

Identification des problèmes d'annotation pour l'extraction de relations

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

L'annotation d'un corpus est une tâche difficile et laborieuse, notamment sur des textes de spécialité comme les textes biomédicaux. Ainsi, dans un contexte comme l'extraction des interactions aliment-médicament (FDI), l'annotation du corpus POMELO a été réalisée par un seul annotateur et présente des risques d'erreur. Dans cet article, nous proposons d'identifier ces problèmes d'annotation en utilisant un corpus Silver Standard (CSS) que nous établissons à partir d'un vote majoritaire parmi les annotations proposées par des modèles entraînés sur un domaine similaire (interaction médicament-médicament – DDI) et l'annotation manuelle à évaluer. Les résultats obtenus montrent que l'annotation dans POMELO est considérablement éloignée du CSS. L'analyse des erreurs permet d'en identifier les principales causes et de proposer des solutions pour corriger l'annotation existante.

ABSTRACT

Identification of annotation problem for the relation extraction.

Annotating a corpus is a difficult and time-consuming task, especially on texts of specific field such as biomedical texts. Thus, in the context of the extraction of food-drug interactions (FDI), the POMELO corpus annotation has been performed by a single annotator and presents risks of errors. In this article, we propose to identify these annotation problems by referring to a Silver Standard (CSS) corpus that we establish from a majority vote among the annotations proposed by models trained on a similar domain (drug-drug interaction) and the manual annotation to be evaluated. The obtained results show that the annotation in POMELO is far from the CSS. Error analysis helps to identify the main causes and to suggest solutions to correct the existing annotation.

MOTS-CLÉS : Annotation, extraction de relation, corpus biomédical, interaction aliment-médicament.

KEYWORDS: Annotation, relation extraction, biomedical corpus, food-drug interaction.

1 Introduction

Il peut être difficile dans de nombreuses circonstances d'établir un corpus selon les pratiques habituelles, c'est-à-dire une double annotation avec un consensus. C'est notamment le cas dans les domaines spécialisés où mobiliser plusieurs experts pour réaliser des annotations peut être compliqué. En pratique, les corpus annotés sont développés lors des travaux préliminaires, comme preuve de concept ou pour définir le guide d'annotation. Étant donné qu'aucun consensus n'a été effectué, on peut s'attendre à ce que de tels corpus comportent des problèmes d'incohérence ou d'ambiguïté.

Dans ce contexte, il est possible d'envisager d'améliorer l'annotation effectuée par un seul annotateur pour réduire le coût de l'annotation. Pour atteindre cet objectif, nous considérons la tâche d'extraction de relations spécifiques au domaine, en particulier l'extraction de l'interaction entre aliments et médicaments (*Food-drug interaction* – FDI) pour laquelle il existe très peu de corpus annotés ou de modèles d'extraction automatique.

À l'instar des interactions entre médicaments (*Drug-drug interaction* – DDI), les FDI correspondent à l'absorption, distribution ou l'effet inattendu d'un médicament, provoqués par l'ingestion d'un aliment (Doogue & Polasek, 2013). 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).

Toutefois, à la différence des DDI (Aagaard & Hansen, 2013) ou les effets indésirables des médicaments (Aronson & Ferner, 2005) qui sont répertoriés dans des bases de données telles que DrugBank¹ ou Thériaque², les FDI figurent rarement dans les bases de connaissances et sont généralement dispersées dans des sources hétérogènes (Wishart *et al.*, 2017). En outre, dans DrugBank par exemple, les informations sont principalement stockées sous forme de phrases.

Des travaux précédents ont été réalisés pour l'extraction automatique de ces informations (Hamon *et al.*, 2017) et ont conduit à la définition d'un ensemble de 639 résumés Medline (corpus POMELO) annotés par un externe en pharmacie selon 21 types de relation avec un grand déséquilibre en termes de nombre d'exemples par type. Les tentatives d'entraînement d'un système de détection automatique des interactions entre aliments et médicaments sur ce corpus ont conduit à de faibles performances. ((Randriatsitohaina & Hamon, 2019)) Nous émettons l'hypothèse que l'annotation du corpus, étant faite par une seule personne, peut présenter des incohérences. Afin de les identifier, nous proposons de tirer parti du corpus existant sur les DDI qui est un domaine similaire afin d'évaluer l'annotation POMELO selon un schéma d'annotation de plus haut niveau (4 types de relations DDI). Nos contributions se concentrent sur la définition d'un corpus Silver Standard (CSS) à partir d'un consensus avec les annotations des modèles d'extraction des DDI, l'analyse de l'annotation POMELO par rapport au CSS, l'identification des incohérences et la proposition de solutions pour corriger l'annotation existante.

