

# Biomedical Relation Extraction With Knowledge Graph-Based Recommendations

Diana Sousa  and Francisco M. Couto 

**Abstract**—Biomedical Relation Extraction (RE) systems identify and classify relations between biomedical entities to enhance our knowledge of biological and medical processes. Most state-of-the-art systems use deep learning approaches, mainly to target relations between entities of the same type, such as proteins or pharmacological substances. However, these systems are mostly restricted to what they directly identify on the text and ignore specialized domain knowledge bases, such as ontologies, that formalize and integrate biomedical information typically structured as direct acyclic graphs. On the other hand, Knowledge Graph (KG)-based recommendation systems already showed the importance of integrating KGs to add additional features to items. Typical systems have users as people and items that can range from movies to books, which people saw or read and classified according to their satisfaction rate. This work proposes to integrate KGs into biomedical RE through a recommendation model to further improve their range of action. We developed a new RE system, named K-BiOnt, by integrating a baseline state-of-the-art deep biomedical RE system with an existing KG-based recommendation state-of-the-art system. Our results show that adding recommendations from KG-based recommendation improves the system's ability to identify true relations that the baseline deep RE model could not extract from the text. The code supporting this system is available at <https://github.com/lasigeBioTM/K-BiOnt>.

**Index Terms**—Biomedical relation extraction, information retrieval, knowledge graphs, text mining.

## I. INTRODUCTION

THE exponential growth in scientific literature does not allow researchers to keep up with all recent advances in their respective fields and in adjacent areas that could be of interest to their research [1]. To this end, the field of Natural Language Processing (NLP) is mostly focused on automatic means to identify and extract relevant information from unstructured text [2], [3]. One of the prominent tasks of the NLP field is Relation Extraction (RE), which aims at extracting and

Manuscript received 27 August 2021; revised 18 December 2021 and 15 April 2022; accepted 3 May 2022. Date of publication 10 May 2022; date of current version 9 August 2022. This work was supported by FCT through Deep Semantic Tagger (DeST) Project under Grant PTDC/CCI-BIO/28685/2017 (<http://dest.rd.ciencias.ulisboa.pt/>), in part by LASIGE Research Unit under Grants UIDB/00408/2020 and UIDP/00408/2020, and in part by FCT and FSE through Ph.D. Scholarship under Grant SFRH/BD/145221/2019. (Corresponding author: Diana Sousa.)

The authors are with LASIGE, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal (e-mail: [dfsousa@lasige.di.fc.ul.pt](mailto:dfsousa@lasige.di.fc.ul.pt); [fjcouto@edu.ulisboa.pt](mailto:fjcouto@edu.ulisboa.pt)).

Digital Object Identifier 10.1109/JBHI.2022.3173558

classifying relations between entities of interest. Most Biomedical RE studies focus on extracting relations between the same and different type entities, such as diseases, genes, phenotypes, and pharmacological substances. Recently, there have been several advances regarding this task, mostly using deep learning techniques [4]. However, few make use of external sources of knowledge that are openly available such as domain-specific ontologies, which are highly popular in the biomedical domain. Using additional sources of knowledge can also result in more system explainability by facilitating the re-traceability of AI decisions to specific components of the models [5].

An ontology is a structured way of providing a common vocabulary in which shared knowledge is represented [6]. Biomedical ontologies are usually structured as directed acyclic graphs. Each node corresponds to an entity and the edges correspond to known relations between those entities of type *is-a*. Some of the most prominent ever-evolving biomedical ontologies are the Gene Ontology (GO) [7], the Chemical Entities of Biological Interest (ChEBI) [8], the Disease Ontology (DO) [9], and the Human Phenotype Ontology (HPO) [10]. Yet, researchers do not incorporate this structured information in most biomedical RE deep learning models.

On the other hand, Knowledge Graph (KG)-based Recommendation systems already showed the importance of external sources of knowledge to add additional features to items when using deep learning models [11]–[13]. This focus on external sources of knowledge could be highly relevant for the biomedical RE field since most researchers focus their work on already known relations between biomedical entities, which implies that a large volume of relations are not explicitly described in the literature. Consequently, systems that rely solely on available literature to identify these relations do not have enough information to establish more complex interactions. Therefore, we need to go beyond the text and uncover how to integrate entity annotations knowledge into our systems, as in most recent recommendation systems.

In this work, we propose to integrate a KG-based recommendation model into biomedical RE to answer the following question: *Can recommendations add value to the biomedical RE task, enhancing their range of action?*

The first step towards biomedical RE based on recommendation was to adapt three publicly available RE datasets into the standardized recommender systems format of <user-item-rating>. We chose the PGR-crowd [14] dataset that describes relations between human phenotypes and genes (Fig. 1), the DDI Corpus [15]) that describes relations between drugs/chemicals,

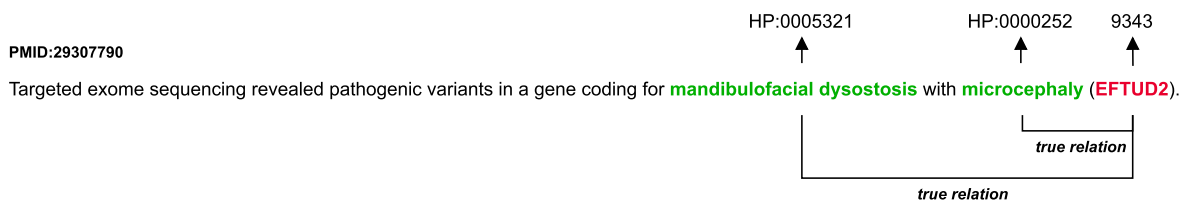


Fig. 1. An example of an annotated sentence retrieved from the PGR-crowd dataset from article PMID:29307790. The phenotype entities are linked to the HPO ontology (HP:0005321 and HP:0000252), and the gene entity is linked to the NCBI gene database (9343). We simplified the sentence to facilitate reading comprehension.

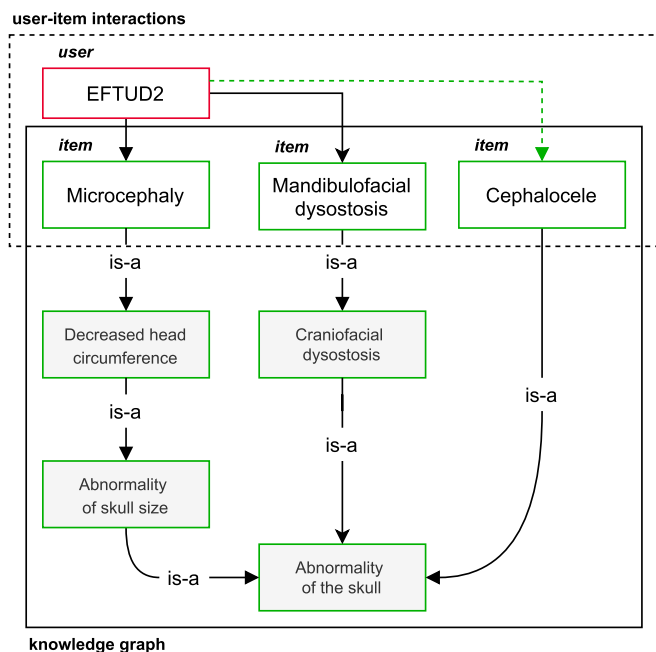


Fig. 2. An example of a relation recommendation retrieved from the adapted PGR-crowd dataset [14]. This example illustrates how a graph connection can contribute to a new relation recommendation.

and the BC5CDR Corpus [16], that describes drugs/chemicals interactions with diseases, to demonstrate the range of our approach.

