On the Generative Discovery of Structured Medical Knowledge

Chenwei Zhang[†], Yaliang Li[‡], Nan Du[‡], Wei Fan[‡], Philip S. Yu[†§]

[†]University of Illinois at Chicago, Chicago, IL 60607
[‡]Tencent Medical AI Lab, Palo Alto, CA 94301
[§]Institute for Data Science, Tsinghua University, Beijing, China
{czhang99,psyu}@uic.edu,{yaliangli,ndu,davidwfan}@tencent.com

ABSTRACT

Online healthcare services can provide the general public with ubiquitous access to medical knowledge and reduce medical information access cost for both individuals and societies. However, expanding the scale of high-quality yet structured medical knowledge usually comes with tedious efforts in data preparation and human annotation. To promote the benefits while minimizing the data requirement in expanding medical knowledge, we introduce a generative perspective to study the relational medical entity pair discovery problem. A generative model named *Conditional Relationship Variational Autoencoder* is proposed to discover meaningful and novel medical entity pairs by purely learning from the expression diversity in the existing relational medical entity pairs. Unlike discriminative approaches where high-quality contexts and candidate medical entity pairs are carefully prepared to be examined by the model, the proposed model generates novel entity pairs directly by sampling from a learned latent space without further data requirement. The proposed model explores the generative modeling capacity for medical entity pairs while incorporating deep learning for hands-free feature engineering. It is not only able to generate meaningful medical entity pairs that are not yet observed, but also can generate entity pairs for a specific medical relationship. The proposed model adjusts the initial representations of medical entities by addressing their relational commonalities. Quantitative and qualitative evaluations on real-world relational medical entity pairs demonstrate the effectiveness of the proposed method in generating relational medical entity pairs that are meaningful and novel.

CCS CONCEPTS

• **Computing methodologies** → **Knowledge representation and reasoning**; **Neural networks**; • **Applied computing** → **Health care information systems**; • **Mathematics of computing** → *Variational methods*;

KEYWORDS

Knowledge Discovery; Generative Modeling; Medical Entity Pair

KDD 2018, August 19–23, 2018, London, United Kingdom

ACM ISBN 978-1-4503-5552-0/18/08…$15.00
https://doi.org/10.1145/3219819.3220010

ACM Reference Format:
Chenwei Zhang, Yaliang Li, Nan Du, Wei Fan, Philip S. Yu. 2018. On the Generative Discovery of Structured Medical Knowledge. In *KDD 2018: 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3219819.3220010

1 INTRODUCTION

Increasingly, people engage in health services on the Internet [12, 50]. The healthcare services can provide the general public with ubiquitous access to medical knowledge and reduce the information access cost significantly. The structured medical knowledge discussed in this paper are binary ones. A relational medical entity pair, which consists of two medical entities with a semantic connection between them, is an intuitive representation that distills human medical knowledge [34]. For example, the $Disease \xrightarrow{Cause} Symptom$ relationship indicates a "Cause" relationship from a disease entity to a symptom entity which is caused by this disease, such as the medical entity pair <*synovitis*, *joint pain*>. Table 1 shows some relational medical entity pairs for common medical relationships.

The ability to understand, reason and generalize is central to human intelligence [29]. However, it possesses significant challenges for machines to understand and reason about the relationships between two entities [33]. Specifically, real-world relational medical entity pairs posses certain challenging properties to deal with: First, various linguistic expressions are usually used for a medical entity. For example, *nose plugged*, *blocked nose* and *sinus congestion* are symptom entities that share the same meaning but expressed very differently. Second, one medical relationship may be instantiated by entity pairs in varying granularities or different relationship strength. For instance, Disease $\xrightarrow{Cause}$ Symptom relationship may include coarse-grained entity pairs like <*rhinitis*, *nose plugged*>, while <*acute rhinitis*, *nose plugged*>, <*chronic rhinitis*, *nose plugged*> are considered as fine-grained entity pairs. As for the relationship strength, <*cold*, *fatigue*> has greater relationship strength than <*cold*, *ear infections*> as cold rarely cause serious

Relational Medical Entity Pairs	Medical Relationship
<*synovitis*, *joint pain*>	$Disease \xrightarrow{Cause} Symptom$
<*stiffness of a joint*, *orthopedics*>	$Symptom \xrightarrow{Belongto} Department$
<*muscular contusion*, *disinsertion*>	$Symptom \xrightarrow{RelatedTo} Symptom$

Table 1: Sample medical relationships and relational medical entity pairs.

complications such as ear infections. It is straightforward for human beings yet still challenging for a machine to give insights on their relational commonalities.

To expand the scale of relational medical entity pairs, discriminative methods for relation extraction [2, 3, 16, 23, 26, 32, 42] or knowledge graph completion [6, 13, 22, 38, 43, 44] can be adopted. Despite achieving decent performance in identifying the relationship given candidate entity pairs, their performances capitalize on well-prepared context as external resources, or high-quality candidate entity pairs given for testing. For relation extraction methods which aim to examine whether or not a semantic relationship exists between two entities given the context, they require a substantial collection of contexts over a full spectrum of relationships we would like to work on: e.g. contexts obtained from free-text corpora where two entities co-occur in the same sentence with a relationship between them, from existing domain-specific knowledge graphs [1, 28], or from web tables and links [21]. As medical relationships in the real-world are becoming more and more complex and diversely expressed [49], such context is hard to obtain. Knowledge graph completion methods tell what kind of relationship, and how likely a relationship is going to be formulated between two given entities. They usually do not require contexts for training. However, they are vulnerable to the "garbage-in, garbage-out" situation: we can not obtain the rational medical entity pairs for a specific relationship when no high-quality entity pairs having that relationship are among the candidate entity pairs. The choice of candidates may involve additional human annotation; otherwise, any dyadic combinations of medical entities need to be annotated and tested by the model, which is tedious and labor-intensive.

In both tasks mentioned above, the lacking preparation of external resource or additional human annotation is fatal to successfully discover the structured medical knowledge. Therefore, it is crucial for us to discover relational medical entity pairs without substantial data requirement.

Problem Studied: We propose a novel research problem called RElational Medical Entity-pair DiscoverY (REMEDY), which aims at understanding the medical relationship solely from the existing medical entity pairs via their diverse expressions. We aim to discover meaningful and novel entity pairs of a specific medical relationship in a generative fashion, without sophisticated feature engineering and substantial data requirement such as large-scale free-text as contexts, or further data preparation.

