Extracting Food-Drug Interactions from Scientific Literature: Relation Clustering to Address Lack of Data

Tsanta Randriatsitohaina[1] and Thierry Hamon[1,2(✉)]

[1] LIMSI, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay, France
{tsanta,hamon}@limsi.fr
[2] Université Paris 13, Sorbonne Paris Cité, 93430 Villetaneuse, France

Abstract. Food-Drug Interaction (FDI) occurs when food and drug are taken simultaneously and cause unexpected effect. This paper tackles the problem of mining scientific literature in order to extract these interactions. We consider this problem as a relation extraction task which can be solved with classification method. Since Food-Drug Interactions need a fine-grained description with many relation types, we face the data sparseness and the lack of examples per type of relation. To address this issue, we propose an effective approach for grouping relations sharing similar representation into clusters and reducing the lack of examples. Cluster labels are then used as labels of the dataset given to classifiers for the FDI type identification. Our approach, relying on the extraction of relevant features before, between, and after the entities associated by the relation, improves significantly the performance of the FDI classification. Finally, we contrast an intuitive grouping method based on the definition of the relation types and a unsupervised clustering based on the instances of each relation type.

Keywords: Food-Drug interactions · Information extraction · Text mining

1 Introduction

Although knowledge bases or terminologies exist in specialized domains, updating these information often requires to access unstructured data such as scientific literature. The problem deeply occurs when focusing on a new knowledge which has no recording in terminological resources yet. Thus, while drug interactions [1] or drug adverse effects [2] are listed in databases such as DrugBank[1] [14] or Thériaque[2], other information such as interactions between drug and food is barely listed in knowledge and mainly scattered in heterogeneous sources [14]. Besides, information is mainly stored as sentences. Actually, while food-drug interactions can correspond to various types of adverse drug effects and lead to harmful consequences on the patient's health and well-being, they are less known and studied and consequently very sparse in the scientific

---

[1] https://www.drugbank.ca/.
[2] http://www.theriaque.org.

literature. Similarly to interactions between drugs, Food-Drug Interaction (FDI) corresponds to the appearance of an unexpected effect. For example, grapefruit is known to inhibit the effect of an enzyme involved in the metabolism of several drugs [7]. Other foods may affect the absorption of a drug or its distribution in the organism [5].

The relation extraction task in biomedical texts generally consists in the identification of the related entities and the recognition of the relation category. In this article, we address the automatic identification of interaction statements between drug and food in abstracts of scientific articles issued from the Medline database.

To extract this information from the abstracts, we face several difficulties: (1) drugs and foods are very variable in the summaries. Drug can be mentioned by its common international name or active drug substances, while foods may be referenced by a particular nutrient, component or food family; (2) the interactions are described in a rather precise way in the texts, which leads to a limited number of examples; (3) the available set of annotations does not include the different types of interaction homogeneously and the learning set is often unbalanced.

Our contributions focus on FDI extraction and improvement of previous classification results by proposing a relation representation which addresses the lack of data, applying clustering method on type of relations, and using cluster labels in a classification step for identification of FDI type.

2   Related Work

Various types of approaches have been explored to extract relations from biomedical texts. Some approaches combine patterns and CRF for recognition of symptoms in biomedical texts [8]. Other approaches generate automatically lexical patterns for processing free text in clinical documents relying on a multiple sequential alignment to identify similar contexts [12]. Sentence's verb is compared to a list of verbs known as indicating relation to determine the relation between entities [15]. Then they construct the syntax dependency tree around the verbs to identify the related entities.

Drug-drug Interaction (DDI) extraction described by [3] is similar to our food-drug interaction extraction problem even if we need to identify much more types of relation (see Sect. 3). Our method joins their two steps approach for DDI detection and classification in which we added a relevant sentences selection step as proposed in [10]. [10] focus on the identification of relevant sentences and abstracts for extraction of pharmacokinetic evidence of DDI. [9] built two classifiers for DDI extraction: a binary classifier to extract interacting drug pair and a DDI type classifier to associate the interacting pairs with predefined relation categories. [4] consider the extraction of protein localization relation as a binary classification. All the protein-location pairs appearing in the same sentence are considered as positive instances if they are related, and negative otherwise. In contrast, we use multi-class classification for relation type recognition. [11] propose a CNN-based method for DDI extraction. In their model, drug mentions in a sentence are normalized in the following way: the two considered drug names are replaced by `drug1` and `drug2` according to the occurrence order in the sentence, respectively, and all the other drugs are replaced by `drug0`. Other works use recurrent neural network model with multiple attention layers for DDI classification [17].