To make the adaption of biomedical RE datasets to the <user-item-rating> recommendation format, we had to first decide which entities would be considered the users (*user* entities) and the items (*item* entities), for the datasets that had different type entities (PGR-crowd and BC5CDR dataset). Our choices for *item* entities, described in detail in Section III under the A. Datasets sub-section, were to give priority to entities that were covered by ontological KGs (phenotypes in the PGR-crowd by the HPO) and to diversify the type of ontological KGs chosen (the DO for diseases in the BC5CDR Corpus). Therefore, as KGs, we used the ontologies HPO [10], ChEBI [8], and DO [9], linked to the *item* entities when possible.

Fig. 2 illustrates an example of a relation recommendation where the user *EFTUD2* is related to *Microcephaly* and *Mandibulofacial dysostosis* in our RE dataset of reference (Fig. 1). Using the KG we can determine that one of the ancestor connections for *Microcephaly* and *Mandibulofacial dysostosis* is *Abnormality of the skull*. By sharing an ancestor connection,

these two items reinforce the connection between other descendants and the user *EFTUD2*. Thus, we can recommend a relation between our user *EFTUD2* and our item *Cephalocele* (i.e., the green dashed line).

In real-case scenarios, our approach would imply adapting existing and future RE biomedical datasets so that all entities are linked to a KG identifier, which by itself would enhance the applicability of such datasets. Moreover, linking identified entities to KGs is already a widely disseminated NLP task called Named-Entity Linking (NEL) or Concept/Entity Normalization, and most of the time, a natural precedent to RE [17]. As stated previously, a large volume of biomedical ontologies covers most types of studied entities. Therefore, the biggest hurdle would be to guarantee high KG coverage for the original datasets, while our adaptation of the datasets to a recommendation format is highly generalizable. For instance, given a biomedical sentence within a dataset, where the offsets of the entities of interest are identified and linked to KGs (Fig. 1), the next step would be to give a rating to the possible relations considering the whole training dataset and identifying if there are other entities within the KG ancestry line that exhibit similar relations that can further support a true relation. Thus, we expect that the degree of coverage of each KG over each dataset dictates the effectiveness of the approach as well as other factors that we will discuss throughout this manuscript.

After dataset re-formation, we adapted a state-of-the-art recommendation system to recommend relations between biomedical entities considering the specificities of the domain and evaluated the system's added value to a standard state-of-the-art deep learning biomedical RE system.

In this paper, we present the results for relation recommendation on its own, and then the added value of these recommendations to a deep learning biomedical RE system. Our results show that adding KG-based recommendations improves RE systems' ability to identify true relations in high KG coverage settings that baseline deep RE models could not extract from text, indicating that the recommendation model adds value to the RE task. This work resulted in the following main contributions:

- Pipeline for integrating a KG-based recommendation model in a biomedical RE system, including the adaptation of RE datasets for training.
- Biomedical RE deep learning system with added knowledge in the form of KG-based recommendations (K-BiOnt).

The following section will present the related work for RE and recommender systems that specifically target the biomedical domain. We will then proceed to methodology, where we

describe the dataset construction and the different training stages on their own and their joint evaluation. Next, we present results and discuss the effects of the KG-based recommendation on biomedical RE. Finally, we finish with the main conclusions and future work.

## II. RELATED WORK

Biomedical RE is a task in NLP that usually follows Named-Entity Recognition (NER) and Named-Entity Linking (NEL) or Concept/Entity Normalization [18], the identification and mapping of entities in unstructured text, respectively [19]. This information extraction task mostly focuses on relations within the same sentence [20], with approaches that range from co-occurrence to a variety of Machine Learning methods. However, in recent years, deep learning approaches became state-of-the-art for most domains achieving high levels of precision. However, many biomedical relations are still hard to extract even when exploring the full documents [21]. This can be explained by the complexity of the domain that requires a extensive domain knowledge to be correctly perceived. One of those models, inspired by the BO-LSTM and BI-LSTM models [22], [23] is the BiOnt model [24]. The BiOnt model uses deep learning methods along with domain-specific ontologies, achieving state-of-the-art results. Thus, it is a model that already explores external sources of knowledge to perform biomedical RE even if with less KG depth than our proposal. Therefore, in this work, we propose a more in-depth use of KGs with recommendation approaches to detect missed true annotations by BiOnt, which we detail in the methodology section. We chose BiOnt over other biomedical or even non-biomedical [25], [26] RE systems [27], due to its use of similar external entity knowledge. If we can effectively improve the BiOnt model, it shows that their use of knowledge is not sufficient to fully grasp less frequent relations. Nevertheless, we also use the state-of-the-art BioBERT system [27] on the same datasets to provide an extra comparison baseline. The BioBERT system is a BERT-based contextualized word representation model based on a masked language model and pre-trained using bidirectional transformers on large-scale biomedical corpora.

Item recommendation initially focused on similarity-based methods that aimed at extracting features of users and items, computing their similarity, recommending similar users or items to a target user. Similarity-based methods using Neural Network (NN) models effectively extract latent features of users and items for recommendations [28]–[30]. However, they deal with issues such as data sparsity [31], cold-start [32], and lack of explainability [33] (i.e., an user understanding why an item is being recommended). Content-based methods introduce additional information, such as relational data [34] and knowledge graphs [35], [36], and help relieve those issues. Therefore, recently, researchers have focused their attention on generating recommendations using knowledge graphs as additional information [13], [37], such as the TUP system created by Cao *et al.* [38].

Current recommender systems that deal with biomedical data, target different tasks, such as recommending ontologies to annotate biomedical text [39], model biological processes [40],

TABLE I  
STATISTICS OF PGR-CROWD, DDI CORPUS, AND BC5CDR CORPUS REGARDING RELATION EXTRACTION

Dataset	Annotations	Relations		
		True	False	Total
PGR-crowd	4150	5501	626	6127
DDI Corpus	18491	5000	29449	34449
BC5CDR Corpus	32527	1448	2294	3742

recommending drugs to target SARS-CoV-2 regarding the COVID-19 pandemic [41], recommending entities of potential interest to specific researchers [12], or even recommending articles to expand existing biomedical datasets [42]. There is also a focus on recommending articles and venues to researchers to limit their search space [43], [44], for instance, by performing keyword-based recommendation [45]. Further, there is a significant amount of work done on biomedical KG completion [46]–[48], including trying to depend less on domain-specific labeling and going through a minimum supervision route that can scale with the volume of literature available [49].

## III. METHODOLOGY

To demonstrate the benefits of allying KG-based Recommendation to deep Biomedical RE, we used three publicly available datasets describing relations between different types of biomedical entities: the PGR-crowd [14], [50], the DDI Corpus [15], and the BC5CDR Corpus [16]. The first step was to convert these RE datasets into a format compatible with KG-based Recommendation, which required several adjustments and rating assessment described in detail in the following section. Moreover, in this section, we will provide the training and evaluation details for each component system: deep biomedical RE and KG-based Recommendation on their own, and as added features in a deep biomedical RE system plus recommendations. All baseline systems used or adapted throughout our work are openly available through their respective authors, including original configuration details.