2 État de l'art

Le développement d'un corpus Silver Standard (CSS) a été introduit pour réduire les coûts d'annotation de corpus. L'idée est d'appliquer des systèmes d'annotation sur un corpus et d'harmoniser l'annotation finale en utilisant un système de vote. Rebholz-Schuhmann *et al.* (2010) a produit un corpus pour l'identification des entités nommées à partir de textes biomédicaux. Lu *et al.* (2011) utilisent un algorithme espérance-maximisation (Expectation Maximisation) pour déduire la vérité terrain, basé uniquement sur les soumissions des équipes participantes à un défi sur le normalisation de gènes. Suivant le même principe, nous établissons un CSS obtenu par un système de vote sur les annotations des modèles d'extraction des DDI pour identifier ensuite les problèmes d'annotation des relations.

Certaines approches utilisent l'apprentissage actif afin d'améliorer l'annotation d'un corpus bruyant

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

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

((Rehbein & Ruppenhofer, 2017)). Le processus est basé sur le choix d'une instance mal étiquetée par détection d'erreur, en donnant les instances sélectionnées à l'oracle (l'annotateur humain), pour être manuellement désambiguïsées et ajoutées aux données d'apprentissage.

Une approche de détection des erreurs d'annotation est proposée par Ménard & Mougeot (2019), basée sur le regroupement des instances correspondant à chaque valeur d'étiquette "silver" unique, puis sur la considération des valeurs aberrantes comme des étiquettes éventuellement bruyantes qui doivent être ré-annotées par l'expert. Dans notre cas, nous utilisons le résultat de modèles entraînés sur les DDI pour identifier les éventuelles erreurs d'annotation.

Différents types d'approches ont été explorés pour extraire les DDI. Parmi elles, Kolchinsky *et al.* (2015) se concentrent sur l'identification des phrases pertinentes et des résumés pour l'extraction des preuves pharmacocinétiques. Étant donné que nous nous concentrons sur la correction d'annotations, nous proposons de travailler sur les phrases du corpus POMELO qui ont été annotées comme pertinentes. Ben Abacha *et al.* (2015) proposent une approche basée sur un SVM combinant : (i) les caractéristiques décrivant les mots dans le contexte des relations à extraire, (ii) les noyaux composites utilisant des arbres de dépendance. Leur 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. Dans notre tâche, nous utilisons un SVM linéaire combiné avec des descripteurs sémantiques et syntaxiques qui nous permettent d'obtenir des résultats comparables. 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 un CNN pour l'extraction des DDI. 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) proposent une méthode hybride combinant un réseau de neurones récurrent et convolutif induisant une amélioration de 3%. Le réseau convolutif profond de (Dewi *et al.*, 2017) permet de couvrir de longues phrases qui ont des jeux de données de DDI typiques et d'obtenir une performance de 0,86. BERT (Devlin *et al.*, 2019), un modèle de représentation de mots contextualisés permet obtenir les meilleurs résultats sur les tâches de TAL y compris l'extraction de relations. Lee *et al.* (2019) ont entraîné et adapté BERT sur des corpus biomédicaux pour être efficaces dans les tâches d'exploration de textes biomédicaux. Nous avons adapté BioBert pour l'extraction des DDI.

3 Corpus

Notre approche est mise en œuvre sur le corpus POMELO qui contient des annotations de FDI mais dont la qualité doit être améliorée, et le corpus DDI fournissant des annotations d'interactions entre médicaments et dont la qualité est avérée suite à son utilisation dans des campagnes d'évaluation.

Corpus POMELO Des études ont déjà été menées sur les FDI durant lesquelles l'ensemble de données POMELO 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) en anglais, collectés à partir

du portail PubMed³. 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. Étant donné que nous examinons les interactions aliment-médicament, 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 900 phrases étiquetées avec 13 types de relations (avec le nombre d’instances) : *decrease absorption* (53), *slow absorption* (15), *slow elimination* (15), *increase absorption* (39), *speed up absorption* (1), *new side effect* (4), *negative effect on drug* (88), *worsen drug effect* (8), *positive effect on drug* (21), *improve drug effect* (6), *no effect on drug* (109), *without food* (13), *unspecified relation* (528).