Proposed Model: A generative model named Conditional Relationship Variational Autoencoder (CRVAE) is introduced for relational medical entity pair discovery. The proposed model exploits the generative modeling capacity roots in Bayesian inference while incorporating deep learning for powerful hands-free feature engineering. The model takes entity pairs and their medical relationship types as the input. It encodes relational medical entity pairs into a latent space conditioned on the relationship type. Based on pre-trained entity representations, the encoding process further addresses relationship-enhanced entity representations, entity interactions, and expressive latent variables. The latent variables are decoded to reconstruct entity pairs. Once trained, the generator samples directly from the learned latent variables and decodes them into novel medical entity pairs having a specific relationship.

Overall, CRVAE has three notable strengths:

CRVAE significantly lower the data requirement for structured medical knowledge discovery by learning directly from the existing medical entity pairs. In the medical domain, the diversely expressed medical entity pairs offer significant advantages for the model to understand their commonalities from various medical expressions without additional contexts.

CRVAE generates novel, meaningful entity pairs by its generative nature. The generator adopts a density-based sampling strategy that decodes the sampled latent variables into entity pairs. Unlike discriminative methods which learn the discrepancies among different medical relationships, the CRVAE models the distribution of diversely expressed entity pairs within each relationship, so as to generate new ones rationally.

CRVAE can generate entity pairs for a particular relationship without additional data preparation. Discriminative models rely on the quality of candidate entity pairs to obtain novel, meaningful entity pairs efficiently. CRVAE's conditional inference ability makes it more efficient to discover structured medical knowledge for specific medical relationships.

The contributions of this paper can be summarized as follows:

- We introduce a generative perspective to study the Relational Medical Entity-pair Discovery (REMEDY) problem, which aims to expand the scale of high-quality yet novel structured medical knowledge with minimized data requirement.
- We propose a model named Conditional Relationship Variational Autoencoder (CRVAE) that generatively discover relational medical entity pairs for a specific relationship, by solely learning the commonalities from diversely expressed entity pairs without sophisticated feature engineering.
- We obtain relationship-enhanced entity representations as a byproduct of the encoding process of the model.

2 RELATED WORKS

Deep Generative Models: Recent years have witnessed an increasing interest in deep generative models that generate observable data based on hidden parameters. Unlike Generative Adversarial Networks (GANs) [31] which generate data based on arbitrary noises, the Variational Autoencoders (VAEs) [18] setting we adopted is more expressive since it tries to model the underlying probability distribution of the data by latent variables so that we can sample from that distribution to generate new data accordingly. An increasing number of models and applications are proposed which consider data in different modalities, such as generating images [15, 30] or natural language [7, 24, 45]. [46] works on generative relation discovery with a probabilistic graphic model that requires hand-crafted relation-level features. As far as we know, the relational medical entity pair discovery problem we studied in this paper, which is suitable for deep generative modeling, has not been studied in a generative perspective with restricted data requirement.
Knowledge Graph Completion: Existing knowledge graph completion methods [6, 13, 22, 38, 43, 44] are discriminative models. During training, those methods are trained to distinguish entity pairs of one relationship from another [22, 48], or to identify meaningful entity pairs from randomly sampled negative entity pairs with no relationships [5, 35]. During testing, some candidate entity pairs are prepared ahead of time and given to the model. The

model examines what kind of, and how likely there is a relationship for each candidate entity pair. The proposed model can be seen as augmenting an existing knowledge graph in a generative way. Although both knowledge graph completion task and our task provide additional entity pairs as their results, they adopt entirely different approaches. The knowledge base completion models rely on the discrepancies among entity pairs of different relationships to distinguish one from another. Otherwise, random negative samples are used for discriminative training. Our model does not rely on discrepancies among relationships: it exploits the commonalities from diverse expressions within each relationship for a rational generation. Knowledge graph completion methods are also vulnerable to low-quality candidate entity pairs during testing: the truly meaningful entity pairs cannot be even obtained when they are not a part of the candidate entity pairs that discriminative models examine from. The choice of candidates involves additional human annotation to improve efficiency; otherwise, any dyadic combinations of medical entities need to be annotated and tested by the model. While the generative nature of our model makes it only generate rational entity pairs by learning from existing ones: no additional data needs to be prepared for generative discovery.

Relationship Extraction: There is another related research area that studies relation extraction [2, 3, 16, 23, 26, 32, 42], which usually amounts to examining whether or not a relation exists between two given entities in a context [10]. Most relationship extraction methods require large amounts of high-quality external information, such as a large text corpus [2, 3, 20, 32] and knowledge graphs [8, 39, 41]. However, in the medical domain, it is tedious and label-intensive to obtain a free-text which contains the co-occurrence of all kinds of relational medical entity pairs. Thus, we propose an effective generative method that learns from the existing medical entity pairs directly. Pre-trained word vectors are used in our model to provide initial entity representations, which do not introduce further labeling cost.

3 PRELIMINARIES

We briefly review preliminaries that relate to the proposed model.

3.1 Autoencoder (AE)

The traditional autoencoder [4] is a multi-layer non-recurrent neural network architecture which has been widely used for unsupervised representation learning. When given an input data x, the autoencoder starts with an encoder net where the input is mapped into a low-dimensional latent variable $z = encoder\_net(x)$ through one or more layers of non-linear transformations, followed by a decoder net where the resulting latent variable z is mapped to an output data $x' = decoder\_net(z)$ which has the same number of units as the input data x, via one or more non-linear hidden layers. The objective of the AE is to minimize the data reconstruction loss:

$$\mathcal{L}_{AE}(x) = \left\|x - x'\right\|^2 = \left\|x - decoder\_net(encoder\_net(x))\right\|^2, \quad (1)$$

and the resulting latent variable z is the low-dimensional latent feature learned from the data x in a totally unsupervised fashion.

