé pour mettre au point notre approche afin d'extraire des interactions entre médicament et aliment (section 4). Puis nous présentons et discutons les résultats obtenus (section 6) et nous concluons (section 7).

2 Etat de l'art

Différents types d'approches ont été explorés pour extraire les relations à partir des textes biomédicaux. Certaines approches combinent les patrons avec du CRF pour la reconnaissance des symptômes

1. <https://www.drugbank.ca/>

2. <http://www.theriaque.org>

dans les textes biomédicaux (Holat *et al.*, 2016). D'autres approches génèrent automatiquement des données lexicales pour le traitement du texte libre dans les documents cliniques en s'appuyant sur un modèle alignement séquentiel multiple pour identifier des contextes similaires (Meng & Morioka, 2015).

Afin d'extraire les interactions médicament-médicament (DDI) (Kolchinsky *et al.*, 2015) se concentrent sur l'identification des phrases pertinentes et des résumés pour l'extraction des preuves pharmacocinétiques. (Ben Abacha *et al.*, 2015) proposent une approche basée sur du SVM combinant : (i) les caractéristiques décrivant les mots dans le contexte des relations à extraire, (ii) les noyaux composites utilisant des arbres de dépendance. L'union et l'intersection des résultats permettent d'obtenir des F1-mesures de 0,5 et 0,39 respectivement. Une méthode en deux étapes basée également sur le classifieur SVM est proposée par (Ben Abacha *et al.*, 2015) pour détecter les DDI potentiels, puis classer les relations parmi les DDI déjà identifiées. Cette seconde approche permet d'obtenir des F1-mesures de 0,53 et 0,40 sur des résumés Medline et 0,83 et 0,68 sur des documents de DrugBank. (Kim *et al.*, 2015) ont construit deux classifieurs pour l'extraction DDI : un classifieur binaire pour extraire les paires de médicaments en interaction et un classifieur de types DDI pour identifier les catégories de l'interaction. (Cejuela *et al.*, 2018) considèrent l'extraction de la relation de localisation des protéines comme une classification binaire. (Liu *et al.*, 2016) proposent une méthode basée sur CNN pour l'extraction des DDI. Dans leur modèle, les mentions de médicaments dans une phrase sont normalisées de la manière suivante : les deux médicaments considérés sont remplacés par `drug1` et `drug2` respectivement dans l'ordre de leur apparition, et tous les autres médicaments sont remplacés par `drug0`. D'autres travaux utilisent un modèle de réseau neuronal récurrent avec plusieurs couches d'attention pour la classification DDI (Yi *et al.*, 2017; Zheng *et al.*, 2017), ou utilisant des récurrences au niveau des mots et des caractères (Kavuluru *et al.*, 2017) produisant une performance de 0,72. Sun *et al.* (2019) propose une méthode hybride combinant un réseau de neurone récurrent et convolutionnel induisant une amélioration de 3%. Le réseau convolutionnel profond de Dewi *et al.* (2017) permet de couvrir de longues phrases qui ont des jeux de données de DDI typique et obtenir une performance de 0,86.

Une méthode simple d'adaptation de domaine consiste à construire un jeu de données étiqueté pour le domaine cible, puis à ajuster un modèle entraîné avec ce dernier (Jiang & Zhai, 2007; Blitzer *et al.*, 2006; Daumé III *et al.*, 2010). Comme notre problème d'extraction d'interaction médicament-aliment est similaire à l'extraction des DDI et les données DDIs sont mieux fournies, nous proposons une approche basée sur l'adaptation de domaine en projetant des modèles entraînés à partir du DDI sur des données FDI afin d'obtenir de nouvelles étiquettes.

3 Corpus

Des études ont déjà été menées sur les FDI durant lesquelles l'ensemble de données POMÉLO a été développé (Hamon *et al.*, 2017). Ce corpus est constitué de 639 résumés d'articles scientifiques du domaine médical (269 824 mots, 5 752 phrases), collectés à partir du portail PubMed³ avec la requête : ("FOOD DRUG INTERACTIONS" [MH] OR "FOOD DRUG INTERACTIONS*") AND ("adverse effects*"). Les 639 résumés sont annotés selon 9 types d'entités et 21 types de relations avec Brat (Stenetorp *et al.*, 2012) par un étudiant en pharmacie. Les annotations se concentrent sur des informations sur la relation entre aliments, médicaments et pathologies.