Corpus DDI Le corpus DDI⁴ est composé de 4 037 instances de relations issues de la base de données DrugBank et des résumés Medline. Il a été utilisé lors de SemEval 2013 (Segura-Bedmar *et al.*, 2013) et propose quatre types d’annotation : *Conseil* (819 instances) pour une recommandation concernant l’utilisation concomitante de deux médicaments, *Effet* (1 700 instances) pour l’effet de la DDI, *Mécanisme* (1322 instances) pour la pharmacodynamique et la pharmacocinétique du médicament, *Interaction* (188 instances) si aucune information sur l’interaction n’est fournie.

4 Méthode

Afin d’évaluer la qualité de l’annotation du corpus POMELO, nous proposons une approche en trois étapes décrite à la figure 1 .

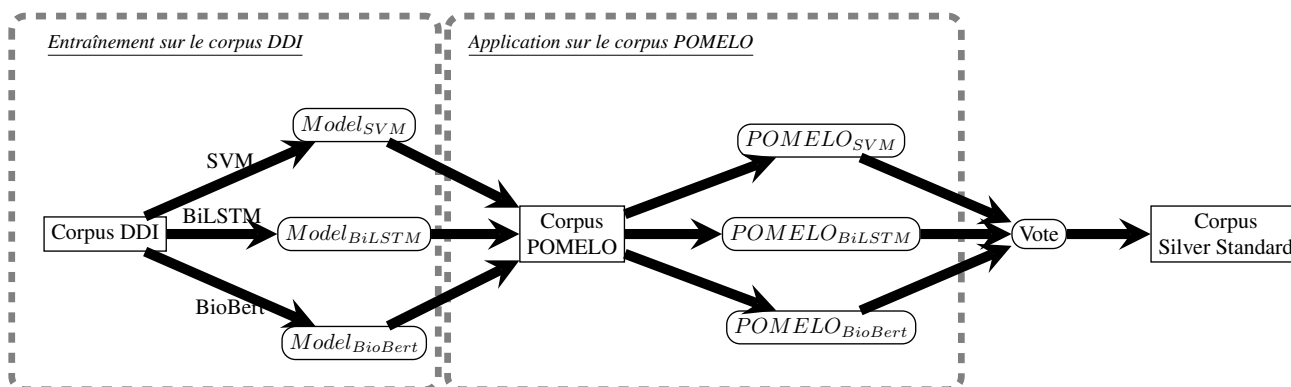


FIGURE 1 – Architecture de l’approche.

Étape 1 : Apprentissage sur le corpus DDI Nous proposons de tirer parti des méthodes d’extraction de DDI pour obtenir des annotations sur le corpus POMELO. Pour ce faire, nous avons entraîné 3 types de modèles sur le corpus DDI :

- **SVM.** SVM (Vapnik, 1995) ayant montré de bonnes performances dans l’état de l’art, nous proposons d’entraîner un modèle SVM linéaire avec les TF/IDF des formes fléchies, lemmes,

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

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

catégories morpho-syntaxiques des mots, catégories sémantiques des termes (aliment, médicament, supplément alimentaire) des contextes avant, entre et après les deux arguments de la relation.

- **BiLSTM.** Nous entraînons un modèle récurrent BiLSTM (Gers *et al.*, 2000) avec les plongements de mots Skipgram fastText (Bojanowski *et al.*, 2017) de taille 300 et les plongements de position.
- **BioBert.** Nous utilisons BioBert version 1.1 Lee *et al.* (2019) entraîné avec un vocabulaire de 4,5 milliards de mots issus de PubMed.

