

Demographic Aware Probabilistic Medical Knowledge Graph Embeddings of Electronic Medical Records

Aynur Guluzade, Endri Kacupaj, and Maria Maleshkova

University of Bonn, Bonn, Germany
s6aygulu@uni-bonn.de
{kacupaj,maleshkova}@cs.uni-bonn.de

Abstract. Medical knowledge graphs (KGs) constructed from Electronic Medical Records (EMR) contain abundant information about patients and medical entities. The utilization of KG embedding models on these data has proven to be efficient for different medical tasks. However, existing models do not properly incorporate patient demographics and most of them ignore the probabilistic features of the medical KG. In this paper, we propose DARLING (**D**emographic **A**ware **p**RobabiListic **m**edical **k**nowledge **e**mbeddin**G**), a demographic-aware medical KG embedding framework that explicitly incorporates demographics in the medical entities space by associating patient demographics with a corresponding hyperplane. Our framework leverages the probabilistic features within the medical entities for learning their representations through demographic guidance. We evaluate DARLING through link prediction for treatments and medicines, on a medical KG constructed from EMR data, and illustrate its superior performance compared to existing KG embedding models.

Keywords: Demographics, Probabilistic Medical Knowledge Graph, Knowledge Graph Embedding, Electronic Medical Records

1 Introduction

In recent years, knowledge graphs (KGs) have been established for medical assistance as the underlying core component of clinical decision support systems (CDSSs) [23,17] and self-diagnostic symptom checkers [15]. Those KGs are often extracted from sources such as Electronic Medical Records (EMR) and store clinical information into a set of triples for representing medical entities as nodes and relations as the edges between them. Medical KG-based applications have been reported in different scenarios, such as treatment recommendations [3], medicine recommendations [7], and drug-to-drug similarity measurements [2]. Those applications usually are performed through a link prediction process and can be divided into two steps: 1) learn embeddings of medical entities and relations, 2) make predictions/recommendations according to these embeddings.

Several approaches rely on medical KG embeddings for recommendation tasks, work from Gong et al. [7] proposed a medicine recommendation framework that embeds medical entities such as diseases, medicines, patients, and their corresponding relations into a shared lower dimensional space. The authors use the embeddings to decompose the task into a link prediction process while considering the patient’s diagnosis and adverse drug reactions. Chen et al. [3] proposed a framework operating on medical KG for thyroid treatment recommendation with cold start based on TransD [10] and network embeddings with a hierarchical structure.

While recent research [20,12,7,3,2] employs traditional KG embedding methods [1,19,13,10] as a first step for representing patients and medical entities, it lacks the consideration of demographic meta-data. However, such information is very advantageous and even necessary for medical tasks. Several works [22,9,18,14] have investigated and shown the importance and development of demographics in different medical tasks and challenges. Considering this, we argue that incorporating demographics as part of medical KGs allows us to retain patients’ generic information and construct more accurate representations for medical entities.

Hence, in this paper, we propose DARLING (**D**emographic **A**ware **p**Robabi**L**-istic **m**ed**I**cal **k**Nowledge **e**mbeddin**G**) – the first demographic-aware medical KG embedding framework that explicitly incorporates demographics in the medical entities space by associating patient demographics with a corresponding hyperplane. Our framework leverages probabilistic features within medical entities for learning their representations through demographic guidance. We evaluate DARLING on link prediction for treatments and medicines, where it achieves improved results compared to multiple existing KG embedding models on standard metrics. Furthermore, for evaluation purposes, we construct a medical KG from the MIMIC-III [11] data, which comprises medical elements relating to patient admissions, such as demographics, disease diagnosis, treatment procedures, etc. The medical KG contains diseases, treatments and medicines, where our construction method automatically links all extracted entities with existing biomedical knowledge graphs, including ICD-9 [16] ontology and DrugBank [21]. With our work we make the following key contributions to the state of the art:

- We propose DARLING, the first demographic-aware framework for learning probabilistic medical KG embeddings.
- We provide a method to construct a medical KG with demographics meta-data that links all extracted entities with existing biomedical knowledge graphs.
- We demonstrate DARLING’s effectiveness through extensive experiments and show its superior performance via link prediction on treatments and medicines. We also illustrate the sensitivity of different demographic categories in our framework.

To facilitate reproducibility and reuse our framework implementation, alongside the medical knowledge graph construction method, the results are also publicly available¹.