3   Dataset

Studies have already been conducted considering Food-Drug Interactions in which the POMELO dataset was developed [6]. This dataset consists of 639 abstracts of scientific articles from the medical field (269,824 words, 5,752 sentences). They were collected from the PubMed portal[3] by the query: (`"FOOD DRUG INTERACTIONS"[MH] OR "FOOD DRUG INTERACTIONS*"`) AND (`"adverse effects*"`). All 639 abstracts were annotated according to 9 types of entities and 21 types of relation in Brat [16] by a pharmacy resident. The annotations focus on information about relation between food, drug and pathologies.

Since we are considering Food-Drug Interactions in this paper, we construct our dataset by taking into account every couple of *drug* and *food* or *food-supplement* from POMELO dataset. The resulting dataset is composed of 831 sentences labelled with 13 types of relations: *decrease absorption*, *slow absorption*, *slow elimination*, *increase absorption*, *speed up absorption*, *new side effect*, *negative effect on drug*, *worsen drug effect*, *positive effect on drug*, *improve drug effect*, *no effect on drug*, *without food*, *non-precised relation*. The statistics of the dataset is given in Table 1. Meaning of the relation type is detailed in Sect. 4.1.

Table 1. Statistics of annotated relations by initial types

Relation type	#	Percentage
Non-precised relation	476	57.3%
Decrease absorption	49	5.9%
Positive effect on drug	19	2.3%
Negative effect on drug	85	10.2%
Increase absorption	38	4.6%
Slow elimination	15	1.8%
Slow absorption	15	1.8%
No effect on drug	109	13.1%
Improve drug effect	6	0.7%
Without food	9	1.1%
Speed up absorption	1	0.1%
Worsen drug effect	5	0.6%
New side effect	4	0.5%
Total	831	100%

4   Grouping Types of Relation

The distribution of our dataset is very unbalanced as shown in the Table 1. For instance, *speed up absorption* relation has only one example, which do not permit efficient generalization of the represented relation. This lack of examples is due to the fine-grained

---

[3] https://www.ncbi.nlm.nih.gov/pubmed/.

description of the relations. To solve this problem, we propose two methods for grouping relations sharing similarities in order to obtain more examples per group of relations. The first method relies on the definition of relation types (intuitive grouping) while the second one is based on unsupervised clustering of the relation instances.

4.1   Intuitive Grouping

In this section, we propose a very intuitive way for grouping Food-Drug relations. FDI identification task presents similarity with Drug-Drug Interaction, where two drugs taken together lead to a modification of their effects. ADME [5] (absorption, distribution, metabolism and excretion) relations are involved, but applying this grouping in POMELO dataset would require supplementary annotation process.

The intuitive grouping is done as below:

1. **Non-precised relation.** Instances labelled with '*non-precised relation*' do not give more precision about the relation involved. While we do not have information that would permit to combine them with another relation, they will be considered as one individual group, especially since they represent more than half of the data.
2. **No effect.** '*No effect on drug*' instances represent food-drug relations in sentences where it is explicitly expressed that the considered food has no effect on the drug, unlike other relations that express actual food-drug interactions. As a result, these instances are represented as one individual group.
3. **Reduction.** Since instances labelled with '*decrease absorption*', '*slow absorption*', '*slow elimination*' express diminution of action of drug under the influence of a food, they are grouped to form the *reduction* relation.
4. **Augmentation.** Similarly to *reduction* relation, instances labelled with '*increase absorption*', '*speed up absorption*' are grouped to form the *augmentation* relation.
5. **Negative.** This group includes instances labelled with '*new side effect*', '*negative effect on drug*', '*worsen drug effect*', '*without food*'. *negative effect on drug* express explicitly a negative effect of food on drug, '*worsen drug effect*' expresses a negative effect of the drug, *side effect* is generally an adverse effect of the drug that join a negative connotation, the same to '*without food*' that prevents from taking food with the considered drug.
6. **Positive.** By analogy with the *negative* relation, '*positive effect on drug*', '*improve drug effect*' are grouped to form the *positive* relation.