### A. Datasets

To take advantage of KG-based Recommendation into RE, we had to create standard <user-item-rating> datasets using the PGR-crowd, DDI Corpus, and BC5CDR Corpus original RE datasets. These datasets describe relations between human phenotypes and genes, using NCBI gene database<sup>1</sup> and HPO identifiers (PGR-crowd), between drugs/chemicals, that can be linked to ChEBI ontology identifiers (DDI Corpus), and interactions between drugs/chemicals and diseases, that can be linked to the ChEBI and DO ontologies (BC5CDR Corpus). The PGR-crowd and DDI datasets are available in the same XML format, and the BC5CDR Corpus is available in a text format, for standard RE applications. Table I presents the RE datasets’ general statistics, including counts for the total number of entity annotations and the distribution of true and false

<sup>1</sup><https://www.ncbi.nlm.nih.gov/gene>

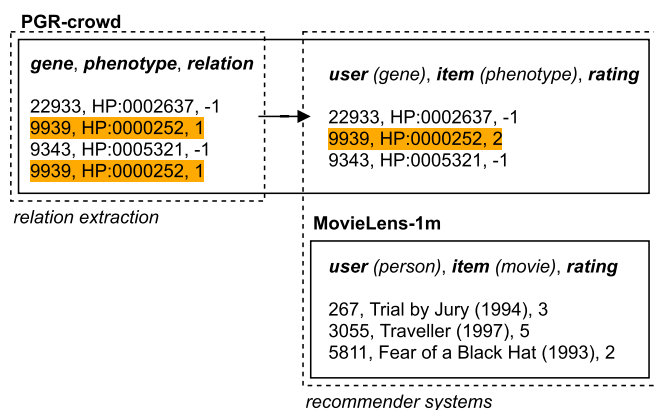


Fig. 3. The conversion from a relation extraction type dataset to a recommender system type dataset (PGR-crowd), including the comparison with examples from the MovieLens-1 m dataset [11].

relations. A relation is considered true if semantically there is an implication of an association between two entities considered in the same sentence, and false if there is no semantic relation or a semantic connection negates the relation between the entities. We did not consider the DDI Corpus relations' different labels classifications for this work, only the binary classification of true/false.

Although the protocols were similar for the three datasets, we had to consider that in the PGR-crowd and BC5CDR datasets, each relation had two distinct entities (genes and human phenotypes, and drugs/chemicals and diseases, respectively). To cast our *item* roles for the PGR dataset, we chose the phenotype entities since these were already mapped to an ontological knowledge graph (HPO). To diversify our approach, for the BC5CDR Corpus, we decided to map disease entities to the DO ontology as *item* entities. While on the DDI Corpus dataset, we were dealing with relations between the same type of entities (i.e., drugs/chemicals) that could both be mapped to a knowledge graph (ChEBI).

In the PGR-crowd dataset, we considered our *users* as genes and our *items* as human phenotypes. Each relation can appear more than once in a RE dataset since different sentences/articles can describe the same relation. We can have multiple instances of the same relation with the same or different labels. Therefore, we attributed 1 to true relations and  $-1$  to false relations and considered the rating the sum of all occurrences of the same relation within the training dataset. This process allowed us to have only one occurrence of each relation as expected in recommender systems type datasets, where the user only rates an item once. Fig. 3 further elucidates the process by direct comparison with an example of the MovieLens-1 m dataset [11]. We followed the same procedure for the BC5CDR Corpus, where we considered our *users* drugs/chemicals and our *items* diseases.

The DDI Corpus was not as straightforward to assign *user* and *item* roles to the drug/chemical entities. As each relation has two entities of the same type, we had to verify the symmetry between relations (e.g., is entity one an effect of entity two and vice-versa, or is it just a one-sided relationship?). For this we

considered the classification done by the creators of the DDI Corpus dataset, where each true relation could be of type *effect* (asymmetric), *mechanism* (asymmetric), *advice* (symmetric), or *int* (symmetric). While the other types are intuitive, the *int* type refers to the default positive interaction for which there is no additional information. So, we disregarded the entities' order for false and symmetric relations and maintained the order assigned for true asymmetric relations when adapting the RE dataset for recommendation. The process of calculating the ratings was identical to the previously described for the PGR-crowd and BC5CDR datasets.

For model training, we converted all entities to an internal identifier. Also, the existing ratings were treated as positive interactions while negative interactions were generated randomly by corrupting items following other models that target implicit feedback [51]. In the work done by Cao *et al.* [38] the negative sampling was done by corrupting items that were less commonly used by users, which could not be applied to datasets with low average ratings.

While previous works [52], [53] mapped publicly available datasets such as MovieLens-1m [11] and DBbook2014<sup>2</sup> to DBpedia [54] entities, whenever a mapping was available, we mapped our datasets to three publicly available biomedical ontologies (HPO for PGR-crowd, ChEBI for DDI Corpus, and DO for BC5CDR Corpus). For the PGR-crowd dataset, since the pre-existing entity identifiers already linked to HPO, our coverage was 100%. However, for the other two datasets (BC5CDR and DDI), the coverage was 26.0% and 32.1%, respectively, which is expected since the creators did not rely on the DO or ChEBI ontology to identify the original entities. The mapping was done automatically by exact matching, allowing for a Levenshtein distance of 1. Thus, particularly in the DDI and BC5CDR Corpora, we did not match possible synonym entities. Doing a more detailed mapping would require either the usage of external normalization tools [18] or domain expertise to review all entities, which would be time and cost-intensive. However, we recognize the limitation in our entity normalization stage, that could be improved in future work. In contrast with previous work [38], we did not preprocess the datasets to filter low-frequency users and items or performed editing on the type of entities or relations in triples, due to our universe being considerably smaller and the reassurance of using domain-specific ontologies instead of the generic domain of DBpedia.

Table II describes the statistics for the three datasets (PGR-crowd, DDI Corpus, and BC5CDR Corpus) regarding the KG-based recommendation format. The data sparsity issue is prevalent in all datasets due to the low number of average ratings.

## B. Training

The deep biomedical RE system BiOnt [24] worked as our baseline since it achieved state-of-the-art performance in the datasets used in this work and also uses knowledge graphs as added information layers. We designed experiments regarding relation recommendation, following the work of Cao *et al.* [38],

<sup>2</sup><http://2014.eswc-conferences.org/important-dates/call-RecSys.html>

TABLE II  
STATISTICS OF PGR-CROWD, DDI CORPUS, AND BC5CDR CORPUS REGARDING KNOWLEDGE GRAPH-BASED RECOMMENDATION

		PGR-crowd	DDI Corpus	BC5CDR Corpus
<b>User-Item Interactions</b>	Users	2176	2538	593
	Items	409	2563	599
	Ratings	4308	25973	2080
	Average Ratings	2	10	4
	Sparsity	99.5%	99.6%	99.4%
<b>Knowledge Graph</b>	Entity	15671	497046	13355
	Relation	1	1	1
<b>Multi-Tasks</b>	Item-Entity Alignments	409	1225	156
	Coverage	100%	47.8%	26.0%

and the incorporation of those recommendations into BiOnt (K-BiOnt). These systems were chosen due to their state-of-the-art results but also their availability and in-depth documentation. For all our experiments, we divided the datasets into a 6:1:3 ratio, corresponding to the training set, the validation set, and the test set, respectively. We used the original datasets as provided for the Deep Learning component, making the appropriate parsing for the system specifications. While for the Recommendation component, we used the re-formatted datasets, as described in the previous sub-section.