The rest of the paper is structured as follows: Section 2 summarises the related work and Section 3 presents the proposed DARLING framework. Section 4 describes the experiments, including the experimental setup and the evaluation. Finally, we conclude in Section 5.

2 Related Work

Our work lies at the intersection of medical KG embeddings and graph-based embedding approaches that employ patient demographics. In this section, we describe previous efforts and refer to different approaches.

Current efforts on knowledge graphs have concentrated on automatic knowledge base completion and population. Multiple KGs [6,5] have been constructed from vast volumes of medical databases over the last years. Medical KGs contain medical facts of medicines and diseases and provide a pathway for medical discovery and applications, such as effective treatment and medicine recommendation. Unfortunately, such medical KGs suffer from severe data incompleteness problems, which impedes their application in clinical medicine. Celebi et al. [2] proposed a KG embedding approach for drug-to-drug interaction prediction in a realistic scenario. Hettige et al. [8] proposed an EMR embedding framework that introduces a graph-based data structure to capture visit-code associations in an attributed bipartite graph and the temporal sequencing visits through a point process. Moreover, Choi et al. [4] proposed another approach that learns the representations for both medical codes and visits from large datasets. Both works [8,4] consider patient demographics on medical codes; however, they only focus on visit-code embeddings and do not directly harness medical KG embeddings or any medical task. Hence, specific medical information is insufficiently tailored. To the best of our knowledge, none of the existing approaches employs demographics as an essential factor for medical KG embeddings.

3 DARLING

This section presents a detailed description of DARLING (Figure 1). In this work, we use a boldface lower-case letter \mathbf{x} to denote a vector, $\|\mathbf{x}\|_p$ to represent its l_p norm and d for the embedding dimension. Please refer to the appendix for the background of our framework, detailed experiment results, etc.

3.1 Probabilistic Medical Knowledge Graph with Demographics

A KG can be denoted as a set of triples $\mathcal{K} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ where \mathcal{E} and \mathcal{R} are the set of entities and relations respectively. More precisely, the KG comprises

¹ <https://github.com/AynurGuluzade/DARLING>

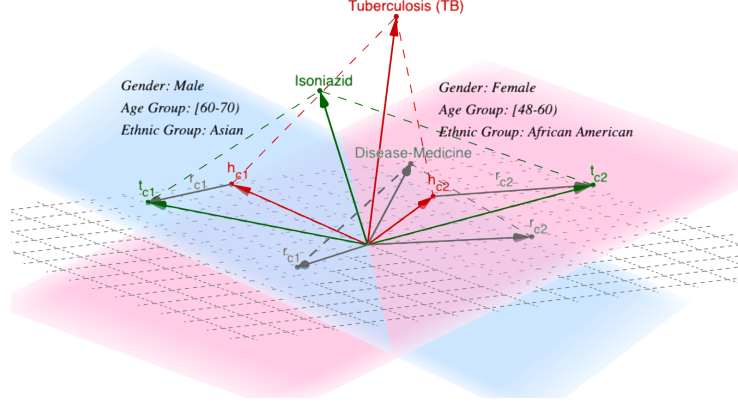


Fig. 1. The medical triple (Tuberculosis (TB), Disease-Medicine, Isoniazid) is valid for both demographic sets c_1 and c_2 . $\mathbf{h}_{(c1)}$, $\mathbf{r}_{(c1)}$ and $\mathbf{t}_{(c1)}$ are the projections of the triple onto the demographic hyperplane c_1 . As we can observe, the hyperplanes encode demographic information such as gender, age group, and ethnic group. DARLING learns demographic-aware representations of medical entities and relations by minimizing the translational distance $\|\mathbf{h}_c + \mathbf{r}_c - \mathbf{t}_c\|_2$ based on the triple probability score.

triples $(h, r, t) \in \mathcal{K}$ in which $h, t \in \mathcal{E}$ represent a triples' respective head and tail entities and $r \in \mathcal{R}$ represents its relation. The direction of a relationship indicates the roles of the entities, i.e., head or tail entity. In our scenario, we consider medical entities, extracted from EMR data, such as disease diagnosis, treatment procedures, and medicines; therefore $\mathcal{E} \subseteq \mathcal{D} \cup \mathcal{P} \cup \mathcal{M}$, where \mathcal{D} is the set of diagnosis, \mathcal{P} is the set of procedures and \mathcal{M} is the set of medicines. In particular, we construct triples with $h \in \mathcal{D}$ and $t \in \mathcal{P} \cup \mathcal{M}$. Hence, our medical KG is denoted as $\mathcal{K} \subseteq \mathcal{D} \times \mathcal{R} \times \mathcal{P} \cup \mathcal{M}$.