For the rest of the paper, we will note this intuitive grouping method ARNP that stands for Augmentation, Reduction, Negative and Positive.

At the end, we get 6 Food-Drug relation types with relatively balanced number of examples. Statistics of this new distribution are given in Table 2.

4.2   Unsupervised Clustering

Clustering is a data mining method that aims at dividing a set of data into different homogeneous groups, in that the data of each subset shares common characteristics, which most often correspond to similarity criteria defined by measures of distance

Table 2. Statistics of annotated relations by grouped types

Relation type	#	Percentage
Non-precised relation	476	57.3%
No effect on drug	109	13.1%
Reduction	79	9.5%
Augmentation	39	4.7%
Negative	103	12.4%
Positive	25	3%
Total	831	100%

between elements. To obtain a good partitioning, it is necessary to minimize intra-class inertia to obtain clusters as homogeneous as possible and maximize inter-class inertia in order to obtain well-differentiated subsets. In this section, we propose to use clustering method to group Food-Drug relations involving food effect on drug.

Relation Representation. In our case, the data to be clustered is Food-Drug relations. For this purpose, each relation should be represented by a set of features such that the resulting data $D = [F_1, F_2, ..., F_n]$, should be a vector of size n, where n is the number of relations to be clustered, F_i is a set of features representing relation R_i. The most natural way to get features F_i is to group every sentences S_i labelled by relation R_i in the initial dataset D_S: F_i = Concatenation(S_i) for S_i in D_S. We assume this representation as the baseline of our task.

To improve the relation representation, we propose a supervised approach to extract the more relevant features for relation R_i by training a n-classes SVM Classifier on the initial dataset D_S. SVM decision is based on an hyperplane that maximizes the margin between the samples and the separator hyperplane represented by $h(x) = w^T x + w_0$, where $x = (\mathbf{x}_1, ..., \mathbf{x}_N)^T x = (\mathbf{x}_1, ..., \mathbf{x}_N)^T$ is the vector of features and $w = (\mathbf{w}_1, ..., \mathbf{w}_N)^T$ the vector of weights. From these weights, we can determine the importance of each feature on SVM decision given by a matrix of feature coefficients C of size $n \times nf$ where n is the number of classes (here relations) and nf the number of features.

We propose to extract the nm most important features for each relation to represent the considered relation such that the relation R_i is represented by a vector of nm features which corresponds to the nm first positive features of the ith vector of C. The resulting dataset is a matrix $D = [F_1, F_2, ..., F_n]$ of size $n \times nm$ where n is the number of classes (here relations), nm is the number of features to extract, and F_i is the feature extracted to represent relation i.

However, relation representation is quite more complicated than word representation since the meaning of the sentence relies entirely on the two related arguments considered. In order to capture more accurately the expression of the relation in a sentence, we propose to use as features, lemmas before the first argument of the relation, lem-

mas between the two arguments, and lemmas after the second argument for the SVM classification. For the rest of the paper, we will note this method BBA-SVM.

Relation Clustering and FDI-Classification. Following the relation description in Sect. 4.1, we consider *non-precised relation* and *no effect on drug* relation as is, but the 11 others will be grouped into 4 clusters. We apply the approach proposed in Sect. 4.2 on sentences labelled by the 11 effect relations in POMELO dataset. The resulting data is a matrix D of size $11 \times nm$ where nm is the number of features extracted to represent a relation, that is given to an unsupervised clustering algorithm to be grouped into 4 clusters. The results is a vector of cluster labels $Cl = [Cl_1, Cl_2, ..., Cl_{11}]$ that contains 4 unique values where Cl_i is the cluster to which the relation R_i belongs. Once clusters defined, labels of sentences from the initial dataset are replaced by the cluster labels associated to the relation. Finally, we perform a 6-classes classification to identify FDI type.