3.2 Variational Autoencoder (VAE)

The concept of automatic encoding and decoding makes AE suitable for generative models. Unlike the traditional autoencoder [4] where the hidden variable z has unspecified distributions, the variational autoencoder (VAE) [19] roots in Bayesian inference and inherits the architecture of AE to encode the Bayes automatically for an expressive generation. VAE assumes that the input data x can be encoded into a set of latent variables z with certain distributions, such as multivariate Gaussian distributions. The resulting Gaussian latent variables z are generated by the generative distribution $P_\theta(z)$ and x' is generated with a Bayesian model by a conditional distribution on z: $P_\theta(x'|z)$. VAE infers the latent distribution $P(z)$ using $P_\theta(z|x)$. $P_\theta(z|x)$ can be considered as some mapping from x to z, which is inferred by variational inference as one of the popular Bayesian inference methods. In VAE, $P_\theta(z|x)$ is usually inferred using a simpler distribution $Q_\phi(z|x)$ such as a Gaussian distribution. The objective of VAE is to optimize its variational lower bound:

$$\mathcal{L}_{VAE}(x, y; \theta, \phi) = -KL\left[Q_\phi(z|x)\,||P_\theta(z|x)\right] + \log\left(P_\theta(x)\right), \quad (2)$$

where the first term uses the KL-divergence to minimize the difference between the simple distribution $Q_\phi(z|x)$ and its true distribution $P_\theta(z|x)$, while the second term maximizes the $log\left(P_\theta(x)\right)$.

3.3 Conditional Variational Autoencoder (CVAE)

Although the VAE can generate data that belongs to different types, the latent variable z is only modeled by x in $P_\theta(z|x)$ without knowing the type of it. Thus it cannot generate an output x' that belongs to a particular type y. The conditional variational autoencoder (CVAE) [36] is an extension to VAE that generates x' with conditions. CVAE models both the data x and latent variables z. However, both x and z are conditioned on a class label y:

$$\mathcal{L}_{CVAE}(x, y; \theta, \phi) = -KL\left[Q_\phi(z|x, y)\,||P_\theta(z|x)\right] + \log\left(P_\theta(x|y)\right). \quad (3)$$

In this way, the real latent variable is distributed under $P_\theta(z|y)$ instead of $P_\theta(z)$. With such appealing formulation, we can have a separate $P_\theta(z|y)$ for each class y.

4 CONDITIONAL RELATIONSHIP VARIATIONAL AUTOENCODER

In this section, we introduce the Conditional Relationship Variational Autoencoder (CRVAE) model for the REMEDY problem. The proposed model consists of three modules: encoder, decoder, and generator. The encoder module takes relational medical entity pairs and a relationship indicator as the input, trained to enhance medical entity representations and encode the diversely expressed entity pairs for each medical relationship to a latent space as Q_ϕ. The decoder is jointly trained to reconstruct the entity pairs as P_θ. The generator model shares the same structure with the decoder. However, instead of reconstructing the relational medical entity pair given in the input, it directly samples from the learned latent variable distribution to generate meaningful medical relational entity pairs for a particular relationship. Figure 1 gives an overview of the proposed model.

The model takes a tuple $<e_h, e_t>$ and a relationship indicator r as the input, where e_h and e_t are head and tail medical entity of a relationship r. For example, e_h ="*synovitis*" and e_t="*joint pain*", while the corresponding r is an indicator for $Disease \xrightarrow{Cause} Symptom$.

To effectively represent medical entities, pre-trained word embeddings that embody rich semantic information can be obtained

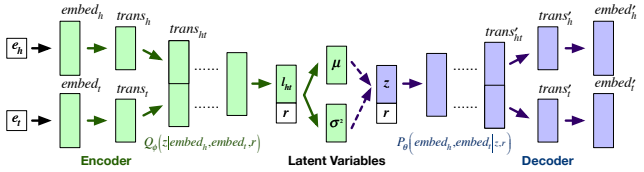

Figure 1: An overview of Conditional Relationship Variational Autoencoder (CRVAE) for Relational Medical Entity-pair Discovery during training. The encoder module is show in green color and the decoder module is show in blue. Model inputs are in white color.

as initial entity representations for e_h and e_t. For simplicity, we adopt 200-dimensional word embeddings pre-trained using Skip-gram [25]. After a table lookup on the pre-trained word vector matrix $W_{embed} \in \mathbb{R}^{V \times D_E}$ where V is the vocabulary size (usually tens of thousands) and D_E is the dimension of the initial entity representation (usually tens or hundreds), $embed_h \in \mathbb{R}^{1 \times D_E}$ and $embed_t \in \mathbb{R}^{1 \times D_E}$ are derived as the initial embedding of medical entities.

4.1 Encoder

With the initial entity representation $embed_h$ and $embed_t$ and their relationship indicator r, the encoder first translates and then maps entity pairs to a latent space as $Q_\phi(z|embed_h, embed_t, r)$.

4.1.1 Translating for Relationship-enhancing. The initial embedding obtained from word embedding reflects semantic and categorical information. However, it is not specifically designed to model the medical relationship among medical entities.

To get entity representations that address relationship information, the encoder learns to translate each medical entity from its initial embedding space to a relationship-enhanced embedding space that distills relational commonalities. For example, a non-linear transformation can be used: $translate(x) = f(x \cdot W_{trans} + b_{trans})$ where f can be an non-linear activation function such as the Exponential Linear Unit (ELU) [9]. $W_{trans} \in \mathbb{R}^{D_E \times D_R}$ is the weight variable and $b_{trans} \in \mathbb{R}^{1 \times D_R}$ is the bias where D_R is the dimension for relationship-enhanced embeddings.

$$trans_h = translate(embed\_h), \tag{4}$$

$$trans_t = translate(embed\_t) \tag{5}$$

are obtained as relationship-enhanced embeddings for e_h and e_t.

4.1.2 Mapping to Latent Variables. The relationship-enhanced entity representation $trans_h$ and $trans_t$ are concatenated

$$trans_{ht} = [trans_h, trans_t] \tag{6}$$

and mapped to the latent space by multiple fully connected layers. For example, we can obtain a variable l_{ht} that addresses the relationship information, as well as entity interactions from two medical entities, by applying six consecutive non-linear fully connected layers on $trans_{ht}$.