We incorporate patient demographic meta-data such as gender, age group and ethnic group, by adding a new demographic dimension to the KG triples. We define the demographic set as $\mathcal{C} \subseteq \mathcal{G} \times \mathcal{A} \times \mathcal{T}$, where \mathcal{G} is the set of genders, \mathcal{A} the set of age groups and \mathcal{T} the set of ethnic groups. Our medical KG contains quadruples $(h, r, t, c) \in \mathcal{K}$ where $c \in \mathcal{C}$ is a demographic set for which the corresponding medical triple (h, r, t) holds. We aim to incorporate the demographic meta-fact c directly into our learning algorithm, to learn demographic-aware embeddings of the KG elements.

Moreover, inspired by [12], we strengthen each quadruple (h, r, t, c) existence by introducing a statistical probability that indicates how likely is the particular triple (h, r, t) to appear with the demographic set c . Specifically, we associate each quadruple with the probability $p(h, r, t, c)$ which is calculated as, $p(h, r, t, c) = \mathcal{N}_{(h, r, t, c)} / \mathcal{N}_h$, where $\mathcal{N}_{(h, r, t, c)}$ is the number of EMR admissions that the quadruple (h, r, t, c) was extracted, while \mathcal{N}_h is the number of admissions that contain the medical entity h . In contrast to [12], our probability value

is calculated by considering also the demographic set c and not only the triple (h, r, t) .

3.2 Demographic-Guided Translation

DARLING operates as an interaction model $f : \mathcal{D} \times \mathcal{R} \times \mathcal{P} \cup \mathcal{M} \rightarrow \mathbb{R}$ that computes a real-value score representing a medical KG quadruple’s plausibility, given the embeddings for the entities, relations, and demographic sets. In our approach, we want the medical entity to have a distributed representation associated with different demographic sets. We achieve that by representing a demographic set as a hyperplane i.e., we will have $|\mathcal{C}|$ number of different hyperplanes represented by normal vectors $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{|\mathcal{C}|}$, where $|\mathcal{C}|$ denotes the total number of demographic sets. Therefore, we attempt to segregate the space into different demographic zones with the help of the hyperplanes. In this way, medical triples valid with demographic set c are projected onto the demographic-specific hyperplane $\mathbf{w}_c \in \mathbb{R}^d$, where their translational distance (in our case similar to [1]) is minimized. Figure 1 illustrates an example where the medical triple (h, r, t) is valid for both demographic sets c_1 and c_2 . Hence it is projected onto the hyperplanes corresponding to those demographic sets. Using the triple embeddings $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$, the projected representations on \mathbf{w}_c are computed as:

$$\begin{aligned} \mathbf{h}_c &= \mathbf{h} - \mathbf{w}_c^\top \mathbf{h} \mathbf{w}_c, \\ \mathbf{r}_c &= \mathbf{r} - \mathbf{w}_c^\top \mathbf{r} \mathbf{w}_c, \\ \mathbf{t}_c &= \mathbf{t} - \mathbf{w}_c^\top \mathbf{t} \mathbf{w}_c, \end{aligned} \tag{1}$$

where $\mathbf{h}_c, \mathbf{r}_c, \mathbf{t}_c \in \mathbb{R}^d$. Consequently, we expect that a positive triple, valid with demographic set c , will have the mapping as $\mathbf{h}_c + \mathbf{r}_c \approx \mathbf{t}_c$. Accordingly, our scoring function is defined as:

$$f_c(h, r, t) = \|\mathbf{h}_c + \mathbf{r}_c - \mathbf{t}_c\|_p, \tag{2}$$

where $p \in \{1, 2\}$ is a hyper-parameter.

Alongside the entity and relation embeddings, we also learn $\{\mathbf{w}_c\}_{c=1}^{|\mathcal{C}|}$ for each demographic set c . Furthermore, by projecting the triple onto its demographic hyperplane, we incorporate demographic knowledge into the entity and relation embeddings, i.e., the same distributed representation will have a different role in different demographic sets.