1) *Baseline Deep Learning Model*: The BiOnt model uses ontologies as external sources of knowledge to add information layers to a baseline deep learning model, following the work of Lamurias *et al.* [22]. An ontology is a formal definition of concepts related to a specific subject. It can be represented by a tuple  $\langle C, R \rangle$ , where  $C$  represents the set of concepts in an ontology and  $R$  the set of relations between the same ontology concepts. Similar to our dataset construction, the type of ontology relations considered by Sousa and Couto [24] is subsumption relations, *is-a* due to its transitive aspect. For instance, with  $(c_1, c_2) \in R$ , and  $(c_2, c_3) \in R$ , the authors assume that  $(c_1, c_3)$  is a valid relation within the ontology. The ancestors of each concept  $c$  are given by:

$$Anc(c) = a : (c, a) \in T \quad (1)$$

where  $T$  is the transitive closure of  $R$ . The authors define the common ancestors between the concepts  $c_1$  and  $c_2$  as:

$$CommRA(c_1, c_2) = Anc(c_1) \cap Anc(c_2) \quad (2)$$

A relation between different ontology concepts can be represented by  $(x_1, y_1)$ , where  $x_1 \in X$  and  $X$  represents the set of concepts in the first ontology, and  $y_1 \in Y$  and  $Y$  represents the set of concepts in the second ontology. For instance, with  $(x_2, y_2) \in RA$ , where  $RA$  is the set of relations between ancestors, and  $(x_2 \text{ is-a } x_1)$ , and  $(y_2 \text{ is-a } y_1)$ , their model assumes that  $(x_1, y_1)$  is a valid relation. The concatenation of the relations between the ancestors of concepts  $x_2$  and  $y_2$  is defined using:

$$ConcRA(x_2, y_2) = Anc(x_2) \bowtie Anc(y_2) \quad (3)$$

Since the common ancestors' channel could only be used for relations between the same type of biomedical entities (i.e., DDI Corpus), we only use the concatenation of ancestors channel for the relations between different biomedical entities (i.e., PGR-crowd).

Each ontology concept corresponds to one-hot vector  $v_c$ , a vector of zeros except for the position corresponding to the concepts' ID. An embedding matrix  $M \in \mathbb{R}^{D \times C}$  transforms these sparse vectors into dense vectors, where  $D$  is the dimensionality of the embedding layer and  $C$  is the number of concepts of the ontologies. Then, the output of the embedding layer is given by:

$$f(c) = M \cdot v_c \quad (4)$$

Next, the ontology embedding layer, with a dimensionality of 50 (as suggested by [22]), initializes its values randomly to be later tuned through back-propagation. Then, the vectors' sequence representing the relations between the ancestors of the terms is fed into a Long short-term memory (LSTM) layer, ordered from the more general concepts to the terms themselves. Finally, the system uses a max pool layer fed into a dense layer through a sigmoid activation function, and a softmax layer outputs the probability for each class.

The model was trained using a stochastic gradient descent optimization algorithm where weights were updated using the back-propagation of error algorithm. At each iteration, the model with a given set of weights creates predictions and computes the error for those predictions. The optimization algorithm seeks to alter the weights to reduce that error in the next evaluation. The relevant hyperparameters of this model tuned for our experiments were mini-batch gradient descent optimization algorithm (RMSprop), learning rate (0.0001), loss function (categorical cross-entropy), and dropout rate (0.500) for every layer except the penultimate and output layers.

We used the three standard evaluation metrics for RE models: *Precision*: Expresses how often the results are correct; *Recall*: It is the number of correct results identified; *F1 score*: Expresses overall performance by the harmonic mean of precision and recall.

2) *Item Recommendation*: The TUP model created by Cao *et al.* [38] takes a list of user-item pairs  $\mathcal{Y} = (u, i)$  as input, and outputs a relevance score  $g(u, i; p)$ , indicating the likelihood that  $u$  likes  $i$ , given the preference  $p \in \mathcal{P}$ . In this work, instead of the terminology being *the likelihood that  $u$  likes  $i$*  is the likelihood that  $u$  as a biomedical entity is related to  $i$  as another biomedical entity. For each user-item pair, the TUP model induces a preference. The authors designed two strategies for preference induction: a hard approach that selects one out of the  $P$  preferences and a soft way that combines all preferences with attentions. The soft strategy yielded a better performance

for both authors' datasets. Therefore, we opted for this strategy to create our models. The preferences constitute the motives for each an *user* entity may be connected to an *item* entity. For traditional recommendation set-ups, these usually lack depth in explainability. To avoid that, in this work, we provided them explicit semantics by aligning them with the KG relations, capturing the intuition that the types of item attributes play a crucial role in user assignment. Considering that an entity might be related to another entity according to various factors, which can not be restricted by a firm boundary, instead of selecting the most prominent preference, the soft strategy combines multiple preferences via an attention mechanism:

$$p = \sum_{p' \in \mathcal{P}} \alpha_{p'} p' \quad (5)$$

where  $\alpha_{p'}$  is the attention weight of preference  $p'$ , and defined as proportional to the similarity score:

$$\alpha_{p'} \propto \phi(u, o, p') \quad (6)$$

To deal with the issue where one entity (i.e., *user*) might be associated with multiple entities (i.e., *items*), and also, several entities (i.e., *users*) may be associated with a single entity (i.e., *item*) (1-to-N and N-to-1 issues), the authors introduce preference hyperplanes assigning each preference with two vectors (inspired by TransH [55]):  $w_p$  for the projection to a hyperplane,  $p$  for the translation between users and items. The authors define the hyperplane-based translation function as follows:

$$g(u, i; p) = \|u^\perp + p - i^\perp\| \quad (7)$$

$$u^\perp = u - w_p^T u w_p, \quad i^\perp = i - w_p^T i w_p \quad (8)$$

where  $u^\perp$  and  $i^\perp$  are projected vectors of the user and the item, and are obtained through the induced preference  $p$ .  $w_p$  is the projection vector that is obtained along with the induction process of preferences  $p$  through attentive addition of all projection vectors based on the induced attention weights in the soft strategy:

$$w_p = \sum_{p' \in \mathcal{P}} \alpha_{p'} p' w_{p'} \quad (9)$$

Then, the authors encourage the translation distances of the interacted items to be smaller than random ones for each user through BPR Loss function:

$$\mathcal{L}_p = \sum_{(u, i) \in \mathcal{Y}} \sum_{(u, i') \in \mathcal{Y}'} -\log \sigma[g(u, i'; p') - g(u, i; p)] \quad (10)$$

where  $\mathcal{Y}'$  contains negative interactions by randomly corrupting an interacted item to a non-interacted one for each user.

The relevant hyperparameters for TUP tuned for our experiments were learning rate (0.005),  $L_2$  coefficient ( $10^{-5}$ ), optimization method (Adagrad), batch size (512), embedding size (64), and the number of preferences (1). The number of preferences corresponds to the number of different relations within the KGs attributed to each dataset (Table II). Since this model is inspired by the TransH model described previously, some functionalities could not be explored due to the lack of relation diversity.

The TUP model uses as evaluation metrics the following:

- *Precision@10*: The fraction of the of *item* entities recommended that are relevant for establishing relations with our *user* entities, within the top 10 recommendations. The final precision corresponds to the mean of all *user* entities.
- *Recall@10*: The proportion of *item* entities relevant to the *user* entities that have been successfully recommended, within the top 10 recommendations. The final recall corresponds to the mean of all *user* entities.
- *F1 score@10*: The harmonic mean of precision at rank 10 and recall at rank 10.
- *Hit ratio@10*: It is 1 if the gold *item* entities are recommended within the top 10 *item* entities, otherwise it is 0. The final hit ration score corresponds to the mean of all *user* entities.
- *nDCG@10*: The Normalized Discounted Cumulative Gain (nDCG) is a standard measure of ranking quality that considers the graded relevance among positive and negative *item* entities within the top 10 of the ranked list.