Étape 2 : Application des modèles sur le corpus POMELO Une fois les modèles entraînés sur les données DDI, nous les appliquons sur le corpus POMELO. Cette étape permet d’obtenir des annotations dites artificielles des phrases pertinentes de POMELO selon les types de relations DDI. À ce niveau, il nous faut une correspondance des types de relations DDI et FDI pour pouvoir évaluer les annotations manuelles faites sur POMELO. Pour ce faire, nous avons établi une correspondance entre les types de DDI et les types de FDI, validée par les experts, comme suit : les interactions impliquant la pharmacocinétique des médicaments (absorption, élimination, métabolisme) sont étiquetées *Mécanisme* (123 instances), les interactions impliquant les effets des médicaments (positifs, négatifs, effet secondaire) sont étiquetées *Effet* (236 instances), le type de relation *Without food* qui est une contre-indication à la prise d’aliment avec un médicament est étiqueté *Conseil* (13 instances) et les interactions *Unprecised relation* sont étiquetées *Interaction* (528 instances). Nous appellerons cette correspondance POMELO-DDI.

Étape 3 : Vote majoritaire parmi les annotations obtenues Dans cette section, nous proposons d’établir une annotation de corpus Silver Standard (CSS) en appliquant un système de vote majoritaire sur les annotations obtenues par les modèles SVM, BiLSTM et BioBert ainsi que l’annotation POMELO-DDI décrite à la section 4. Nous prenons en compte l’annotation POMELO-DDI étant donnée qu’elle a été faite par un expert donc peut apporter des informations pertinentes sur certains types de relation. Ainsi, pour chaque instance de relation, nous gardons le type de relation le plus fréquent parmi les quatre types d’annotations. En cas d’absence d’accord, nous privilégions l’annotation POMELO-DDI qui est basée sur l’annotation humaine.

Étape 4 : Analyse des erreurs d’annotation dans POMELO-DDI Nous utilisons l’annotation Silver Standard obtenue à la section 4 comme pré-annotation et nous identifions les problèmes d’annotation des relations où l’annotation manuelle et le Silver Standard diffèrent.

5 Expériences et résultats

Pour l’étape d’apprentissage sur les données DDI, nous évaluons les modèles selon la macro-moyenne des F1-mesures obtenues par validation croisée en 3 échantillons. De manière générale, les résultats obtenus par les modèles sont prometteurs, avec des F1-mesures allant de 0,78 avec SVM à 0,91 avec BioBert. C’est d’autant plus important que ces performances décriront la fiabilité des annotations que les modèles fourniront. Les performances du modèle BioBert surpassent de loin celle des autres modèles. Cela confirme l’efficacité des plongements de mot contextualisé appris sur des documents du domaine spécifique.

Une fois les modèles entraînés, nous les appliquons sur le corpus POMELO en gardant les mêmes configurations (paramètres et descripteurs). Cela nous permet d'obtenir 4 annotations de type DDI sur les données POMELO selon les modèles SVM, BiLSTM et BioBert ainsi que l'annotation POMELO-DDI décrite à la section 4. Le vote majoritaire appliqué sur ces 4 annotations nous permet d'avoir une sorte de consensus (corpus Silver Standard) que nous utilisons comme annotation de référence pour évaluer la performance des modèles obtenu à l'étape 1 sur l'annotation du corpus POMELO selon les types DDI. Les résultats de cette évaluation indiquent que BioBert fournit l'annotation la plus en accord avec le consensus (kappa de Cohen 0,86) tandis que les annotations manuelles POMELO-DDI en semblent être les plus éloignées (kappa de Cohen 0,15).

Nous proposons alors d'effectuer une analyse des erreurs afin d'identifier la source de ces différences d'annotation. Une des erreurs les plus fréquentes est l'annotation des titres ou des descriptifs des expériences avec le type *interaction* alors que la phrase contient des vocabulaires spécifiques au mécanisme ou aux effets de médicaments, d'où la faible précision (cf. tableau 1). Par exemple l'instance "*Nous décrivons ici deux études pharmacocinétiques pour quantifier l'impact des **aliments** sur l'absorption de la **ziprasidone** chez des volontaires sains.*" a été annotée manuellement *interaction* et *mécanisme* par le vote. Étant donné que la conclusion de l'étude qui devrait fournir plus d'information sur le médicament et l'aliment considérés est donnée un peu plus tard dans le résumé, il est possible de ne pas annoter les titres ou les descriptifs réduisant ainsi l'ambiguïté qui peut avoir lieu. Nous avons également observé que le type *interaction* est utilisé comme type par défaut et n'est pas précisé dans certains cas alors qu'il est préférable de privilégier des types plus précis, ce que font les modèles. Par exemple, l'instance "*Bloc cardiaque auriculo-ventriculaire complet développé en raison de l'utilisation du **vérapamil** et de la consommation de **miel***" a été annoté manuellement *interaction* et *effet* par le vote. Dans certains cas, l'annotateur utilise le type *interaction* faute de type plus précis plus adapté. Par exemple l'instance "*En ce qui concerne les autres facteurs alimentaires pouvant interagir avec les **anticoagulants oraux**, le patient doit être mis en garde concernant les suppléments de **vitamines A, E et C et d'alcool** utilisés de façon chronique ou ingérés en grande quantité.*" a été annoté manuellement *interaction* et *conseil* par le vote. Ce type de relation pourrait donc être intégré dans le guide d'annotation, ce qui permettrait d'améliorer le rappel (cf. tableau 1).