3.3 Optimization through Probability Score

DARLING employs a margin-based pairwise ranking loss to differentiate between correct/positive and incorrect/negative triples. The negative triples are obtained

by corrupting the positive one; thus, the pairs often share common head or tail entities and relations. Formally, we aim to minimize the following loss function:

$$\mathcal{L} = \sum_{c \in [C]} \sum_{x \in \mathcal{D}_c^+} \sum_{y \in \mathcal{D}_c^-} \max(0, g_c(x) - g_c(y) + \gamma), \quad (3)$$

with respect to the entity, relation and demographic set vectors. \mathcal{D}_c^+ is the set of valid triples with demographic set c , the negative triples are drawn from the set \mathcal{D}_c^- , and γ is a margin separating correct and incorrect triples.

Unlike existing approaches that employ pairwise ranking loss, DARLING does not directly use the score function $f_c(h, r, t)$. Instead, we utilize probability values p for estimating the scores for optimization. First, we introduce a mapping function or probability score function f_p that allows us to map each quadruple probability value $p(h, r, t, c)$ into a score. This function is defined as:

$$f_p(h, r, t, c) = \lambda \ln p(h, r, t, c)^{-1}, \quad (4)$$

where λ is a scaling factor. To avoid a 0 denominator for negative quadruples, we introduce a constant probability value $\varepsilon_n > 0$. Furthermore, we set a minimum probability value for positive quadruples as ε_p , where $\varepsilon_p > \varepsilon_n$. Second, we define the function $g_c()$ as the absolute difference of the probability score f_p and the score function f_c . Formally, this is described as:

$$g_c(h, r, t) = |f_p(h, r, t, c) - f_c(h, r, t)|. \quad (5)$$

In this way, the DARLING optimization process allows us to learn representations that would satisfy each quadruple probability value. Specifically, the quadruple entities with a high probability value will have representations closer in space compared to those with a low probability.

4 Experiments

4.1 Datasets and Medical Knowledge Graph Construction

We perform experiments on a real EMR dataset – MIMIC-III [11], and two biomedical knowledge graphs, DrugBank [21] and the ICD-9 [16] ontology. MIMIC-III (Medical Information Mart for Intensive Care) comprises information related to patients admitted to critical care units at a large tertiary care hospital. The dataset contains distinct information about 46,520 patients, 58,976 admissions, and 1,517,702 prescription records associated with 6,985 diagnosis, 2,032 procedures and 4,525 medicines. For our work, we extract patient demographics, disease diagnosis, treatment procedures, and medicines. We link the extracted medicines to DrugBank, which is a bioinformatics resource that consists of medicine-related entities. The DrugBank KG contains 8,054 medicines, 4,038 other related entities (e.g., protein or drug targets) and 21 relationships. Moreover, we link extracted diseases and treatments with ICD-9 ontology (International Classification of Diseases, Ninth Revision) which contains 13,000 international standard codes of diagnosis and procedures. We connect MIMIC-III,

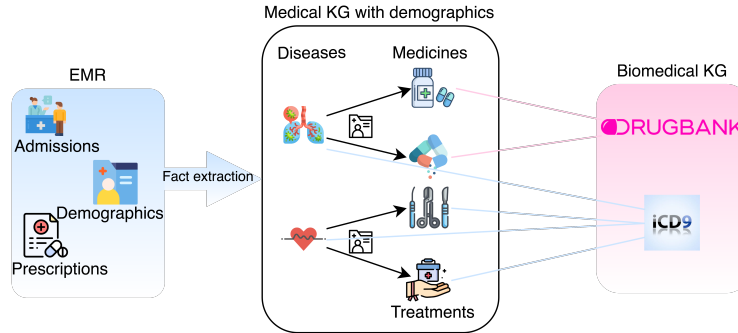


Fig. 2. Overview of our Medical KG construction process.

Table 1. Medical KG number of entities, relations and demographics.

Entities		Relations		Demographics	
#Disease	6,968	#Disease_to_Treatment	58,225	#Gender	2
#Treatment	1,475	#Disease_to_Medicine	67,916	#Age group	6
#Medicine	846			#Ethnic group	7
#Total	9,289	#Total	126,141	#Total (sets)	79

DrugBank, and the ICD-9 ontology by constructing the medical KG (c.f. Figure 2). Our medical KG includes 9,289 distinct entities and two types of relations – Disease_to_Treatment and Disease_to_Medicine. Regarding demographics, we end up with 79 different demographic set combinations (gender, age group and ethnic group). Finally, our KG contains 126,141 distinct quadruples, 100,912 of which we use for training, 10,091 for validation and 15,138 for testing. Table 1 gives details on the constructed KG.