This pipeline is summarized in Fig. 1. In this paper, we address the issue of lack of number of examples per relation by grouping relations with similar features. We assume an intuitive clustering way on relations involving food effect on drug, leading to 4 relations that are Augmentation, Reduction, Negative and Positive. Then we use clustering method to automatically group such relations according to features selected preliminary by a SVM classifier. Once effect relation clustered, we carry out a 6-classes-classification to identify the type involved in each sentence. So a configuration is composed of:

1. a relation representation step with number of features and features extraction method as parameters
2. a clustering step with clustering algorithm as parameter
3. a classification step with classifier and features as parameters
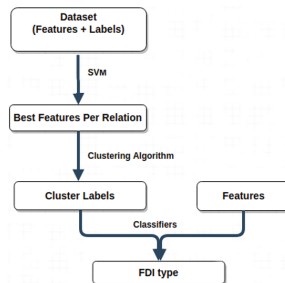
Fig. 1. Architecture of the approach - Relation Representation - Relation Clustering - FDI Classification

5   Experiments

Since our objective is to determine Food-Drug Interaction type, our experiments are focused on the performance of the relation classification from the POMELO dataset.

5.1   Clustering

Relation Representation. We experiment the impact of relation representation on the classification performance by varying the approach used to represent relation: (1) baseline - a relation R is represented by set of words of all sentences labelled by the relation R; (2) a lemma-SVM approach - lemmas are given to a SVM classifier and the more relevant features are extracted; (3) our BBA-SVM approach - lemmas before the first argument of the relation, lemmas between the two arguments, and lemmas after the second argument are given to a SVM classifier and best features are extracted; (4) inflected forms and lemmas, lemmas before the first argument of the relation, lemmas between the two arguments, and lemmas after the second argument are given to a SVM classifier and the more relevant features are extracted (ILBBA).

Clustering Algorithms. To evaluate our approach, we compare the performance of 4 clustering algorithms from Scikit-learn [13] implementation: (1) KMeans - to divide data into k subsets, central points k called centroids of partitions are identified such that the distance between the centroid and the points inside each partition is minimum; (2) Mini Batch K-Means - a variant of the KMeans algorithm which uses mini-batches to reduce the computation time, Mini-batches are subsets of the input data, randomly sampled in each training iteration; (3) Spectral clustering does a low-dimension embedding of the affinity matrix between samples, followed by a KMeans in the low dimensional space; (4) Agglomerative Clustering performs a hierarchical clustering using a bottom up approach: each observation starts in its own cluster, and clusters are successively merged together.

Clustering Evaluation. We use 4 metrics to evaluate the clustering assignment compared with the intuitive ARNP grouping method. Among them, we have (1) Adjusted Rand index that measures the similarity of the two assignments, ignoring permutations and with chance normalization; (2) Homogeneity - each cluster contains only members of a single class; (3) Completeness - all members of a given class are assigned to the same cluster; (4) Calinski-Harabaz Index is used to evaluate the model, where a higher Calinski-Harabaz score relates to a model with better defined clusters.

5.2   FDI Type Classification

Preprocessing. Each sentence of the dataset is preprocessed as following: numbers were replaced by the character '#' as proposed in [10], other special characters are removed, each word is converted to lower case.

Features. To evaluate the efficiency of our proposed approach, features are composed of inflected forms, lemmas, POS-tag of words, lemmas before the first argument of the relation, lemmas between the two arguments, and lemmas after the second argument.

Classification Models. Several classes of classification algorithms exist according to their mode of operation: (1) Linear models make classification decision based on value of a linear combination of the features; (2) Neighbors-based models classify an object

according to the vote of common classes of its nearest neighbors; (3) Tree-based models represent features as nodes of a decision tree with leaves as class labels; (4) Ensemble models combine the decision of multiple algorithms to obtain better classification performance; (5) Bayesian models are probabilistic classifiers based on Bayes theorem assuming independence between features. Classification models can be combined with preprocessing methods to improve the quality of the features thus facilitating decision-making.

In this experiment, we evaluate the performance of at least one classifier of each classes from Scikit-learn [13] implementation: (1) a Decision Tree (DTree), (2) a l2-linear SVM classifier (LSVC-l2), (3) a Logistic Regression (LogReg), (4) a Multinomial Naive Bayes (MNB), (5) a Random Forest Classifier (RFC), (6) a K-Nearest-Neighbors (KNN), and (7) a SVM combined with Select From Model feature selection algorithm (SFM-SVM).