3) *K-BiOnt*: The general approach pipeline that joins BiOnt to a TUP adaptation for RE is presented in Fig. 4.

First, we proceed with the standard training process of the BiOnt model, which can be divided into three main stages after sentence tokenization: WordNet classes [56], word embeddings, and ontology embeddings. The ontology embedding stage represents the relations between the ancestors for each ontology concept corresponding to an entity. For instance, for the PGR-crowd dataset, the system links the entities to the HPO and GO biomedical ontologies, the DDI Corpus to the ChEBI ontology, and the BC5CDR Corpus to the ChEBI and DO ontologies with different coverage degrees, as discussed previously. The system uses a max pool layer fed into a dense layer through a sigmoid activation function, and a softmax layer outputs the probability for each class. The BiOnt model adds external entity knowledge through two channels of common ancestry and concatenation of ancestors. These knowledge channels aim to answer the questions: i) Do the entities in question share ancestors? (only applicable to relations between the same type of entities); and ii) Do the entities in question have ancestors that have established relations? The BiOnt model uses the answer to these two questions to support or discard a relation. However, our K-BiOnt knowledge layer goes deeper into the inferences that can support or discard a relation by answering: Do we have entities outside of the ones considered in the relation that we know that establish relations with one of the entities ancestors? If yes, how many? In what capacity (true/negative)? And in what degree (e.g., 2, 3, or -5)?

Thus, going back to the example in Fig. 4, our goal is to support or discard a relation between a gene *HDAC4* and a human phenotype, *Mandibular prognathism*. Yet, since their ancestors do not have known relations (excluding the BiOnt concatenation of ancestors' channel) and the entities are from different types (excluding the usage of the common ancestors' channel), the BiOnt knowledge layer does not provide information to support or discard this relation. However, since we know from the training set that the gene, *HDAC4*, shares a true relation with both the *Jaw ankylosis* and the *Trimus* phenotype, and both of these entities have the same ancestor as the *Mandibular prognathism*

input sentence - PMID:28570402

**HDAC4** expression is suspected to be involved in the epigenetic regulations behind the **mandibular prognathism** phenotype.

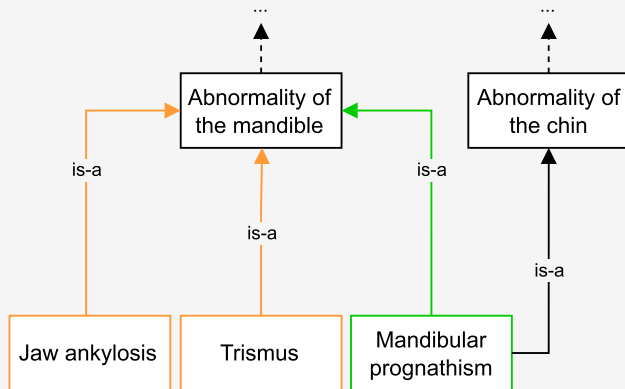
baseline BiOnt

Gene Ontology

Human Phenotype Ontology

**HDAC4**<sub>noun.e1</sub> expression<sub>noun</sub> suspected<sub>verb</sub> involved<sub>verb</sub> in<sub>adverb</sub> regulations<sub>noun</sub> behind<sub>adverb</sub> **mandibular prognathism**<sub>noun.e2</sub>

K-BiOnt knowledge layer



Human Phenotype Ontology

Known interactions from relations present in the training dataset have two knowledge graph interactions with the same ancestor:

**HDAC4, jaw ankylosis, true**

**HDAC4, trismus, true**

**supports  
true relation**

Fig. 4. General approach pipeline regarding the joint approaches BiOnt and the knowledge layer based on recommendation. We simplified the input sentence to facilitate reading comprehension (PMID:28570402).

phenotype, we can support a true relation between *HDAC4* and *Mandibular prognathism*. Our knowledge layer considers translational relationships between *users* and *items*. Which in our example means that for each human phenotype (*item*) related to a gene (*user*), we consider the whole subsequent ancestry of the phenotype until root to provide information on our gene.

### C. Joint Evaluation

To evaluate our approach, we created a confusion matrix table to compare the BiOnt model's output and the adjusted TUP model results against our gold standard test sets. This table served as a way for us to detect the main contributions of adding KG-based recommendation to a deep RE model (K-BiOnt) to all datasets.

One caveat is that an extracted relation between two entities is specific to the text where they are mentioned; however, in KG-based recommendation, the relation tag is specific to the entities it refers. To overcome this issue, we primarily considered the label generated through the application of the baseline system to the normal text-bound RE dataset. We only changed the output if the modules (Deep Learning and Recommendation) disagreed with the label. Upon disagreement, we only altered

the label if the Recommendation module assigned true and the baseline system false. Thus, only considering the Recommendation module input to capture an undetected connection. Table III presents an example of a confusion matrix table for five distinct scenarios using relations from the PGR-crowd dataset. All true relations captured by KG-based recommendation were attributed with the final judgment of true independently of the Top@N. We only considered false relations in the final judgment if all model components agreed on the label false. Therefore, the BiOnt module component was preferred for the attributed label since it is based on the linguistic context of the relationship. The Recommendation module, based exclusively on knowledge regarding the target entities, was only considered for potential true labels that we hypothesis could not be retrieved solely on the linguistic context or were less frequent in the training data.

## IV. RESULTS AND DISCUSSION

This section presents our assessment of the benefits of using KG-based recommendation as an added resource for RE systems in the biomedical domain. As baseline, we compared the results of the baselines deep learning models BiOnt [24] and BioBERT [27] for RE with the adjusted TUP model [38] for

TABLE III

AN EXAMPLE OF A CONFUSION MATRIX TABLE FOR FIVE DISTINCT SCENARIOS USING RELATIONS FROM THE PGR-CROWD DATASET

Relation (user-item)	BiOnt	Top@3	Top@5	Top@10	Final Judgement	Gold Standard
WNT7B-cancer	True	True	True	True	True	True
Adcy1-seizures	False	False	False	True	True	True
TRAPPC2-developmental delay	True	False	False	False	True	True
IARS2-neuropathy	False	False	False	False	False	True
SLC12A6-autosomal recessive	False	True	True	True	True	False

TABLE IV

OVERALL RESULTS REGARDING THE BIOnt AND BIOBERT SYSTEMS. P STANDS FOR PRECISION AND R FOR RECALL

Dataset	System	P (%)	R (%)	F1 (%)
PGR-crowd	BioBERT	80.04	99.22	88.60
	BiOnt	81.30	97.55	88.69
DDI Corpus	BioBERT	86.61	87.23	86.92
	BiOnt	61.85	73.24	67.06
BC5CDR Corpus	BioBERT	59.78	72.67	65.60
	BiOnt	63.96	71.73	67.62

item recommendation, and the integration of BiOnt with TUP (K-BiOnt).

### A. Deep Learning Model

Table IV presents the results for the application of our three datasets to the deep learning models BiOnt and BioBERT.