Nous avons également observé que certaines instances impliquent plusieurs types de relation simultanément, ce qui provoque une divergence entre les différentes annotations. Par exemple l'instance "*Ces résultats démontrent que l'administration de **ziprasidone** avec des **aliments** est cruciale pour assurer une biodisponibilité optimale, fiable et dépendante de la dose et donc un contrôle et une tolérabilité prévisibles des symptômes.*" mentionne à la fois le mécanisme du médicament (biodisponibilité) ainsi que son effet (contrôle, tolérabilité). Les 2 types de relation ont généralement un lien de cause à effet, donc plusieurs solutions sont possibles : (1) privilégier une relation par rapport aux autres, (2) annoter les différents types en même temps (multi-étiquette), (3) scinder l'instance en plusieurs instances de relation différentes.

6 Conclusion

Dans cet article, nous proposons d'identifier les problèmes durant l'annotation des relations dans des textes biomédicaux, en particulier les interactions médicament-aliment (FDI) rassemblées dans le corpus POMELO. Pour ce faire, nous proposons de tirer parti des données d'un domaine similaire (interactions médicament-médicament – DDI) afin d'établir un corpus Silver Standard CSS) obtenu par

Relation	P	R	F1	Nb instance
Conseil	0,54	0,11	0,19	61
Effet	0,64	0,57	0,60	264
Interaction	0,12	0,98	0,22	65
Mécanisme	0,95	0,23	0,37	510
moy / total	0,56	0,48	0,52	900

TABLE 1 – Macro précision, rappel, F1-mesure entre le corpus POMELO-DDI et le corpus Silver Standard.

un vote majoritaire parmi les annotations des modèles d'extraction automatique ainsi que l'annotation à évaluer. Entraînés sur le corpus DDI, les performances des modèles allant de 0,78 jusqu'à 0,91 de F1-mesure obtenues avec des modèles SVM, BiLSTM et BioBert sont prometteuses et indicatrices de la fiabilité des annotations que ces modèles fourniront. Une fois les modèles entraînés, nous les appliquons sur les instances POMELO puis nous appliquons un système de vote majoritaire pour garder le type de relation affecté le plus souvent à chaque instance constituant ainsi le CSS. Le CSS servira de référence afin d'identifier les sources de divergences avec les annotations manuelles faites sur le corpus POMELO. Les résultats obtenus ont montré que l'annotation dans le corpus POMELO est considérablement éloignée du corpus Silver Standard. L'analyse des erreurs a permis d'en identifier les principales causes et de proposer des solutions à savoir limiter les annotations aux phrases informatives et fiables de FDI, privilégier les types de relation précis au type générique, ajouter d'autres types de relation comme *conseil*, dans le cas de l'implication de plusieurs types de relation dans une instance, privilégier une relation par rapport à l'autre ou annoter les types en même temps (multi-étiquette) ou scinder l'instance en plusieurs instances de relation différentes. Dans la suite du travail, nous proposons d'appliquer ces solutions dans une phase de correction d'annotation en utilisant le Silver Standard comme pré-annotation et entraîner des modèles d'extraction automatique sur l'annotation obtenue.

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.
- BOJANOWSKI P., GRAVE E., JOULIN A. & MIKOLOV T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, **5**, 135–146.