3. <https://www.ncbi.nlm.nih.gov/pubmed/>

Étant donné que nous examinons les interactions aliment-médicament dans cet article, nous avons construit notre ensemble de données en tenant compte de tous les couples de *drug* et *food* ou *food-supplement* à partir des données POMELO. L'ensemble de données qui en résulte est composé de 902 phrases étiquetées avec 13 types de relations : *Relation (Rel)*, *Decrease absorption (Dec)*, *No effect on drug (No)*, *Increase absorption (Inc)*, *Negative effect on drug (Neg)*, *Positive effect on drug (Pos)*, *New side effect (New)*, *Without food (Wout)*, *Improve drug effect (Imp)*, *Slow elimination (Sl-e)*, *Slow absorption (Sl-a)*, *Worsen drug effect (Wors)*, *Speed up absorption (Spd)*. Les statistiques de l'ensemble de données sont données dans la table 1.

Relation	#	Pourcentage	Relation	#	Pourcentage
unspecified relation	530	58,8%	no effect on drug	109	12,1%
decrease absorption	53	5,9%	improve drug effect	6	0,7%
positive effect on drug	21	2,3%	without food	13	1,4%
negative effect on drug	88	9,8%	speed up absorption	1	0,1%
increase absorption	39	4,3%	worsen drug effect	8	0,9%
slow elimination	15	1,7%	new side effect	4	0,4%
slow absorption	15	1,7%			
Total			902		100%

TABLE 1: Statistique des données - Interactions aliment-médicament

4 Méthode

Comme les interactions entre aliments-médicaments sont décrites de manière très fine selon de nombreux types de relations, nous sommes confrontés au manque de données et au manque d'exemples par type de relation comme indiqué dans le tableau 1. Par exemple, la relation *speed up absorption* est représenté par un seul exemple, ce qui ne permet pas d'établir une méthode efficace pour extraire automatiquement la relation considérée. Par ailleurs, la tâche d'identification des FDI présente des similitudes avec la tâche d'extraction des DDI, où deux médicaments pris ensemble mènent à une modification de leurs effets. Dans cet article, nous étudions la projection des connaissances sur les DDI afin d'obtenir de nouveaux types correspondant aux FDI.

4.1 Correspondance de types DDI-FDI

Afin de vérifier la cohérence entre deux domaines, nous proposons une approche permettant de trouver une correspondance de type DDI pour chaque type FDI. Pour ce faire, un modèle est entraîné sur les données DDI. Ensuite chaque type de relation est représenté par un ensemble de descripteurs, formant ainsi une instance par type de relation. Puis nous projetons un modèle entraîné avec les données DDI sur cette représentation afin de déterminer à quel type de DDI correspond la relation. L'ensemble de données de relation est donc un vecteur $D_R = [D_1, D_2, \dots, D_n]$ de taille n , où n est le nombre de types de relations FDI, D_i est un ensemble de descripteurs représentant la i^{th} relation. Dans cet article, nous construisons les descripteurs D_i à partir des descripteurs de chaque phrase P_i étiquetée par la relation R_i dans l'ensemble de données initial D . Ainsi, le modèle affecte une étiquette de type DDI pour chaque type FDI.

4.2 Contraste des étiquettes d’instances

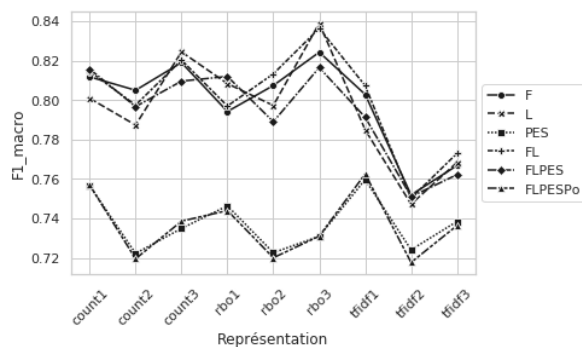
Dans cette section, nous analysons individuellement les étiquettes DDI affectées aux instances FDI afin de déterminer si tous les membres d’une classe FDI donnée sont affectés au même type DDI et appuyer la correspondance obtenu dans la section 4.1. Pour ce faire, nous retenons le meilleur modèle obtenu pour la classification des DDIs et l’appliquons sur les données FDI afin d’obtenir de nouvelles étiquettes pour chaque instance. Cette méthode nous permet de contraster les annotations des données POMELO et une annotation basée sur les types DDIs. Une méthode de classification des instances FDI est ensuite mise en place en utilisant les étiquettes DDI.