For BiOnt, the results were slightly different from the performances reported on the original work for the DDI Corpus [24], but almost identical for the PGR-crowd dataset [14]. We can justify the DDI Corpus performance differences ( $\approx 4\%$  in F1) with our use of the updated ChEBI version that has fewer alignments with the entities in the original dataset. The PGR-crowd as a significant imbalance of true/false relations with the majority of relations being true and the DDI and BC5CDR Corpora share the same imbalance but in favour of the false relations (Table I), which can affect the performance of these datasets differently, despite the BiOnt model ability to assign class weights.

As for the BioBERT system, the PGR-crowd dataset and the BC5CDR Corpus results were very similar to the BiOnt's model performance. However, given the class imbalances of all three datasets, it should be possible to alter class weights. Still, BioBERT's loss function does not allow this flexibility, possibly undermining their results. The BioBERT system significantly outperformed the BiOnt model for the DDI Corpus, despite the class imbalances.

### B. Knowledge Graph-Based Recommendation

Table V presents the results for the adapted TUP model using the soft item recommendation strategy mentioned in Section III-B3. TUP authors [38] state that the peak performance for their model is when the average number of ratings for user ranges from 100 to 200. This range is far from our average number of ratings for both datasets (2 for PGR-crowd, 10 for DDI Corpus, and 4

for the BC5CDR Corpus). Thus, we believe that more training data allied with less sparsity would enhance our results further. Also, the higher overall results for the PGR-crowd demonstrate the importance of item-entity alignments since all items (i.e., human phenotypes) were linked to the HPO [10]. In contrast, on the DDI and BC5CDR Corpora, only 47.8% and 26.0% of the items could be linked to the ChEBI [15] and DO [9] ontologies, respectively, leading to a drop in performance.

### C. Joint Evaluation

Table VI presents the final results by adding the adjusted TUP model recommendation to the BiOnt model (K-BiOnt), considering top@3, top@5, and top@10 recommendations. These results are a reflection of the results of the confusion matrix tables created as described in the example of Table III.

For the PGR-crowd dataset, the average number of ratings per *user* entity is 2. A false relation usually appears only once being rated with  $-1$  and not a lower number which is not sufficient to indicate to the model that the *user* entity is unrelated to an *item* entity. An approach that we could study in the future is creating negative sampling using false relations, not the traditional random sampling for negative observations associated with implicit feedback. The added performance of TUP over the BiOnt model (K-BiOnt) holds for all Top@N. However, the number of false positive relations increases with the subsequent decrease in performance for Top@5 and Top@10. Although, after closer inspection to the added false positives, for the majority of them, the *item* entity human phenotype is under the *Mode of inheritance* category of HPO, not under *Phenotypic abnormality*. This last branch is the most developed branch within the HPO and of more interest to researchers. Likewise, the BC5CDR Corpus also increased performance compared to the BiOnt baseline for Top@5 and Top@10 despite the low TUP performance, which indicates potential for a more impactful approach following improvement in linking the entities to KG identifiers.

For the DDI Corpus, the results are identical across all Top@N (the BioBERT baseline) since we could not capture a true positive through item recommendation within the first ten recommendations.

### D. Ablation Study

To study the impact of knowledge graph coverage, we chose the dataset with the lowest value in coverage (BC5CDR Corpus).



TABLE V  
OVERALL RESULTS REGARDING THE TUP MODEL FOR ITEM RECOMMENDATION

Dataset	Precision@10 (%)	Recall@10 (%)	F1@10 (%)	Hit ratio@10 (%)	nDCG@10 (%)
PGR-crowd	3.72	32.10	6.59	35.96	19.33
DDI Corpus	1.90	2.40	1.90	18.70	8.20
BC5CDR Corpus	0.53	2.24	0.79	5.28	2.37

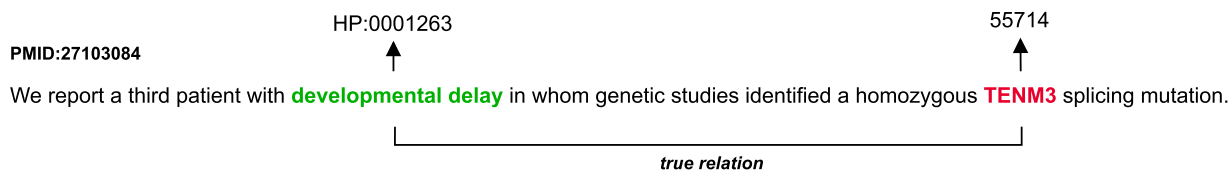


Fig. 5. An example of a true relation detected by the recommendation module from the PGR-crowd dataset from article PMID:27103084. The phenotype entity is linked to the HPO ontology (HP:0001263), and the gene entity is linked to the NCBI gene database (55714). We simplified the sentence to facilitate reading comprehension.

TABLE VI  
FINAL RESULTS BY ADDING THE ADJUSTED TUP MODEL RECOMMENDATIONS TO THE BiOnt MODEL (K-BiOnt), CONSIDERING TOP@3, TOP@5, AND TOP@10 RECOMMENDATIONS. P STANDS FOR PRECISION AND R FOR RECALL

Dataset	Configuration	P (%)	R (%)	F1 (%)
PGR-crowd	BiOnt	81.30	97.55	88.69
	+ Top@3	81.23	97.90	<b>88.79</b>
	+ Top@5	80.98	<b>98.17</b>	88.75
	+ Top@10	80.98	<b>98.17</b>	88.75
DDI Corpus	BiOnt	<b>61.85</b>	<b>73.24</b>	<b>67.06</b>
	+ Top@3	61.85	73.24	67.06
	+ Top@5	61.85	73.24	67.06
	+ Top@10	61.85	73.24	67.06
BC5CDR Corpus	BiOnt	63.96	71.73	67.62
	+ Top@3	63.96	71.73	67.62
	+ Top@5	64.15	72.34	68.00
	+ Top@10	<b>64.45</b>	<b>73.25</b>	<b>68.56</b>

We created the recommendation module only taking into account the 156 items linked to DO ontological concepts. Table VII presents the results for this study.

By Table VII, we can verify that the presence of only ontological covered items, even if in a small number, is enough to impact the performance of item recommendation, almost doubling our previous results. Even if there was no significant impact on the K-BiOnt model due to the small number of items, we know that augmenting the covered items through a more robust concept normalization step can improve the K-BiOnt performance.

### E. Impact on RE

We decided to perform error analysis on the performance of the PGR-crowd dataset, comparing the baselines BioBERT and BiOnt to our approach K-BiOnt to measure their actual impact on the RE task. We found that more true relations were identified by considering relations that the KG-based recommendation model recommended.

In the PGR-crowd dataset, the *item* entities (i.e., human phenotypes) are all linked to the HPO, with subsequent complete coverage of the KG entities over the *item* entities. The full coverage translated a higher contribution of the recommendation module to the RE task. Fig. 5 illustrates one of those true relations detected by the recommendation module and missed by the BiOnt model. Note that our models only added true relations recommended with the adjusted TUP at Top@3. All other experiments also recommended false positives, undermining the recommendation module benefits.

These results show the advantage of adding recommendations to RE, mainly to populate knowledge bases of gold standard relations, where the goal is not only to identify the relation that is explicitly mentioned in the text but to find every true relation that we can derive from it. The success of the recommendation module is explained by the exploration of KGs that allows the RE process to consider the connections between the associated KG.