As a variational inference model, we assume a simple Gaussian distribution of $Q_\phi(z|embed_h, embed_t, r)$ for the relational medical entity pair $<e_h, e_t>$ with a relationship r. Therefore, for each relational medical entity pair $<e_h, e_t>$ and a relationship indicator r, a mean vector μ and a variance vector σ^2 can be learned as latent variables to model $Q_\phi(z|embed_h, embed_t, r)$:

$$\mu = [l_{ht}, r] \cdot W_\mu + b_\mu, \tag{7}$$

$$\sigma^2 = [l_{ht}, r] \cdot W_\sigma + b_\sigma, \tag{8}$$

where a one-hot indicator $r \in \mathbb{R}^{1 \times |R|}$ is used for the medical relationship r and $|R|$ is the number of all relationships. $W_\mu, W_\sigma \in \mathbb{R}^{\left(D_{l_{ht}} + |R|\right) \times D_L}$ are weight terms and $b_\mu, b_\sigma \in \mathbb{R}^{1 \times D_L}$ are bias terms. D_L is the dimension for latent variables and $D_{l_{ht}}$ is the dimension for l_{ht}. To stabilize the training, we model the variation vector σ^2 by its log form $\log \sigma^2$ (to be explained in Equation (15)).

4.2 Decoder

Once we obtain latent variables μ, σ^2 for an input tuple $<e_h, e_t>$ with an relationship r, the decoder uses latent variables and the relationship indicator r to reconstruct the relational medical entity pair. The decoder implements the $P_\theta(embed_h, embed_t|z, r)$.

Given μ, σ^2, it is intuitive to sample the latent value z from the distribution $N(\mu, \sigma^2)$ directly. However, such operator is not differentiable thus optimization methods failed to calculate its gradient. To solve this problem, a reparameterization trick is introduced in [19] to divert the non-differentiable part out of the network. Instead of directly sampling from $N(\mu, \sigma^2)$, we sample from a standard normal distribution $\epsilon \sim N(0, \mathrm{I})$ and convert it back to z by $z = \mu + \sigma\epsilon$. In this way, sampling from ϵ does not depend on the network.

Similarly as the use of multiple non-linear fully connected layers for the mapping in the encoder, multiple non-linear fully connected layers are used for an inverse mapping in the decoder. After the inverse mapping we obtain $trans'_{ht} \in \mathbb{R}^{1 \times 2D_R}$. The first D_R dimensions of $trans'_{ht}$ are considered as a decoded relationship-enhanced embedding for e_h, while the last D_R dimensions are for e_t:

$$trans'_h = trans'_{ht}[:D_R], \tag{9}$$

$$trans'_t = trans'_{ht}[D_R:], \tag{10}$$

where $trans'_h, trans'_t \in \mathbb{R}^{1 \times D_R}$. $trans'_h$ and $trans'_t$ are further inversely translated back to the initial embedding space $\mathbb{R}^{D_E}$:

$$embed'_h = f(trans'_h \cdot W_{trans\_inv} + b_{trans\_inv}), \tag{11}$$

$$embed'_t = f(trans'_t \cdot W_{trans\_inv} + b_{trans\_inv}), \tag{12}$$

where $embed'_h, embed'_t \in \mathbb{R}^{1 \times D_E}$ are considered as reconstructed representations for $embed_h$ and $embed_t$.

4.3 Training

Inspired by the loss function of CVAE, the loss function of CRVAE is formulated to minimize the variational lower bound:

$$\begin{aligned}&\mathcal{L}_{CRVAE}(embed_h, embed_t, r; \theta, \phi) = \\ &- KL\left[Q_\phi\left(z|embed_h, embed_t, r\right)||P_\theta\left(z|embed_h, embed_t, r\right)\right] \\ &+ \log\left(P_\theta\left(embed_h, embed_t|r\right)\right).\end{aligned} \tag{13}$$

The first term minimizes the KL divergence loss between the unknown true distribution $P_\theta\left(z|embed_h, embed_t, r\right)$ and a simple distribution $Q_\phi\left(z|embed_h, embed_t, r\right)$. The second term models the

entity pairs by $\log\left(P_\theta\left(embed_h, embed_t|r\right)\right)$. The above equation can be reformulated as:

$$\begin{aligned}&\mathcal{L}_{CRVAE}(embed_h, embed_t, r; \theta, \phi) = \\ &- KL\left[Q_\phi\left(z|embed_h, embed_t, r\right)||P_\theta\left(z|r\right)\right] \\ &+ \mathbb{E}\left[\log\left(P_\theta\left(embed_h, embed_t|z, r\right)\right)\right],\end{aligned} \tag{14}$$

where $P_\theta\left(z|r\right)$ describes the true latent distribution z given a certain relationship r and $\mathbb{E}\left[\log\left(P_\theta\left(embed_h, embed_t|z, r\right)\right)\right]$ estimates the maximum likelihood. Since we want to sample from $P_\theta(z|r)$ in the generator, the first term aims to let $Q_\phi(z|embed_h, embed_t, r)$ be as close as possible to $P_\theta(z|r)$ which has a simple distribution $N(0, \mathrm{I})$ so that it is easy to sample from. Furthermore, if $P_\theta(z|r) \sim N(0, \mathrm{I})$ and $Q(z|embed_h, embed_t, r) \sim N(\mu, \sigma^2)$, then a close-form solution for the first term in Equation (14) can be derived as:

$$\begin{aligned}-KL\left[Q_\phi\left(z|embed_h, embed_t, r\right)||P_\theta\left(z|r\right)\right] &= -KL\left[N(\mu, \sigma)||N(0, \mathrm{I})\right] \\ &= -\frac{1}{2}(tr(\sigma^2) + \mu^T\mu - D_L - \log\det(\sigma^2)) \\ &= -\frac{1}{2}\sum_l^{D_L}(\sigma_l^2 + \mu_l^2 - 1 - \log\sigma_l^2),\end{aligned} \tag{15}$$

where l in the subscript indicates the l-th dimension of the vector. Since it is more stable to have exponential term than a log term, $\log\left(\sigma^2\right)$ is modeled as σ^2 which results in the final closed-form of Equation (15):

$$-\frac{1}{2}\sum_l^{D_L}\left(\exp\left(\sigma^2\right)_l + \mu_l^2 - 1 - \sigma_l^2\right). \tag{16}$$