5 Expérimentation

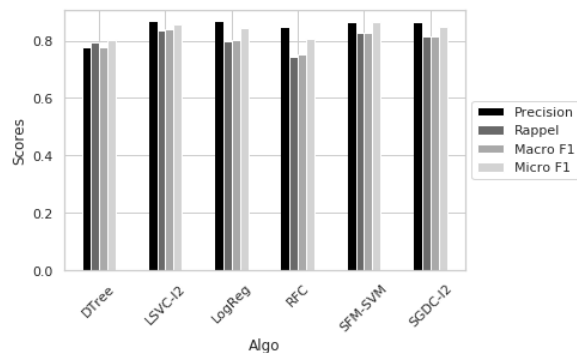
5.1 Modèle DDI

L’adaptation de domaine DDI-FDI est évaluée à travers l’utilisation de classifieurs et de descripteurs.

Corpus Le corpus DDI⁴ est composé de 2280 instances extraites de la base DrugBank et des résumés Medline, similaire au corpus utilisé lors de la compétition (Segura-Bedmar *et al.*, 2013), étiquetées suivant quatre types : *conseil* (540) pour une recommandation concernant l’utilisation concomitante de deux médicaments, *effet* (591) pour l’effet du DDI, *mécanisme* (1006) pour la pharmacodynamique (les effets d’un médicament sont modifiés par la présence d’un autre médicament) ou la pharmacocinétique (les processus par lesquels les médicaments sont absorbés, distribués, métabolisés et excrétés), *interaction* (143) où aucune information sur l’interaction n’est fournie.



(a) Macro F1 obtenu avec SVM selon la représentation des descripteurs.



(b) Performances obtenues selon les modèles, sur l’unigramme, bigramme et trigramme des lemmes.

FIGURE 1: Performances obtenues sur les données DDI.

Descripteurs. Dans chaque phrase, les chiffres sont remplacés par le caractère ‘#’ comme proposé dans (Kolchinsky *et al.*, 2015), les autres caractères spéciaux sont supprimés, et chaque mot est converti en minuscule. Nous analysons l’impact de différents descripteurs : formes fléchies (F) des mots, lemmes (L), étiquettes morpho-syntaxiques (Po), fenêtres de mots précédant le premier argument de la relation (P), mots entre les deux arguments (E) et mots suivant le second argument (S).

4. <https://github.com/dbmi-pitt/public-PDDI-analysis/blob/master/PDDI-Datasets/DDI-Corpus2013>

Classification. La performance de 6 classifieurs est évaluée selon la précision (P), le rappel (R), et la F1-mesure (F_1) obtenus par un processus de validation croisée en 10 échantillons. Nous utilisons l’implémentation Scikit-learn (Pedregosa *et al.*, 2011) des classifieurs : (1) un arbre de décision (DTree), (2) un classifieur SVM linéaire (LSVC-l2), (3) une régression logistique (LogReg), (4) un classifieur bayésien naïf multinomial (MNB), (5) une forêt aléatoire (RFC) et (6) un SVM combiné avec un algorithme de sélection de descripteurs (SFM-SVM).

5.2 Adaptation de domaine

Afin d’étiqueter les données FDI selon les types DDI, la meilleure configuration obtenue à la section 5.1 est appliquée sur les données POMELO selon la méthode décrite dans la section 4. D’après les nouvelles étiquettes, on obtient une répartition différente des instances : interaction (104 instances), mécanisme (397 instances), conseil (72 instances), effet (329 instances) (table 2). Les modèles de classification sont alors évalués sur nos données en utilisant les nouvelles étiquettes afin de déterminer l’effet de l’adaptation de domaine sur l’identification automatique des relations. De la même manière que pour le modèle DDI décrit à la section 5.1, nous évaluons les 6 classifieurs par un processus de validation croisée en 10 échantillons en faisant varier les descripteurs utilisés.