4.2 Experimental Setup

Model Configurations For the experiments, we selected Adam optimizer, and we employ batch sizes of $b = \{128, 256, 512\}$, embedding dimensions of $d = \{128, 256, 512\}$, learning rates $lr = \{0.01, 0.001, 0.0001\}$, a margin $\gamma = 1$ and $p = 2$ for the scoring function. Furthermore, for the probabilistic hyper-parameters λ , e_p and e_n we use the values of 10^{-2} , 10^{-4} , and 10^{-15} respectively. We train DARLING for 100 epochs and select the best state by the corresponding lowest mean rank on the validation set.

Models for Comparison For evaluating the performance of our framework, we compare against the following methods:

Table 2. Results on link prediction for treatments and medicines. DARLING outperforms all other methods.

Task	Disease-Treatment			Disease-Medicine		
Methods	Mean Rank	Hits@3	Hits@10	Mean Rank	Hits@3	Hits@10
TransE [1]	73.94	15.71%	47.40%	27.04	15.89%	54.33%
TransH [19]	75.56	16.51%	48.60%	27.71	16.23%	55.46%
TransR [13]	115.12	12.64%	30.34%	45.74	14.33%	39.16%
TransD [10]	84.66	17.34%	47.64%	33.51	15.97%	55.76%
PrTransE [12]	69.69	16.29%	47.21%	27.51	15.45%	54.80%
PrTransH [12]	69.01	16.89%	47.25%	26.71	16.14%	55.73%
DARLING (ours)	64.65	22.71%	52.19%	22.86	26.90%	61.73%

- **TransE** [1] is a simple but effective translation-based model. A major advantage of TransE is its computational efficiency, which enables its use for large-scale KGs.
- **TransH** [19] is an extension of TransE where each relation is represented by a hyperplane. Our proposed framework, DARLING also adjusts TransE in a similar way by treating the demographic sets as hyperplanes.
- **TransR** [13] explicitly considers entities and relations as different objects and therefore represents them in different vector spaces.
- **TransD** [10] is similar to TransR, however, instead of performing the same relation-specific projection for all entity embeddings, entity-relation-specific projection matrices are constructed.
- **PrTransE & PrTransH** [12] are extensions of TransE and TransH, which introduce triple probabilities. Our optimization via probability score was inspired by this work, however, our approach differs considerably in multiple aspects.

4.3 Results

We apply the following frequently used metrics to summarize the overall performance: 1) Mean rank (MR): which represents the average rank of the test triples, where smaller values indicate better performance. 2) Hits@K: which denotes the ratio of the test triples that have been ranked among the top- k triples, where larger values indicate better performance. We report results for $k = \{3, 10\}$.

Table 2 illustrates the results for treatment and medicine prediction from our constructed medical KG. As indicated, DARLING outperforms all other traditional KG embedding approaches on both mean rank and Hits@K. More precisely, for Hits@3 and Hits@10, our framework has an absolute difference of at least 5 points from the baselines, while for Hits@3 on medicine prediction, this difference is increased to 10 absolute points. DARLING also has lower mean rank compared to all other baselines. The improved performance arises from

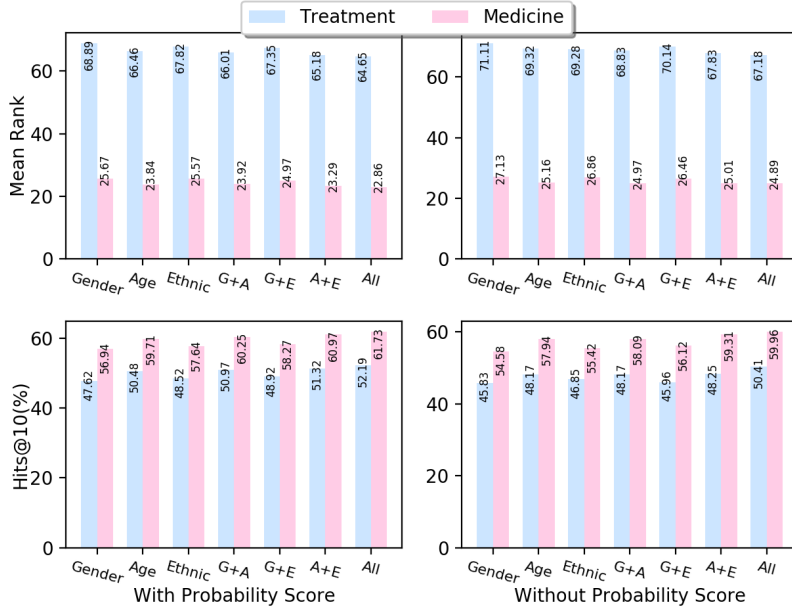


Fig. 3. DARLING results based on different demographic combinations with and without probability score.