The second term in Equation (14) penalizes the maximum likelihood, which is the conditional probability $P_\theta(embed_h, embed_t|z, r)$ of a certain entity pair $<e_h, e_t>$ given the latent variable z and the relationship indicator r. The mean squared error (MSE) is adopted to calculate the difference between $<embed_h, embed_t>$ and $<embed'_h, embed'_t>$:

$$\begin{aligned}&\mathbb{E}\left[\log\left(P_\theta\left(embed_h, embed_t|z, r\right)\right)\right] = \\ &\frac{1}{2D_E}\left(||embed_h - embed'_h||_2^2 + ||embed_t - embed'_t||_2^2\right),\end{aligned} \tag{17}$$

where $\|\cdot\|_2$ is the vector ℓ_2 norm. To minimize the $\mathcal{L}_{CRVAE}$, existing gradient-based optimizers such as Adadelta [47] can be used. Furthermore, a warm-up technique introduced in [37] can let the training start with deterministic and gradually switch to variational, by multiplying β to the first term. The final loss function used for training is formulated as:

$$\begin{aligned}\mathcal{L}_{CRVAE} &= -\frac{\beta}{2}\sum_l^{D_L}\left(\exp\left(\sigma^2\right)_l + \mu_l^2 - 1 - \log\sigma_l^2\right) \\ &+ \frac{1}{2D_E}\left(||embed_h - embed'_h||_2^2 + ||embed_t - embed'_t||_2^2\right),\end{aligned} \tag{18}$$

where β is initialized as 0 and increase by 0.1 at the end of each training epoch, until it reaches 1.0 as its maximum.

4.4 Generator

When we would like to generate relational medical entity pairs of a specific medical relationship, a density-based sampling method is introduced for the generator to sample $\hat{z}$ from the distribution of latent variables conditioned on that relationship r.

Instead of using the latent variable z provided by certain μ and $\log\sigma^2$ in the encoding process from a certain e_h, e_t and r, the generator tries to sample $\hat{z}$ directly from $P_\theta(\hat{z}|r)$ to get the latent space value $\hat{z}$ for a particular relationship r. Once $\hat{z}$ is obtained, the decoder structure is used to decode the relational medical entity pair. Figure 2 illustrates the generative process. The denser region in the

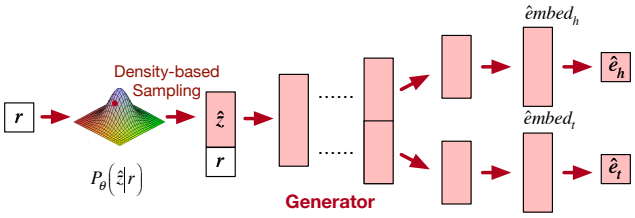

Figure 2: The generator that generate meaningful, novel relational medical entity pairs from the latent space.

latent space $P_\theta(\hat{z}|r)$ indicates that more densely entity pairs are located in the manifold. Therefore, a sampling method that considers the density distribution of $P_\theta(\hat{z}|r)$ samples more often from that region to preserve the true latent space distribution. Specifically, for each relationship r, the density-based sampling samples $\hat{z}$ directly from $P_\theta(\hat{z}|r) \sim N(0, \mathrm{I})$, when trained properly. The resulting vectors $\hat{e}mbed_h$ and $\hat{e}mbed_t$ are mapped back to entity names in natural language, namely $\hat{e}_h$ and $\hat{e}_t$, by finding the nearest neighbor in their initial embedding space $\mathbb{R}^{1\times D_E}$ using W_{embed}. The ℓ-2 distance measure is used for the nearest neighbor search.

Note that the vocabulary of pre-trained word embedding is way more comprehensive than medical entities from labeled entity pairs in training. Using the pre-trained word embedding gives our model the ability to decode unseen medical entities that exist in the vocabulary, but not necessarily in the training data.

5 EXPERIMENTS

5.1 Experiment Settings

5.1.1 Dataset. The dataset consists of 46,018 real-world relational medical entity pairs in Chinese, and it covers six different types of medical relationships, where 70% data are used for training and 30% validation data are used for hyperparameter tuning. Since the proposed model discovers entity pairs by directly sampling from the latent space, not by verifying pre-determined test cases, we evaluate the generated entity pairs directly. Table 2 shows the statistics and representative samples for each medical relationship. We use 200-dimensional word embeddings learned from a Chinese medical corpus on the healthcare forum as the initial entity representation. The vocabulary covers 126,270 words.

5.1.2 Performance Evaluation. Three evaluation metrics are introduced to quantitatively measure the generated relational medical entity pairs: quality, support, and novelty.

Quality Since it is hard for the machine to evaluate whether a relational medical entity pair is meaningful or not, human annotation is involved in assessing the quality of the generated relational

Medical Relationship	Count	Relational Medical Entity Pairs
Disease $\xrightarrow{Cause}$ Body Part	2320	<*tricuspid insufficiency* (三尖瓣闭锁), *tricuspid valve* (三尖瓣)> <*vaginal cancer* (阴道癌), *reproductive system* (生殖)> <*hydrocephaly* (脑积水), *head* (头部)>
Disease $\xrightarrow{RelatedTo}$ Disease	4614	<*infant hydrocephalus* (婴儿脑积水), *congenital hydrocephalus* (先天性脑积水)> <*urethritis* (尿道炎), *cystitis* (膀胱炎)> <*retention of food in the stomach* (食滞胃脘), *infantile indigestion* (小儿消化不良)>
Disease $\xrightarrow{Need}$ Examine	4185	<*salicylates poisoning* (水杨酸类中毒), *routine urianlysis* (尿常规)> <*tetralogy triad* (法洛三联症), *electrocardiogram, ECG* (心电图)> <*epididymitis* (附睾炎) , *cremasteric reflex* (提睾反射)>
Symptom $\xrightarrow{BelongTo}$ Department	8595	<*anchylosis, stiffness of a joint* (关节强直), *orthopedics* (骨科)> <*female lower abdominal pain* (女性小腹疼痛), *gynecology* (妇科)> <*absent infant sucking reflex* (吸吮反射消失), *neonatology* (新生儿科)>
Disease $\xrightarrow{Cause}$ Symptom	16642	<*peritonitis* (腹膜炎), *abdominal venous engorgement* (腹部静脉怒张)> <*urethritis* (尿道炎), *urethra itching* (尿道痒感)> <*radial nerve palsy* (桡神经麻痹), *upper extremity weakness* (上肢无力)>
Symptom $\xrightarrow{RelatedTo}$ Symptom	9662	<*redness and swelling around the umbilicus* (脐周红肿), *periumbilical swelling* (脐周肿胀)> <*muscular contusion* (肌肉挫伤), *disinsertion* (肌腱断裂)> <*fingers benumbed with cold* (手指冻肿), *skin frostbite* (皮肤冻伤)>