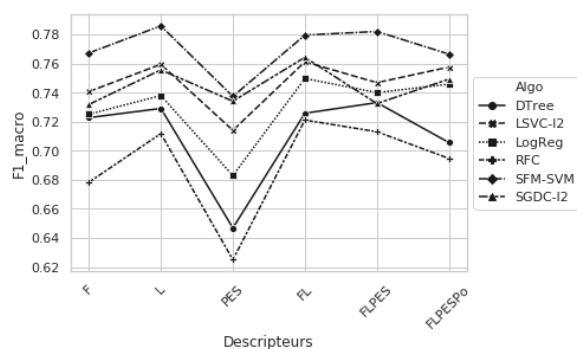
FDI	Rel	Dec	No	Inc	Neg	Pos	New	Wout	Imp	Sl-e	Wors	Sl-a	Spd
DDI	M	M	M	M	E	E	E	C	E	M	E	M	M
Cons	7	2	9	5	9	33	0	54	16	0	0	0	0
Effet	44	2	19	5	53	38	75	23	67	7	50	0	0
Int	17	2	1	0	16	0	0	0	0	0	0	0	0
Méca	32	94	71	90	22	29	25	23	17	93	50	100	100

TABLE 2: Correspondance type DDI-FDI (Ligne 1) et pourcentage d’instances FDI affectées au type DDI (Ligne 2-5) - *Conseil (C)*, *Mécanisme (M)*, *Effet (E)*, *Interaction (Int)*.

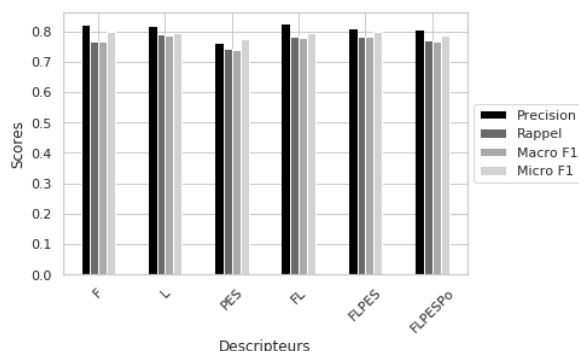
6 Résultat

Pour l’extraction des DDIs, nous avons testé l’impact des unigramme, bigramme et trigramme des descripteurs (figure 1a) en les représentant par leur occurrence binaire (rbo), leur fréquence (count) ou leur TF-IDF (tfidf). La meilleure F1-mesure 0,839 est obtenue avec un SVM linéaire qui présente un bon rappel (figure 1b) en utilisant les trigrammes des lemmes comme descripteurs. Ce résultat est meilleur que celui obtenu par le meilleur modèle sur la même tâche lors de la compétition Semeval 2013 (Segura-Bedmar *et al.*, 2013). Nous avons appliqué ce modèle sur les représentations des relations FDI comme décrit à la section 4.1 afin d’obtenir une correspondance entre les types DDI et FDI. Les résultats (tableau 2, ligne 1) indiquent une cohérence entre les 2 domaines. En effet, les interactions impliquant la cinétique des médicaments (absorption, élimination, métabolisme) sont étiquetées *Mécanisme*, les interactions impliquant les effets du médicament (positif, négatif, secondaire) sont étiquetées *Effet*, la relation *Without food* qui est une contre-indication de prise d’aliment avec le médicament est étiquetée *Conseil*. Cette cohérence des étiquettes démontrent l’efficacité de l’approche par la représentation des relations. Toutefois, nous remarquons que la *relation non-spécifiée* n’a pas été étiquetée *Int*. L’analyse des instances de cette relation (tableau 2,