Classification Quality. Since our goal is to extract Food-Drug interaction from texts, we evaluate our approach by its ability to identify such relations, which is measured by the score of the classifier in each configuration. 3 types of metrics are used in this case: precision (P), recall (R), F1-score (F_1). Considering that one of the challenge of the task is the imbalance of the numbers of examples per class, we compare the macro-scores, that computes scores per class then globally averages the scores, and the micro-scores, that computes scores over all individual decisions. Scores are obtained from a 10-fold cross-validation process.

6   Results and Discussion

Results presented in this section are the performance of a configuration in FDI type identification task using the POMELO dataset and cluster labels. Here the best result is achieved by 200 BBA-SVM features clustered by a Spectral Clustering algorithm, given as label in a SFM-SVM classifier using as features, lemmas before the first argument of the relation, lemmas between the two arguments, and lemmas after the second argument of the relation.

Table 3. Macro F1-score obtained using different methods for relation representation given to clustering algorithms *KMeans*, *MiniBatch-KMeans (MBKM)*, *Spectral Clustering*, *Agglomerative Clustering*

Relation representation	KMeans	MBKM	Spectral	Agglomerative
Baseline	0.362	0.394	**0.522**	0.374
Lemma	0.361	**0.405**	0.373	0.366
BBA-SVM	0.385	0.384	**0.58***	0.361
Inflected+Lemma+BBA	0.473	0.507	0.367	**0.517**
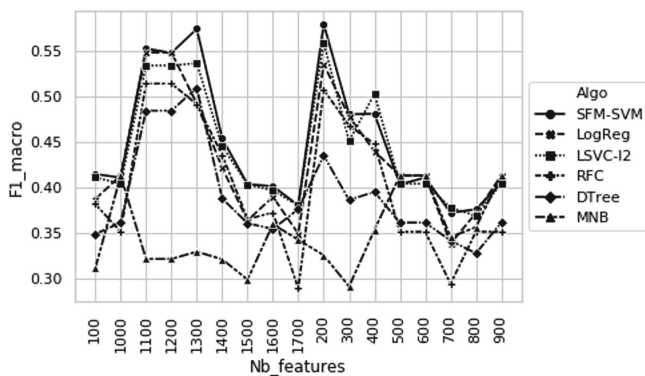
Fig. 2. Macro F1-score obtained on different models while varying the number of features to represent relation for clustering - *Spectral Clustering Model*, *BBA-SVM Method* - BBA features for classification

Table 4. Macro F1-score obtained while varying features used for relation classification after clustering - *KMeans*, *MiniBatchKMeans*, *SpectralClustering*, *AgglomerativeClustering* - BBA-SVM Representation

Feature	Without	ARNP	KMeans	MBKM	SpecC	AggloC
Inflected	0.379	0.378	0.416	**0.425**	0.573	0.422
Inflected + postag	0.388	0.368	0.397	0.413	0.559	0.406
Lemma	**0.41**	0.34	**0.462**	0.396	0.563	**0.449**
Inflected + lemma	0.38	**0.401**	0.423	**0.425**	0.575	0.439
Lemma + postag	0.403	0.392	0.444	0.386	0.56	0.436
Inflected + lemma + postag	0.387	0.395	0.425	**0.425**	0.567	0.443
Before + between + after	0.364	0.364	0.385	0.384	**0.58***	0.361
Inflected + lemma + before + between + after	0.383	0.384	0.394	0.402	0.576	0.403
Inflected + lemma + before + between + after + postag	0.379	0.396	0.383	0.398	0.566	0.404

Our BBA-SVM relation representation approach achieves the best F1-score 0.58 on FDI Classification (Table 3) with a difference of 0.23 from ARNP grouping and non-clustered data (Table 4). Thus, this score is obtained using only 200 features for relation clustering (Fig. 2) from the 1676 features composed by lemmas before the first argument of the relation, lemmas between the two arguments, and lemmas after the second argument for the SVM classification. This result justifies our assumption that a relation is characterized by specific features found in a particular position according to the 2 arguments of the relation. Joining this idea, the fact that feature selection method applied before SVM (SFM-SVM) (Table 6) produces a better performance suggest that some features are more important than others and focusing on them improve