Table 2: Sample Medical Relationships and relational medical entity pairs.

medical entity pairs. We deploy a human annotation task on Amazon Mechanical Turk. Annotators need to pass at least four in five sample cases to qualify the annotation. Majority voting of three annotators is adopted. The quality is measured by:

$$quality = \frac{\text{\# of entity pairs that are meaningful}}{\text{\# of all the generated entity pairs}}. \quad (19)$$

Support Besides human annotations, a support score quantitatively measures the belongingness of an entity pair generated by a specific relationship to existing entity pairs with that relationship. For each generated relational medical entity pair $<\hat{e}_h, \hat{e}_t>$, the support score measures its similarities to known entity pairs of each relationship r_c:

$$support_{<\hat{e}_h,\hat{e}_t,r_c>} = \frac{1}{1 + distance(\hat{embed}_h, \hat{embed}_t, r_c)}, \quad (20)$$

where $distance(\hat{embed}_h, \hat{embed}_t, r_c)$ calculates the distance between the vector $\hat{embed}_h - \hat{embed}_t$ and $NN_{r_c}(\hat{embed}_h - \hat{embed}_t)$ using distance measure such as cosine distance. The NN_{r_c} implements the nearest neighbor search over the $embed_h - embed_t$ space among all the entity pairs having the relationship r_c. For each generated medical entity pair, the support scores of all relationships are normalized:

$$norm\_support_{<\hat{e}_h,\hat{e}_t,r_c>} = \frac{support_{<\hat{e}_h,\hat{e}_t,r_c>}}{\sum_{r_i}^{|R|} support_{<\hat{e}_h,\hat{e}_t,r_i>}}. \quad (21)$$

The generated entity pair $<\hat{e}_h, \hat{e}_t>$ finds support from its estimated relationship which has the highest score, while the relationship r given during the generating process is considered as the ground truth for $<\hat{e}_h, \hat{e}_t>$. The final support value is based on the accuracy of the estimated relationship and the ground truth relationship.

Novelty The ability to generate novel relational medical entity pairs is one of our key contributions. Due to different scope of medical knowledge among individuals, human annotators are not able to precisely evaluate the novelty. We measure the novelty of the generation process by:

$$novelty = \frac{\text{\# of entity pairs that do not exist in the dataset}}{\text{\# of all the generated entity pairs}}. \quad (22)$$

5.1.3 Baselines. Considering that no known methods are currently available for the REMEDY problem, and we consider it unfair to compare with discriminative methods which have external resources, or a test set that is prepared with additional human knowledge, the performance on the following models are compared:

- CRVAE-MONO: The proposed model that works with all entity pairs having the same medical relationship in both training and generation. For each relationship, we train a separate CRVAE with entity pairs having that relationship.
- RVAE: The unconditional version of the model CRVAE where the relationship indicator r is not provided during model training and generation.
- CRVAE-RAND: The proposed model CRVAE with a random sampling based generator. Rather than using the density-based sampling strategy, the generator of CRVAE-RAND samples randomly from the latent space.
- CRVAE: The proposed method where relational medical entity pairs with all types of relationships are used together to train the model. The training is conditioned on relationships, and density-based sampling is used.
- CRVAE-WA: The proposed method with the warm-up strategy introduced in Section 4.3.

5.2 Experiment Results

We generate 1000 entity pairs for each medical relationship for evaluation. Table 3 summarizes the performance of the proposed method when comparing with other alternatives. In summary, CRVAE-MONO demonstrates the power of generative model that

Model Name	Quality	Support	Novelty	Loss (Train / Valid)
CRVAE-MONO	0.6698	0.9550	0.5118	47.3002 / 116.6739
CRVAE-RAND	0.2550	0.3764	0.9952	43.0954 / 83.6589
CRVAE	0.7308	0.9048	0.5682	43.0954 / 83.6589
CRVAE-WA	0.7717	0.9291	0.6193	33.4399 / 57.9470

Table 3: Performance comparison results.

learns commonalities purely from the diversely expressed entity pairs without substantial data requirements. By comparing CRVAE-RAND and CRVAE we show the effectiveness of the density-based sampling in generating high-quality entity pairs. The warm up technique adopted in CRVAE-WA is able to give CRVAE a further performance boost. As a qualitative measure, we also provide relational medical entity pairs generated by the proposed model in Table 4, from which we can see the meaningful and novel structured knowledge discovered in a generative fashion.

Disease $\xrightarrow{Cause}$ Body Part
<*dysentery* (痢疾), *intestine* (肠)>
<*brain tumor* (脑瘤), *head* (头部)>
<*leukopenia* (白细胞减少症), *vascular system* (血液)>
Disease $\xrightarrow{RelatedTo}$ Disease
<*foreign body in esophagus* (食管异物), *bowel obstruction* (肠梗阻)>
<*brain contusion* (脑挫裂伤), *amnesia* (记忆障碍)>
<*respiratory acidosis* (呼吸性酸中毒), *pulmonary edema* (肺水肿)>
Disease $\xrightarrow{Need}$ Examine
<*uremia* (尿毒症), *routine urianlysis* (尿常规)>
<*bacterial meningitis* (细菌性脑膜炎), *cranial CT* (头颅CT)>
<*bowel obstruction* (肠梗阻), *abdominal x-ray* (腹部平片)>
Symptom $\xrightarrow{BelongTo}$ Department
<*retained placenta* (胎盘滞留), *obstetrics* (产科)>
<*fluid retention* (水潴留), *nephrology* (肾内科)>
<*stuffy nose* (鼻塞), *otolaryngology* (耳鼻咽喉科)>
Disease $\xrightarrow{Cause}$ Symptom
<*otogenic brain abscess* (耳源性脑脓肿), *earache* (耳痛)>
<*neuritis* (神经炎), *numbness in the hands* (手麻)>
<*open head injury* (开放性颅脑损伤), *loss of consciousness* (意识模糊)>
Symptom $\xrightarrow{RelatedTo}$ Symptom
<*fatigue* (乏力), *feel wobbly and rough* (四肢无力)>
<*joint pain* (关节痛), *limited joint mobility* (关节活动受限)>
<*blurred vision* (雾视), *eye discomfort* (眼睛不舒服)>

Table 4: Novel and meaningful relational medical entity pairs generated by the proposed method.