lignes 2-5) montre qu'un peu moins de la moitié est considérée comme des effets et le tiers comme des mécanismes. Ces informations pourront servir d'appui pour ajouter des précisions sur les interactions non spécifiées. L'homogénéité élevée pour les relations impliquant des mécanismes suggère une bonne qualité des annotations FDI pour ces types de relation. Les relations impliquant des effets sont beaucoup moins homogènes, mais les nouvelles étiquettes peuvent servir de base pour une ré-évaluation des annotations des FDI. Avec les nouvelles étiquettes obtenues, le résultat obtenu sur la tâche d'extraction des FDI confirme effectivement l'efficacité de la méthode car la F1-mesure résultante est passée de 0,41 sur les étiquettes initiales à 0,79 sur les nouvelles étiquettes (figure 2a). Ce résultat est obtenu par un modèle SVM avec une régularisation l2 précédé d'un processus de sélection de descripteurs SFM, en utilisant les unigramme, bigramme et trigramme des lemmes comme descripteurs. La précision et le rappel sont visiblement équilibrés, ainsi que la macro et micro F1, (figure 2b) ce qui suggère un bon équilibre des données.



(a) Macro F1 selon les différents modèles.



(b) Performances obtenues avec SFM-SVM.

FIGURE 2: Performances obtenues sur les données POMELO.

7 Conclusion et Perspectives

Notre article contribue à la tâche d'extraction des interactions aliments-médicaments (FDI) à partir de la littérature scientifique, que nous traitons comme une tâche d'extraction de relation. L'apprentissage supervisé dans ce but n'est pas très concluant car nous sommes confrontés à un problème de manque d'exemples en raison du nombre élevé de types de relations étant donné que les interactions sont décrites de manière très fine. Pour pallier ce problème, nous proposons d'appliquer une adaptation de domaine à partir des interactions médicaments-médicaments (DDI) qui est une tâche similaire, afin d'établir une correspondance entre les types de relations et étiqueter les instances FDI selon les types DDI. Les résultats obtenus suggèrent que l'approche basée sur la représentation des relations à partir des instances est efficace pour identifier les correspondances des types et évaluer la cohérence entre les 2 domaines. Les nouvelles étiquettes obtenues peuvent servir d'appui pour préciser les interactions non spécifiées et pré-annoter de nouvelles données. Pour la suite du travail, nous envisageons d'utiliser d'autre corpus et méthode DDI et confronter les résultats pour une meilleure annotation des FDI.

Remerciements

Ce travail est financé par l'ANR dans le cadre du projet MIAM (ANR-16-CE23-0012).

Références

- AAGAARD L. & HANSEN E. (2013). Adverse drug reactions reported by consumers for nervous system medications in europe 2007 to 2011. *BMC Pharmacology & Toxicology*, **14**, 30.
- ARONSON J. & FERNER R. (2005). Clarification of terminology in drug safety. *Drug Safety*, **28**(10), 851–70.
- BEN ABACHA A., CHOWDHURY M. F. M., KARANASIOU A., MRABET Y., LAVELLI A. & ZWEIGENBAUM P. (2015). Text mining for pharmacovigilance : Using machine learning for drug name recognition and drug-drug interaction extraction and classification. *Journal of Biomedical Informatics*, **58**, 122–132.
- BLITZER J., MCDONALD R. & PEREIRA F. (2006). Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, p. 120–128 : Association for Computational Linguistics.
- CEJUELA J. M., VINCHURKAR S., GOLDBERG T., PRABHU SHANKAR M. S., BAGHUDANA A., BOJCHEVSKI A., UHLIG C., OFNER A., RAHARJA-LIU P., JENSEN L. J. & ROST B. (2018). LocText : relation extraction of protein localizations to assist database curation. *BMC Bioinformatics*, **19**(1), 15.
- CHOWDHURY F. M., LAVELLI A. & MOSCHITTI A. (2011). A study on dependency tree kernels for automatic extraction of protein-protein interaction. In *Proceedings of BioNLP 2011 Workshop*, p. 124–133 : Association for Computational Linguistics.
- COHEN K. & HUNTER L. (2008). Getting started in text mining. *PLoS Computational Biology*, **4**(1), e20.
- DAUMÉ III H., KUMAR A. & SAHA A. (2010). Frustratingly easy semi-supervised domain adaptation. In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, p. 53–59 : Association for Computational Linguistics.
- DEWI I. N., DONG S. & HU J. (2017). Drug-drug interaction relation extraction with deep convolutional neural networks. *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, p. 1795–1802.
- DOOGUE M. & POLASEK T. (2013). The abcd of clinical pharmacokinetics. *Ther Adv Drug Saf*, **4**(1), 5–7.
- HAMON T., TABANOU V., MOUGIN F., GRABAR N. & THIESSARD F. (2017). Pomelo : Medline corpus with manually annotated food-drug interactions. In *Proceedings of Biomedical NLP Workshop associated with RANLP 2017*, p. 73–80, Varna, Bulgaria.
- HANLEY M., CANCALON P., WIDMER W. & GREENBLATT D. (2011). The effect of grapefruit juice on drug disposition. *Expert Opin Drug Metab Toxicol*, **7**(3), 267–286.
- HOLAT P., TOMEH N., CHARNOIS T., BATTISTELLI D., JAULENT M.-C. & MÉTIVIER J.-P. (2016). Weakly-supervised symptom recognition for rare diseases in biomedical text. In *Proceedings of the 15th International Symposium IDA 2016*, Lecture Notes in Computer Science, 9897, Advances in Intelligent Data Analysis XV, p. 192–203, Stockholm, Sweden.
- JIANG J. & ZHAI C. (2007). Instance weighting for domain adaptation in nlp. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, p. 264–271 : Association for Computational Linguistics.
- KAVULURU R., RIOS A. & TRAN T. (2017). Extracting drug-drug interactions with word and character-level recurrent neural networks. In *Healthcare Informatics (ICHI), 2017 IEEE International Conference on*, p. 5–12 : IEEE.