Table 5. Clusters labels for each relation and scores obtained on different relation representation methods

		ARNP	Baseline	Lemma	BBA-SVM	ILBBA
Cluster labels	Decrease absorption	C1	C1	C1	C1	C1
	Improve drug effect	C2	C1	C3	C1	C3
	Increase absorption	C3	C3	C1	C1	C1
	Negative effect on drug	C4	C1	C1	C1	C1
	New side effect	C4	C1	C1	C4	C4
	Positive effect on drug	C2	C4	C1	C1	C1
	Slow absorption	C1	C1	C1	C1	C1
	Slow elimination	C1	C2	C2	C1	C1
	Speed up absorption	C3	C1	C2	C3	C1
	Without food	C4	C1	C4	C1	C2
	Worsen drug effect	C4	C1	C3	C2	C1
Clustering scores	Adjusted rand score	1.0	0.101	−0.132	−0.043	−0.043
	Calinski-Harabaz score	1.0	0.376	0.999	0.903	1.146
	Completeness score	1.0	0.519	0.312	0.431	0.431
	Homogeneity score	1.0	0.343	0.272	0.284	0.284
Classification scores	Precision macro	0.385	0.547	0.38	**0.589**	0.372
	Recall macro	0.368	0.518	0.381	**0.582**	0.371
	F1 macro	0.364	0.522	0.373	**0.58**	0.367
	F1 micro	0.599	0.656	0.647	**0.67**	0.652

Table 6. Scores obtained using features before + between + after - BBA-SVM representation - Spectral Clustering Algorithm

Model	Precision	Recall	F1-macro	F1-micro
DTree	0.45	0.441	0.435	0.614
LSVC-l2	0.572	0.558	0.56	0.668
LogReg	0.564	0.528	0.536	**0.674**
MNB	0.342	0.344	0.325	0.535
RFC	0.586	0.496	0.508	0.665
SFM-SVM	**0.589**	**0.582**	**0.58**	0.67
kNN	0.393	0.298	0.253	0.589

the decision-making of the classifier. The difference between micro-score and macro-score decrease from 0.13 with ARNP to 0.09, that suggest a reduction of the imbalance of data. Logistic Regression achieves the best result on micro-score but is a little less efficient in macro-score, which means that the model is more sensitive to imbalance of data than SVM models. Besides, the high score of Calinski-Harabaz (Table 5) implying that clusters are dense and well separated support the effectiveness of our approach. Nevertheless, the other clustering scores indicate an independant assignement from the ARNP method. It is explained while analyzing the labels assigned to each relation. In

Table 5, we observe also that 3 relations are represented individually and all others are grouped into one cluster. At first sight, there is no particular reason to explain the grouping. This suggest that the 3 individuals relation are explicitly different from the others but the rest are not sufficiently separable. It is also possible that the POMELO annotated corpus contains mistaken annotations. Actually, the one-annotator annotation can be improved according to our clustering approach, relying on manual validation, and including classification of relations without more precision. Indeed, these data represent more than half of the data, thus creating ambiguities making classification difficult. However, these results show that our approach produces a significant improvement on task of FDI type identification.

7 Conclusion and Future Work

Our paper contributes to the task of extraction of Food-Drug Interaction (FDI) from scientific literature, that we address as a relation extraction task. While applying supervised learning to this purpose, we face the lack of examples because of the high number of relation types. To address this issue, we propose to represent each relation by most important features extracted from SVM classification, then relations are grouped into clusters, and cluster labels are then used as relation labels on the initial dataset. Our approach is based on the assumption that relations are defined by a set of specific feature located in a particular position from the arguments of the relation. Following this idea, we use lemmas before the first argument of the relation, lemmas between the two arguments, and lemmas after the second argument for the SVM classification and extract from them the most important features used by SVM to make a decision of relation assignement. These features are given to clustering algorithms to obtain a cluster label for each relation, that is used as labels of POMELO dataset for FDI identification. Our approach achieves the best performance with 200 features grouped by Spectral Clustering algorithm, and classified by a pipeline of Select From Model feature selection and SVM classification. We get an improvement of 0.23 on F1-score from the ARNP and the non-clustered data. Besides, the decrease in difference between macro and micro average of F1-score suggest a reduction of the imbalance of data. Therefore, experiments results support the effectiveness of our approach. For future work, we will consider FDI type identification as a multilabel classification, using the cluster as first label, or a more domain-based labeling following the ADME classes [5] (Absorption - Distribution - Metabolism - Excretion) of Drug-Drug Interaction by transfer learning.