5.3 Generative Modeling Capacity

Unlike discriminative models which utilize the discrepancies among instances of different classes to discriminate one class from another, the generative nature of the proposed method makes it generate entity pairs only when it fully understands the diverse expressions within each medical relationship. To validate such appealing property, we introduce the baseline CRVAE-MONO which works with all entity pairs having the same medical relationship in both training and generation.

Table 5 compares the fine-grained quality, support and novelty of the generated entity pairs of CRVAE-MONO and CRVAE on each relationship. The CRVAE-MONO on each relationship achieves a reasonable performance, which shows that the generative modeling has the ability to learn directly from the existing medical entity pairs without additional data requirement. Furthermore, when all types of entity pairs are trained altogether in CRVAE, we observe a consistent improvement in not only quality but also novelty.

CRVAE-MONO	Quality	Support	Novelty	Loss (Train/Valid)
Disease $\xrightarrow{Cause}$ Body Part	0.6830	1.0000	0.4880	54.9830 / 126.7426
Disease $\xrightarrow{RelatedTo}$ Disease	0.6890	0.8700	0.4830	51.5131 / 155.0721
Disease $\xrightarrow{Need}$ Examine	0.7080	1.0000	0.5210	54.7635 / 136.4802
Symptom $\xrightarrow{BelongTo}$ Department	0.6870	1.0000	0.4660	39.0959 / 72.5872
Disease $\xrightarrow{Cause}$ Symptom	0.5870	0.9400	0.5730	37.3276 / 83.8797
Symptom $\xrightarrow{RelatedTo}$ Symptom	0.6650	0.9200	0.5400	46.1180 / 125.2818
CRVAE				
Disease $\xrightarrow{Cause}$ Body Part	0.7560	0.9990	0.7240	
Disease $\xrightarrow{RelatedTo}$ Disease	0.6910	0.7440	0.8670	
Disease $\xrightarrow{Need}$ Examine	0.7570	0.9810	0.8710	43.0954 / 83.6589
Symptom $\xrightarrow{BelongTo}$ Department	0.7680	0.9950	0.6130	
Disease $\xrightarrow{Cause}$ Symptom	0.7020	0.8820	0.9270	
Symptom $\xrightarrow{RelatedTo}$ Symptom	0.7110	0.8280	0.8880	

Table 5: Quality, support and novelty metrics of the generated relational medical entity pairs by CRVAE-MONO and CRVAE.

5.4 Effectiveness of Density-based Sampling

To validate the effectiveness of the density-based sampling for the generator, we compare the proposed method with CRVAE-RAND where a random sampling strategy is adopted. From Table 3 we can see that when the distribution of the latent space is not considered, the random sampling strategy in CRVAE-RAND tends to generate more entity pairs that are not seen in the existing dataset. However, the generated entity pairs are of low quality and support.

CRVAE adopts a density-based sampling. The dense region in the latent space indicates that more entity pairs are located. Therefore, in CRVAE, the quality and support of the generated entity pairs benefit from sampling more often at denser regions in the latent space, resulting in less novel but higher quality entity pairs.

5.5 Ability to Infer Conditionally

To effectively discover structured medical knowledge, one of our key contributions is to generate relational medical entity pairs for a specific relationship. That is, the ability to infer new entity pairs for a particular relationship without additional data preparation. Besides seamlessly incorporating this idea in the model design, we also show such conditional inference ability by visualization.

Figure 3 shows the validation samples after being mapped into the μ space using RVAE (left) and CRVAE (right), respectively. The samples are colored based on their ground truth relationship indicators. The left figure indicates that when the relationship indicator r is not given during the training/validation, RVAE is still able to map different relationships into various regions in the latent space, while a single distribution models all types of relationships. Such property is appealing for an unsupervised model, but since the relationship indicator r is not given during training, RVAE fails to generate entity pairs having a particular relationship, unless we manually assign a boundary for each relationship in the latent space. The

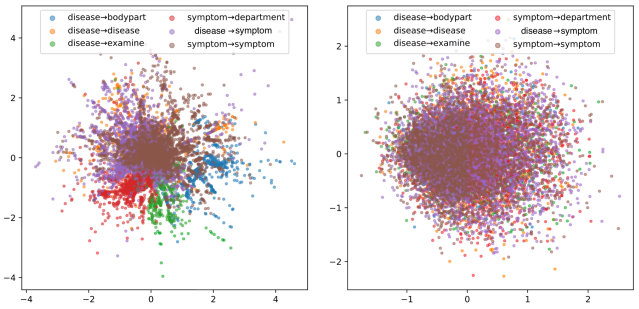

Figure 3: The latent variable μ of RVAE (left) and CRVAE (right) on the validation data, presented in a two-dimensional space after dimension reduction using Primary Component Analysis.

right figure shows that when the relationship indicator r is incorporated during the training, CRVAE learns to let each relationship have a unified latent representation $P_\theta(\hat{z}|r)$. A separate but nearly identical distribution is used to model each medical relationship. Such property may enable the generator of our model to sample the expression variations from a relationship-independent latent space, while the relationship indicator r provides the categorical information regarding what type of medical relationship should the expression variation applies on.

5.6 Relationship-enhancing Entity Adjustment

To show the effectiveness of relationship-enhancement, Table 6 shows the nearest neighbors of a disease entity *genital tract malformation* (生殖道畸形) and a symptom entity *muscle strain* (肌肉拉伤) in their original embedding space, as well as in the space after relationship-enhancing.

From these cases we can see that the original entity representations trained with Skip-gram [25] tend to put entities in proximity when they are mentioned in similar contexts. In the first case, the entity *genital tract malformation* (生殖道畸形) is in close proximity to *infertility* (不孕) and *acyesis* (不孕症). In the second case, entities that have similar context like *pull-up* (引体向上) and *amount of exercise* (运动量) are found near by the entity *muscle strain* (肌肉拉伤).