KIM S., LIU H., YEGANOVA L. & WILBUR W. J. (2015). Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach. *Journal of biomedical informatics*, **55**, 23–30.

KOLCHINSKY A., LOURENÇO A., WU H.-Y., LI L. & ROCHA L. M. (2015). Extraction of pharmacokinetic evidence of drug–drug interactions from the literature. *PloS one*, **10**(5), e0122199.

LIU S., TANG B., CHEN Q. & WANG X. (2016). Drug-drug interaction extraction via convolutional neural networks. *Computational and Mathematical Methods in Medicine*, **2016**.

MENG F. & MORIOKA C. (2015). Automating the generation of lexical patterns for processing free text in clinical documents. *Journal of the American Medical Informatics Association*, **22**(5), 980–986.

PEDREGOSA F., VAROQUAUX G., GRAMFORT A., MICHEL V., THIRION B., GRISEL O., BLONDEL M., PRETTENHOFER P., WEISS R., DUBOURG V., VANDERPLAS J., PASSOS A., COURNAPEAU D., BRUCHER M., PERROT M. & DUCHESNAY E. (2011). Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

RZHETSKY A., SERINGHAUS M. & GERSTEIN M. B. (2009). Getting started in text mining : Part two. *PLoS Comput Biol*, **5**(7), e1000411.

SEGURA-BEDMAR I., MARTÍNEZ P. & HERRERO ZAZO M. (2013). Semeval-2013 task 9 : Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2 : Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, p. 341–350 : Association for Computational Linguistics.

STENETORP P., PYYSALO S., TOPIĆ G., OHTA T., ANANIADOU S. & TSUJII J. (2012). Brat : A web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL '12*, p. 102–107, Stroudsburg, PA, USA : Association for Computational Linguistics.

SUN X., DONG K., MA L., SUTCLIFFE R. F. E., HE F., CHEN S.-S. & FENG J. (2019). Drug-drug interaction extraction via recurrent hybrid convolutional neural networks with an improved focal loss. *Entropy*, **21**, 37.

YI Z., LI S., YU J., TAN Y., WU Q., YUAN H. & WANG T. (2017). Drug-drug interaction extraction via recurrent neural network with multiple attention layers. In *International Conference on Advanced Data Mining and Applications*, p. 554–566 : Springer.

ZHENG W., LIN H., LUO L., ZHAO Z., LI Z., ZHANG Y., YANG Z. & WANG J. (2017). An attention-based effective neural model for drug-drug interactions extraction. In *BMC Bioinformatics*.