References

1. Aagaard, L., Hansen, E.: Adverse drug reactions reported by consumers for nervous system medications in Europe 2007 to 2011. BMC Pharmacol. Toxicol. **14**, 30 (2013)
2. Aronson, J., Ferner, R.: Clarification of terminology in drug safety. Drug Saf. **28**(10), 851–70 (2005)

3. Ben Abacha, A., Chowdhury, M.F.M., Karanasiou, A., Mrabet, Y., Lavelli, A., Zweigen-baum, P.: Text mining for pharmacovigilance: using machine learning for drug name recognition and drug-drug interaction extraction and classification. J. Biomed. Inform. **58**, 122–132 (2015)

4. Cejuela, J.M., et al.: LocText: relation extraction of protein localizations to assist database curation. BMC Bioinform. **19**(1), 15 (2018). https://doi.org/10.1186/s12859-018-2021-9

5. Doogue, M., Polasek, T.: The ABCD of clinical pharmacokinetics. Ther. Adv. Drug Saf. **4**, 5–7 (2013). https://doi.org/10.1177/2042098612469335

6. Hamon, T., Tabanou, V., Mougin, F., Grabar, N., Thiessard, F.: POMELO: medline corpus with manually annotated food-drug interactions. In: Proceedings of Biomedical NLP Workshop Associated with RANLP 2017, Varna, Bulgaria, pp. 73–80 (2017)

7. Hanley, M., Cancalon, P., Widmer, W., Greenblatt, D.: The effect of grapefruit juice on drug disposition. Expert Opin. Drug Metab. Toxicol. **7**(3), 267–286 (2011)

8. Holat, P., Tomeh, N., Charnois, T., Battistelli, D., Jaulent, M.-C., Métivier, J.-P.: Weakly-supervised symptom recognition for rare diseases in biomedical text. In: Boström, H., Knobbe, A., Soares, C., Papapetrou, P. (eds.) IDA 2016. LNCS, vol. 9897, pp. 192–203. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46349-0_17

9. Kim, S., Liu, H., Yeganova, L., Wilbur, W.J.: Extracting drug-drug interactions from literature using a rich feature-based linear kernel approach. J. Biomed. Inform. **55**, 23–30 (2015)

10. Kolchinsky, A., Lourenço, A., Wu, H.Y., Li, L., Rocha, L.M.: Extraction of pharmacokinetic evidence of drug-drug interactions from the literature. PLoS ONE **10**(5), e0122199 (2015)

11. Liu, S., Tang, B., Chen, Q., Wang, X.: Drug-drug interaction extraction via convolutional neural networks. Comput. Math. Methods Med. **2016** (2016). https://doi.org/10.1155/2016/6918381

12. Meng, F., Morioka, C.: Automating the generation of lexical patterns for processing free text in clinical documents. J. Am. Med. Inform. Assoc. **22**(5), 980–986 (2015). https://doi.org/10.1093/jamia/ocv012

13. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)

14. Wishart, S., et al.: DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. **46** (2017)

15. Song, M., Chul Kim, W., Lee, D., Eun Heo, G., Young Kang, K.: PKDE4J: entity and relation extraction for public knowledge discovery. J. Biomed. Inform. **57**, 320–332 (2015)

16. Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., Tsujii, J.: 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 2012, Stroudsburg, PA, USA, pp. 102–107. Association for Computational Linguistics (2012). https://dl.acm.org/citation.cfm?id=2380921.2380942

17. Yi, Z., et al.: Drug-drug interaction extraction via recurrent neural network with multiple attention layers. In: Cong, G., Peng, W.-C., Zhang, W.E., Li, C., Sun, A. (eds.) ADMA 2017. LNCS (LNAI), vol. 10604, pp. 554–566. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-69179-4_39