DARLING’s ability to generate more accurate embeddings of the medical entities. The demographic-based hyperplanes that our framework produces allow it to categorize diseases, treatments and medicines into “subspaces” where it learns their embeddings. Furthermore, the quadruple probability score provides some weighting for the distance between the entities. In contrast, all the baselines project the medical entities in one or two spaces (since the number of relations is two). As demonstrated by our results, this is not sufficient to represent all entities accurately.

4.4 Demographic and Probability Score Sensitivity

We perform experiments to identify the demographic and probability score sensitivity of our framework. In particular, we aim to recognize which demographic category (gender, age group, ethnic group) is more effective and whether the probability score impacts our results. To do so, we adjust DARLING to construct hyperplanes with all demographics individually and with all possible combinations. At the same time, we run the framework by including and excluding the probability scores.

Figure 3 illustrates the results of this experiment for mean rank and Hits@10. Regarding the demographics, we obtain that “Age group” is the most prominent one since the results are better compared to “Gender” and “Ethnic group”. We find this accurate since most diseases are highly correlated with patient

age. Furthermore, we observe that with “Ethnic group” we obtain better results than with “Gender”. When joining the categories “Age group + Ethnic group (A+E)” performs better than “Gender + Age group (G+A)” and “Gender + Ethnic group (G+E)”. It is worth mentioning that even when employing one demographic category together with the probability score, we still acquire better results than any of the baseline methods. We obtain the highest possible results when we utilise all three demographic categories, as we also do for DARLING. Moreover, the results indicate a higher performance by 1-3 points when using the probability score.

5 Conclusions

In this article, we focus on medical knowledge graph embeddings of electronic medical records. We provide a demographic-aware embedding framework that explicitly incorporates demographics in the medical entities’ space by associating patient demographics (gender, age, ethnicity) with a corresponding hyperplane. Our framework leverages probabilistic features of entities for learning their embeddings through demographic guidance. Furthermore, for evaluating our approach, we present a method to construct a medical KG from EMR data and automatically link all extracted entities with existing biomedical knowledge graphs (ICD-9 and DrugBank). We empirically show that our model achieves the best results in link prediction for treatments and medicines, compared to other traditional KG embedding approaches. Moreover, we perform a demographic sensitivity experiment and discover that age is the demographic category that significantly affects our high results. We also show the importance of probability score in our framework. For future work, we intend to employ our framework as an embedding method for existing medical recommendation systems.

References

1. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: NIPS. Curran Associates, Inc. (2013)
2. Celebi, R., Uyar, H., Yasar, E., Gumus, O., Dikenelli, O., Dumontier, M.: Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. BMC Bioinformatics (2019)
3. Chen, D., Ma, C., Wu, Y.: Clinical knowledge graph embeddings with hierarchical structure for thyroid treatment recommendation. In: IEEE DASC/PiCom/CBDCoM/CyberSciTech (2019)
4. Choi, E., Bahadori, M.T., Searles, E., Coffey, C., Thompson, M., Bost, J., Tejedoro-Sojo, J., Sun, J.: Multi-layer representation learning for medical concepts. In: 22nd ACM SIGKDD. Association for Computing Machinery (2016)
5. Dumontier, M., Callahan, A., Cruz-Toledo, J., Ansell, P., Emonet, V., Belleau, F., Droit, A.: Bio2rdf release 3: a larger connected network of linked data for the life sciences. In: ISWC-PD’14 (2014)
6. Ernst, P., Siu, A., Weikum, G.: Knowlife: a versatile approach for constructing a large knowledge graph for biomedical sciences. BMC bioinformatics (2015)