- 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. DOI : [10.1186/s12859-018-2021-9](https://doi.org/10.1186/s12859-018-2021-9).
- DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2019). BERT : Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1 (Long and Short Papers)*, p. 4171–4186, Minneapolis, Minnesota : Association for Computational Linguistics. DOI : [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423).
- 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. *Therapeutic Advances in Drug Safety*, **4**(1), 5–7. PMID 25083246, DOI : [10.1177/2042098612469335](https://doi.org/10.1177/2042098612469335).
- GERS F. A., SCHMIDHUBER J. & CUMMINS F. A. (2000). Learning to forget : Continual prediction with lstm. *Neural Computation*, **12**(10), 2451–2471.
- 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. J., CANCELON P., WIDMER W. W. & GREENBLATT D. J. (2011). The effect of grapefruit juice on drug disposition. *Expert Opinion on Drug Metabolism & Toxicology*, **7**(3), 267–286. PMID 21254874, DOI : [10.1517/17425255.2011.553189](https://doi.org/10.1517/17425255.2011.553189).
- 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. DOI : [10.1371/journal.pone.0122199](https://doi.org/10.1371/journal.pone.0122199).
- LEE J., YOON W., KIM S., KIM D., KIM S., SO C. H. & KANG J. (2019). BioBERT : a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. DOI : [10.1093/bioinformatics/btz682](https://doi.org/10.1093/bioinformatics/btz682).
- LIU S., TANG B., CHEN Q. & WANG X. (2016). Drug-drug interaction extraction via convolutional neural networks. *Computational and Mathematical Methods in Medicine*, **2016**. DOI : [10.1155/2016/6918381](https://doi.org/10.1155/2016/6918381).
- LU Z., KAO H.-Y., WEI C.-H., HUANG M., LIU J., KUO C.-J., HSU C.-N., TSAI R. T.-H., DAI H.-J., OKAZAKI N., CHO H.-C., GERNER M., SOLT I., AGARWAL S., LIU F., VISHNYAKOVA D., RUCH P., ROMACKER M., RINALDI F., BHATTACHARYA S., SRINIVASAN P., LIU H., TORII M., MATOS S., CAMPOS D., VERSPOOR K. M., LIVINGSTON K. M. & WILBUR W. J. (2011). The gene normalization task in BioCreative III. *BMC Bioinformatics*, **12**(S2). DOI : [10.1186/1471-2105-12-S8-S2](https://doi.org/10.1186/1471-2105-12-S8-S2).

- MÉNARD P. A. & MOUGEOT A. (2019). Turning silver into gold : error-focused corpus reannotation with active learning. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, p. 758–767, Varna, Bulgaria. DOI : [10.26615/978-954-452-056-4_088](https://doi.org/10.26615/978-954-452-056-4_088).
- RANDRIATSITOHAINA T. & HAMON T. (2019). Extracting food-drug interactions from scientific literature : relation clustering to address lack of data. In *International Conference on Intelligent Text Processing and Computational Linguistics*, La Rochelle, France. HAL : [hal-02122766](https://hal.archives-ouvertes.fr/hal-02122766).
- REBHOLZ-SCHUHMAN D., JIMENO YEPES A. J., VAN MULLIGEN E. M., KANG N., KORS J., MILWARD D., CORBETT P., BUYKO E., TOMANEK K., BEISSWANGER E. & HAHN U. (2010). The CALBC silver standard corpus for biomedical named entities — a study in harmonizing the contributions from four independent named entity taggers. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta : European Language Resources Association (ELRA).
- REHBEIN I. & RUPPENHOFER J. (2017). Detecting annotation noise in automatically labelled data. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 1160–1170, Vancouver, Canada : Association for Computational Linguistics. DOI : [10.18653/v1/P17-1107](https://doi.org/10.18653/v1/P17-1107).
- 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.
- VAPNIK V. N. (1995). *The Nature of Statistical Learning Theory*. Springer.
- WISHART D., DJOUMBOU Y., GUO A. C., LO E., MARCU A., GRANT J., SAJED T., JOHNSON D., LI C., SAYEEDA Z., ASSEMPOUR N., IYNKKARAN I., LIU Y., MACIEJEWSKI A., GALE N., WILSON A., CHIN L., CUMMINGS R., LE D. & WILSON M. (2017). Drugbank 5.0 : A major update to the drugbank database for 2018. *Nucleic Acids Research*, **46**(D1). DOI : [10.1093/nar/gkx1037](https://doi.org/10.1093/nar/gkx1037).
- 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. *BMC bioinformatics*, **18**(1), 445. DOI : [10.1186/s12859-017-1855-x](https://doi.org/10.1186/s12859-017-1855-x).