However, existing KGs are far from complete, limiting the knowledge we can transfer into RE systems. Considering this limitation, Cao *et al.* [38] aligned item recommendation (TUP) with KG completion (TransH). KG completion is a field in accelerated popularity given its relevance for question answering tasks [57], [58], but also to search entities and their relations in text [59]. This field should be considered for future exploration of added knowledge to biomedical RE, to further enhance the recommendation of less frequent relations.

## V. CONCLUSION AND FUTURE WORK

This paper proposed a new recommendation-based complementary approach to deep learning biomedical RE that considers biomedical ontologies as additional sources of information. The KG-based recommendation pipeline presented in this work takes advantage of *user* entity-*item* entity interactions as well as knowledge graphs that can be linked to *item* entities. In our case study, the biomedical KGs were HPO, ChEBI, and

**TABLE VII**  
ABLATION STUDY RESULTS REGARDING THE TUP MODEL FOR ITEM RECOMMENDATION FOR THE FULL DATASET AND THE KG COVERED SUBSET OF THE BC5CDR CORPUS

Dataset	Precision@10 (%)	Recall@10 (%)	F1@10 (%)	Hit ratio@10 (%)	nDCG@10 (%)
BC5CDR Corpus	0.53	2.24	0.79	5.28	2.37
Subset BC5CDR Corpus	0.79	4.87	1.32	7.92	5.62

DO. We performed experiments using both item recommendation algorithm on its own and as an added module to a deep biomedical RE system. We present the benefits of using both methods simultaneously and the RE task's added value. Our results show that KG-based recommendation can be a valuable asset to biomedical RE by detecting previously undiscovered true relations between biomedical entities. However, the low coverage of the associated KGs damages performance.

Additionally, we produced three recommendation datasets in the format <user-item-rating> for human phenotype and gene relations, drugs/chemicals interactions, and drugs/chemicals and diseases relations attributing a rating for each user-item pair. We also presented a comprehensive pipeline for creating a biomedical RE system using KG-based recommendation (K-BiOnt). Ultimately, we demonstrated that adding recommendations can increase deep biomedical RE models' performance by considering external sources of knowledge when they have sufficient coverage of the domain.

Biomedical RE datasets usually do not describe more than one type of relation. However, upon the availability of a dataset describing more types of labeled relations, a multi-graph approach could be employed linking each *item* entity to their respective ontological identifier. Even though we could argue that our representation of the ratings between user-item pairs is not representative of real-world, it is a cross-approach problem. Current deep learning approaches to biomedical RE also take labeled data to create models where the distribution is not a representation of real-world data and where a lot of less frequent associations are missed. In the future, the approach could be expanded by considering other types of relations between biomedical entities and by applying it to different types of baseline systems (i.e., BioBERT). Another angle to be explored could be adding more biomedical ontologies, including possible interconnections between multiple ontologies, that could expand our KGs even further by increasing the number of preferences. Also, upon availability within biomedical ontologies, another complementary route could be adding informative axioms such as *disjointness* and studying the effect of ontological depth.

## REFERENCES

- [1] N. Fiorini *et al.*, "Best match: New relevance search for pubmed," *PLoS Biol.*, vol. 16, no. 8, pp. 1–12, 2018.
- [2] N. Indurkha and F. J. Damerou, *Handbook of Natural Language Processing*, vol. 2. Cleveland, Ohio, EUA: CRC Press, 2010.
- [3] S. Hasan, D. Rivera, X.-C. Wu, E. B. Durbin, J. B. Christian, and G. Tourassi, "Knowledge graph-enabled cancer data analytics," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 1952–1967, Jul. 2020.
- [4] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [Review Article]," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.
- [5] A. Holzinger, B. Malle, A. Saranti, and B. Pfeifer, "Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI," *Inf. Fusion*, vol. 71, pp. 28–37, 2021.
- [6] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowl. Acquisition*, vol. 5, no. 2, pp. 199–220, 1993.
- [7] G. O. Consortium, "The gene ontology resource: 20 years and still going strong," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D330–D338, 2019.
- [8] J. Hastings *et al.*, "The ChEBI reference database and ontology for biologically relevant chemistry: Enhancements for 2013," *Nucleic Acids Res.*, vol. 41, no. D1, pp. D456–D463, 2012.
- [9] L. M. Schriml *et al.*, "Disease ontology: A backbone for disease semantic integration," *Nucleic Acids Res.*, vol. 40, no. D1, pp. D940–D946, 2012.
- [10] S. Köhler *et al.*, "Expansion of the human phenotype ontology (HPO) knowledge base and resources," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1018–D1027, 2019.
- [11] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *ACM Trans. Interact. Syst.*, vol. 5, no. 4, 2016, Art. no. 19.
- [12] M. Barros, A. Moitinho, and F. M. Couto, "Using research literature to generate datasets of implicit feedback for recommending scientific items," *IEEE Access*, vol. 7, pp. 176668–176680, 2019.
- [13] Q. Guo *et al.*, "A survey on knowledge graph-based recommender systems," in *IEEE Trans. Knowl. Data Eng.*, 2020.
- [14] D. Sousa, A. Lamurias, and F. M. Couto, "A hybrid approach toward biomedical relation extraction training corpora: Combining distant supervision with crowdsourcing," *Database*, vol. 2020, 2020, Art. no. baaa104.
- [15] M. Herrero-Zazo, I. Segura-Bedmar, P. Martínez, and T. Declerck, "The DDI corpus: An annotated corpus with pharmacological substances and drug-drug interactions," *J. Biomed. Informat.*, vol. 46, no. 5, pp. 914–920, 2013.
- [16] J. Li *et al.*, "BioCreative V. CDR task corpus: A resource for chemical disease relation extraction," *Database*, vol. 2016, May 2016, Art. no. baw068.
- [17] M.-S. Huang *et al.*, "Biomedical named entity recognition and linking datasets: Survey and our recent development," *Brief. Bioinf.*, vol. 21, no. 6, pp. 2219–2238, 2020.
- [18] C. Funk *et al.*, "Large-scale biomedical concept recognition: An evaluation of current automatic annotators and their parameters," *BMC Bioinf.*, vol. 15, no. 1, pp. 1–29, 2014.
- [19] D. Sousa, A. Lamurias, and F. M. Couto, "Using neural networks for relation extraction from biomedical literature," in *Artificial Neural Networks*, New York, NY, USA: Springer, 2021, pp. 289–305.
- [20] L. Luo, Z. Yang, M. Cao, L. Wang, Y. Zhang, and H. Lin, "A neural network-based joint learning approach for biomedical entity and relation extraction from biomedical literature," *J. Biomed. Informat.*, vol. 103, 2020, Art. no. 103384.
- [21] P. Verga, E. Strubell, and A. McCallum, "Simultaneously self-attending to all mentions for full-abstract biological relation extraction," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2018, vol. 1, pp. 872–884.
- [22] A. Lamurias, D. Sousa, L. A. Clarke, and F. M. Couto, "BO-LSTM: Classifying relations via long short-term memory networks along biomedical ontologies," *BMC Bioinf.*, vol. 20, no. 1, pp. 1–12, 2019.
- [23] B. Xu, X. Shi, Z. Zhao, and W. Zheng, "Leveraging biomedical resources in Bi-LSTM for drug-drug interaction extraction," *IEEE Access*, vol. 6, pp. 33432–33439, 2018.
- [24] D. Sousa and F. M. Couto, "BiOnt: Deep learning using multiple biomedical ontologies for relation extraction," in *Advances in Information Retrieval*, J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, and F. Martins, Eds., Cham, Switzerland: Springer, 2020, pp. 367–374.
- [25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2019, vol. 1, pp. 4171–4186.