7. Gong, F., Wang, M., Wang, H., Wang, S., Liu, M.: Smr: Medical knowledge graph embedding for safe medicine recommendation. *Big Data Research* (2021)
8. Hettige, B., Wang, W., Li, Y.F., Le, S., Buntine, W.: Medgraph: structural and temporal representation learning of electronic medical records. In: *ECAI Digital - 2020*. IOS Press (2020)
9. Ihra, G.C., Lehberger, J., Hochrieser, H., Bauer, P., Schmutz, R., Metnitz, B., Metnitz, P.G.H.: Development of demographics and outcome of very old critically ill patients admitted to intensive care units. *Intensive Care Medicine* (2012)
10. Ji, G., He, S., Xu, L., Liu, K., Zhao, J.: Knowledge graph embedding via dynamic mapping matrix. In: *53rd ACL-IJCNLP*. Association for Computational Linguistics (2015)
11. Johnson, A.E., Pollard, T.J., Shen, L., Li-Wei, H.L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G.: Mimic-iii, a freely accessible critical care database. *Scientific data* (2016)
12. Li, L., Wang, P., Wang, Y., Wang, S., Yan, J., Jiang, J., Tang, B., Wang, C., Liu, Y.: A method to learn embedding of a probabilistic medical knowledge graph: Algorithm development. *JMIR Med Inform* (2020)
13. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: *28th AAAI*. AAAI Press (2015)
14. Qin, P., Agerbo, E., Mortensen, P.B.: Suicide risk in relation to socioeconomic, demographic, psychiatric, and familial factors: A national register-based study of all suicides in denmark, 1981–1997. *American Journal of Psychiatry* (2003)
15. Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S., Sontag, D.: Learning a health knowledge graph from electronic medical records. *Scientific Reports* (2017)
16. Schriml, L.M., Arze, C., Nadendla, S., Chang, Y.W., Mazaitis, M., Felix, V., Feng, G., Kibbe, W.A.: Disease ontology: a backbone for disease semantic integration. *Nucleic acids research* (2012)
17. Sherimon, P., Krishnan, R.: Ontodiabetic: an ontology-based clinical decision support system for diabetic patients. *Arabian Journal for Science and Engineering* (2016)
18. Standish, L.J., Greene, K.B., Bain, S., Reeves, C., Sanders, F., Wines, R.C.M., Turet, P., Kim, J.G., Calabrese, C.: Alternative medicine use in hiv-positive men and women: Demographics, utilization patterns and health status. *AIDS Care* (2001)
19. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: *28th AAAI*. AAAI Press (2014)
20. Wanyan, T., Kang, M., Badgeley, M.A., Johnson, K.W., De Freitas, J.K., Chaudhry, F.F., Vaid, A., Zhao, S., Miotto, R., Nadkarni, G.N., Wang, F., Rousseau, J., Azad, A., Ding, Y., Glicksberg, B.S.: Heterogeneous graph embeddings of electronic health records improve critical care disease predictions. In: *AIME*. Springer International Publishing (2020)
21. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C., Wilson, M.: DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* (2017)
22. X, H., JD, R., KM, M., JC, F.: Demographics as predictors of suicidal thoughts and behaviors: A meta-analysis. *PLoS ONE* (2017)
23. Xiang, X., Wang, Z., Jia, Y., Fang, B.: Knowledge graph-based clinical decision support system reasoning: A survey. In: *IEEE 4th DSC* (2019)

Appendix

A Background

In this section, we review existing methods for knowledge graph representation learning [1], [19]. A KG is considered as a set of entities \mathcal{E} and relations \mathcal{R} . The set of directed edges, \mathcal{D}^+ comprises triples (h, r, t) where a direction of relation r is from head h to tail t entity

TransE [1] is a simple and efficient translational based distance model. It models the relation as a translation vector between head and tail entity vectors. For the given two entity vectors $\mathbf{h}, \mathbf{t} \in \mathbb{R}^n$, it maps the relation as translation vector $\mathbf{r} \in \mathbb{R}^n$, i.e., $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ for observed triple $\mathbf{h}, \mathbf{r}, \mathbf{t}$. Thus, the distance based scoring function is defined as:

$$f(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{l_1/l_2}, \quad (6)$$

where, $\|\cdot\|_{l_1/l_2}$ is the l_1 or l_2 -norm of the difference vector. $f(h, r, t)$ will be minimized for plausible triples. A margin based pairwise ranking loss is used to differentiate between correct and incorrect triples by minimizing their TransE score difference. Formally, the loss function is defined:

$$\sum_{x \in \mathcal{D}^+} \sum_{y \in \mathcal{D}^-} \max(0, f(x) - f(y) + \gamma), \quad (7)$$

with respect to the entity and relation vectors. γ is a margin hyperparameter. \mathcal{D}^+ stores only positive triples, i.e., observed triples in KG. \mathcal{D}^- is the set with negative examples that are drawn randomly.