- [26] Y. Gu *et al.*, “Domain-specific language model pretraining for biomedical natural language processing,” in *ACM Trans. Comput. Healthcare (HEALTH)*, vol. 3, no. 1, pp. 1–23, 2021.
- [27] J. Lee *et al.*, “Biobert: A pre-trained biomedical language representation model for biomedical text mining,” *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.
- [28] P. Covington, J. Adams, and E. Sargin, “Deep neural networks for youtube recommendations,” in *Proc. 10th ACM Conf. Recommender Syst.*, New York, NY, USA, 2016, pp. 191–198.
- [29] D. Jannach and M. Ludewig, “When recurrent neural networks meet the neighborhood for session-based recommendation,” in *Proc. 11th ACM Conf. Recommender Syst.*, New York, NY, USA, 2017, pp. 306–310.
- [30] Z. Huang, X. Xu, H. Zhu, and M. Zhou, “An efficient group recommendation model with multiattention-based neural networks,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4461–4474, Nov. 2020.
- [31] G. Guo, H. Qiu, Z. Tan, Y. Liu, J. Ma, and X. Wang, “Resolving data sparsity by multi-type auxiliary implicit feedback for recommender systems,” *Knowl.-Based Syst.*, vol. 138, pp. 202–207, 2017.
- [32] Y. Bi, L. Song, M. Yao, Z. Wu, J. Wang, and J. Xiao, “A heterogeneous information network based cross domain insurance recommendation system for cold start users,” in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, New York, NY, USA, 2020, pp. 2211–2220.
- [33] K. Balog and F. Radlinski, “Measuring recommendation explanation quality: The conflicting goals of explanations,” in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, New York, NY, USA, 2020, pp. 329–338.
- [34] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, and T.-S. Chua, “Temporal relational ranking for stock prediction,” *ACM Trans. Inf. Syst.*, vol. 37, no. 2, 2019, Art. no. 27.
- [35] Q. Ai, V. Azizi, X. Chen, and Y. Zhang, “Learning heterogeneous knowledge base embeddings for explainable recommendation,” *Algorithms*, vol. 11, no. 9, 2018, Art. no. 137.
- [36] H. Wang, F. Zhang, M. Zhao, W. Li, X. Xie, and M. Guo, “Multi-task feature learning for knowledge graph enhanced recommendation,” in *Proc. World Wide Web Conf.*, New York, NY, USA, 2019, pp. 2000–2010.
- [37] Y. Feng, B. Hu, F. Lv, Q. Liu, Z. Zhang, and W. Ou, “ATBRG: Adaptive target-behavior relational graph network for effective recommendation,” in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, New York, NY, USA, 2020, pp. 2231–2240.
- [38] Y. Cao, X. Wang, X. He, Z. Hu, and T.-S. Chua, “Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences,” in *Proc. World Wide Web Conf.*, New York, NY, USA, 2019, pp. 151–161.
- [39] M. Martínez-Romero, C. Jonquet, M. J. O’connor, J. Graybeal, A. Pazos, and M. A. Musen, “Ncbo ontology recommender 2.0: An enhanced approach for biomedical ontology recommendation,” *J. Biomed. Semantics*, vol. 8, no. 1, pp. 1–22, 2017.
- [40] J. Wang, D. Tan, B. Cao, J. Fan, and S. Deep, “Independent path-based process recommendation algorithm for improving biomedical process modelling,” *Electron. Lett.*, vol. 56, no. 11, pp. 531–533, 2020.
- [41] L. E. Gates and A. A. Hamed, “The anatomy of the SARS-COV-2 biomedical literature: Introducing the covidx network algorithm for drug repurposing recommendation,” *J. Med. Internet Res.*, vol. 22, no. 8, 2020, Art. no. e21169.
- [42] B. G. Patra *et al.*, “A content-based literature recommendation system for datasets to improve data reusability - A case study on gene expression omnibus (GEO) datasets,” *J. Biomed. Informat.*, vol. 104, 2020, Art. no. 103399.
- [43] J. Guerra, W. Quan, K. Li, L. Ahumada, F. Winston, and B. Desai, “SCOSY: A biomedical collaboration recommendation system,” in *Proc. 40th IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, vol. 2018, pp. 3987–3990, 2018.
- [44] X. Feng *et al.*, “The deep learning-based recommender system “pub-mender” for choosing a biomedical publication venue: Development and validation study,” *J. Med. Internet Res.*, vol. 21, no. 5, 2019, Art. no. e12957.
- [45] F. Yang, J. Zhu, J. Lun, Z. Zheng, Y. Tang, and J. Wu, “A keyword-based scholar recommendation framework for biomedical literature,” in *Proc. IEEE 22nd Int. Conf. Comput. Supported Cooperative Work Des.*, 2018, pp. 247–252.
- [46] X. Zhang and C. Che, “Drug repurposing for Parkinson’s disease by integrating knowledge graph completion model and knowledge fusion of medical literature,” *Future Internet*, vol. 13, no. 1, 2021, Art. no. 14.
- [47] H. Fei, Y. Ren, Y. Zhang, D. Ji, and X. Liang, “Enriching contextualized language model from knowledge graph for biomedical information extraction,” *Brief. Bioinf.*, vol. 22, no. 3, 2021, Art. no. bbaa110.
- [48] S. Biswas, P. Mitra, and K. S. Rao, “Relation prediction of co-morbid diseases using knowledge graph completion,” *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 2, pp. 708–717, Mar./Apr. 2021.
- [49] J. Yuan *et al.*, “Constructing biomedical domain-specific knowledge graph with minimum supervision,” *Knowl. Inf. Syst.*, vol. 62, no. 1, pp. 317–336, 2020.
- [50] D. Sousa, A. Lamurias, and F. M. Couto, “A silver standard corpus of human phenotype-gene relations,” in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2019, pp. 1487–1492.
- [51] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T.-S. Chua, “Explainable reasoning over knowledge graphs for recommendation,” *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, pp. 5329–5336, 2019.
- [52] T. D. Noia, V. C. Ostuni, P. Tomeo, and E. D. Sciascio, “Sprank: Semantic path-based ranking for top-N recommendations using linked open data,” *ACM Trans. Intell. Syst. Technol.*, vol. 8, Sep. 2016, Art. no. 9.
- [53] R. Bhargava, A. Kumar, and S. Gupta, “Collaborative methodologies for pattern evaluation for web personalization using semantic web mining,” in *Proc. Int. Conf. Smart Syst. Inventive Technol.*, 2019, pp. 852–855.
- [54] J. Lehmann *et al.*, “Dbpedia-A large-scale, multilingual knowledge base extracted from wikipedia,” *Semantic Web*, vol. 6, no. 2, pp. 167–195, 2015.
- [55] Z. Wang, J. Zhang, J. Feng, and Z. Chen, “Knowledge graph embedding by translating on hyperplanes,” in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 1112–1119.
- [56] M. Ciaramita and Y. Altun, “Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger,” in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2006, pp. 594–602.
- [57] W. Zheng, H. Cheng, L. Zou, J. X. Yu, and K. Zhao, “Natural language question/answering: Let users talk with the knowledge graph,” in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 217–226.
- [58] X. Huang, J. Zhang, D. Li, and P. Li, “Knowledge graph embedding based question answering,” in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, 2019, pp. 105–113.
- [59] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, “A survey on knowledge graphs: Representation, acquisition and applications,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 494–514, Feb. 2022.