Despite its simplicity and efficiency, TransE cannot model 1-to-N, N-to-1, and N-to-N type of relations as it does not learn a distributed representation of entities. To tackle these flaws, TransH [19] was introduced to model a relation r as a vector on a relation specific hyperplane, and project entities associated with it on the corresponding hyperplane for learning the entities' distributed representation.

B Negative Sampling

For negative sampling, DARLING utilizes a uniform demographic agnostic approach, where it considers the set of all the triples that do not belong to the medical KG, irrespective of the demographic dimension. Formally, for the demographic set c , the negative samples are drawn from the set,

$$\begin{aligned} \mathcal{D}_c^- = & \{(h', r, t) | h' \in \mathcal{E}, (h', r, t) \notin \mathcal{D}^+\} \\ & \cup \{(h, r, t') | t' \in \mathcal{E}, (h, r, t') \notin \mathcal{D}^+\}. \end{aligned} \quad (8)$$

Table 3. Demographic statistics for each category. Our medical KG contains data from 46,520 unique patients.

Gender #Patients		Age Group #Patients		Ethnic Group #Patients	
male	26,121	[0-18)	7,942	white	32,372
female	20,399	[18-48)	7,005	black	3,871
		[48-60)	7,515	asian	1,690
		[60-70)	7,860	hispanic	1,642
		[70-80)	7,939	native	46
		≥ 80	8,259	other	1,489
				unknown	5,410

C Demographic Statistics

Table 3 illustrates the number of unique patients that belong to each demographic category. We constructed our medical KG using data from 46,520 patients and 58,976 admissions related to them. The grouping of age values (years) was done by us, considering that we wanted to distribute the patients equally in different groups. The genders and ethnic groups are adopted from MIMIC-III data [11].

D Inference

For inference, we describe how DARLING can be used for medical recommendation tasks through a link prediction process. Given a query patient with demographic set $c \in \mathcal{C}$ (gender, age, ethnicity) and the query disease diagnosis $d \in \mathcal{D}$, we use DARLING to project the disease d into the hyperplane w_c and recommend top- k treatments and medicines. More precisely, given a query $q = (c, d)$, for each treatment procedure $\forall p \in \mathcal{P}$ and medicine $\forall m \in \mathcal{M}$ we compute its triple score with d (i.e. $f_c(d, r, p)$, $f_c(d, r, m)$) on the demographic hyperplane w_c , and then select the treatment p and medicine m with the top- k highest ranking scores as the recommendation.

E Detailed Results

Table 4 presents detailed results of our framework. In particular, we illustrate results using all possible demographic category combinations, and we further provide results by including and excluding the probability scores. At the same time, we present the results of all other baselines. As we can see, DARLING outperforms all baselines when using all demographic categories and including the probability scores

Table 4. Detailed results of our experiments.

Task	Disease-Treatment				Disease-Medicine			
Methods	Mean Rank		Hits@10		Mean Rank		Hits@10	
TransE [1]	73.94		47.40%		27.04		54.33%	
TransH [19]	75.56		48.60%		27.71		55.46%	
TransR [13]	115.12		30.34%		45.74		39.16%	
TransD [10]	84.66		47.64%		33.51		55.76%	
PrTransE [12]	69.69		47.21%		27.51		54.80%	
PrTransH [12]	69.01		47.25%		26.71		55.73%	
Probability score	with	without	with	without	with	without	with	without
DARLING (Gender)	68.89	71.11	47.62%	45.83%	25.67	27.13	56.94%	54.58%
DARLING (Age)	66.46	69.32	50.48%	48.17%	23.84	25.16	59.71%	57.94%
DARLING (Ethnicity)	67.82	69.28	48.52%	46.85%	25.57	26.86	57.64%	55.42%
DARLING (G+A)	66.01	68.83	50.97%	48.17%	23.92	24.97	60.25%	58.09%
DARLING (G+E)	67.35	70.14	48.92%	45.96%	24.97	26.46	58.27%	56.12%
DARLING (A+E)	65.18	67.83	51.32%	48.25%	23.29	25.01	60.97%	59.31%
DARLING (all)	64.65	67.18	52.19%	50.41%	22.86	24.89	61.73%	59.96%