• *genital tract malformation* (生殖道畸形)	
NN in the relationship-enhanced space $\mathbb{R}^{1\times D_R}$	NN in the initial embedding space $\mathbb{R}^{1\times D_E}$
genital tract (生殖道)	*reproductive system* (生殖系统)
reproductive system (生殖系统)	*reproductive tract tumors* (生殖道肿瘤)
heart malformations (心脏畸形)	*urinary system malformations* (泌尿系畸形)
chromosome abnormalities (染色体异常)	***infertility*** (不孕)
reproductive tract tumors (生殖道肿瘤)	*vaginal atresia* (阴道闭锁)
generative organs (生殖器官)	*genital tract* (生殖道)
urinary system malformations (泌尿系畸形)	*generative organs* (生殖器官)
gastrointestinal malformations (消化道畸形)	***acyesis*** (不孕症)
• *muscle strain* (肌肉拉伤)	
NN in the relationship-enhanced space $\mathbb{R}^{1\times D_R}$	NN in the initial embedding space $\mathbb{R}^{1\times D_E}$
strain (拉伤)	拉伤 (*strain*)
ligament strain (韧带拉伤)	*muscle tear* (肌肉撕裂)
sprain (扭伤)	***pull-up*** (引体向上)
foot pain (足痛)	*sprain* (扭伤)
muscle tear (肌肉撕裂)	*muscle fatigue* (肌肉疲劳)
plantar fasciitis (足底筋膜炎)	*tenosynovitis* (腱鞘炎)
joint sprain (关节扭伤)	*tendonitis* (肌腱炎)
repetitive strain injury, RSI (劳损)	***amount of exercise*** (运动量)

Table 6: The effectiveness of relationship-enhancing adjustment on entity representations.

The translation layer adjusts the original entity representation so that they are more suitable for structured medical knowledge discovery. The nearest neighbors in the adjusted space are not necessarily entities that co-occur in the same context, but more relation-wise similar with the given entity. For example, *heart malformations* (心脏畸形) and *chromosome abnormalities* (染色体异常) may not be semantically similar with the given word *genital tract malformation* (生殖道畸形), but they may serve similar functionalities in a Disease $\xrightarrow{Cause}$ Symptom relationship.

5.7 Hyperparameter Analysis

We train the proposed model with a wide range of hyperparameter configurations, which are listed in Table 7. We vary the batch size from 64 to 256. The dimension D_R for translating the initial entity embeddings is set from 64 to 2048. We try two to seven hidden layers from $trans_{ht}$ to l_{ht} and from $[z, r]$ to $trans'_{ht}$, with different non-linear activation functions. For each hidden layer, the hidden unit number D_H is set from 2 to 1024. The latent dimension D_L is set from 2 to 200.

Parameter	Value
Batch Size	64, 128, 256
D_R	64, 128, 256, 512, 640, 768, 1024, 1280, 1536, 1792, 2048
D_H	2, 4, 8, 16, 32, 64, 128, 256, 512, 640, 768, 1024
D_L	2, 3, 4, 5, 10, 20, 50, 100, 200
Activation	ELU [9], ReLU [27], Sigmoid, Tanh
Optimizer	Adadelta [47], Adagrad [11], Adam [17], RMSProp [40]

Table 7: Hyperparameter configurations.

The top-5 hyperparameter settings with low validation losses are shown in Table 8. Among the combinations of hyperparameter configurations, we find that for fully connected hidden layers from $trans_{ht}$ to l_{ht}, a sequence of six consecutive layers: 1792 ·640·640·512·256·64 works the best for the encoder with ELU as the activation function. For $[z, r]$ to $trans'_{ht}$ in the decoder, such layer setting is organized in a reverse order. A batch size of 64 and the Adadelta optimizer work the best for our task. $D_R = 640$ is used. The latent dimension $D_L = 200$ is adopted for μ and σ^2. We use Xavier initialization [14] for weight variables and zeros for biases. Such configuration achieves a training loss of 43.0954 and a validation loss of 83.6589.

Batch	D_R	$\{D_H\}$	D_L	Act.	Optimizer	Loss(Training /Valid)
64	640	1792·640·640·512·256· 64	200	ELU	Adadelta	43.0954 / 83.6589
64	640	1792·256·640·512·256·128	200	ELU	Adadelta	51.0695 / 86.9153
64	640	1792·256·640·512·256· 64	200	ELU	Adadelta	50.4392 / 88.6438
128	640	1792·640·768·512· 64·128	50	ELU	Adadelta	50.5997 / 89.0125
256	640	512·768·640·256·512	50	ELU	Adam	62.1955 / 89.2014

Table 8: Hyperparameter analysis on the proposed model. Only the top-5 best configurations are shown.

6 CONCLUSION AND FUTURE WORKS

To efficiently expand the scale of high-quality structured medical knowledge while minimizing the effort in date preparation, we introduce a generative perspective to the Relational Medical Entity-pair Discovery (REMEDY) problem. A novel model named Conditional Relationship Variational Autoencoder (CRVAE) is introduced to exploit the generative modeling ability for efficient discovery of

relational medical entity pairs. Unlike traditional discriminative methods which require substantial data as external knowledge, our model purely learns from the commonalities of the existing medical entity pairs by their diverse expressions. It is able to generate meaningful, novel entity pairs of a specific medical relationship by directly sampling from the learned latent space without the requirement of additional context information. The performance of the proposed method is evaluated on real-world medical data both quantitatively and qualitatively. For future works, we would like to extend this framework to more general cases where entity pairs of open-domain knowledge with various granularity are modeled altogether.

7 ACKNOWLEDGMENTS

This work is supported in part by NSF through grants IIS-1526499, IIS-1763325, CNS-1626432, and NSFC 61672313.

REFERENCES

[1] Asma Ben Abacha and Pierre Zweigenbaum. 2011. Automatic extraction of semantic relations between medical entities: a rule based approach. *Journal of biomedical semantics* (2011).
[2] Eugene Agichtein and Luis Gravano. 2000. Snowball: Extracting relations from large plain-text collections. In *DL*.
[3] Ricardo Baeza-Yates and Alessandro Tiberi. 2007. Extracting semantic relations from query logs. In *KDD*.
[4] Yoshua Bengio et al. 2009. Learning deep architectures for AI. *Foundations and trends® in Machine Learning* (2009).
[5] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*.
[6] Antoine Bordes, Jason Weston, Ronan Collobert, Yoshua Bengio, et al. 2011. Learning Structured Embeddings of Knowledge Bases.. In *AAAI*.
[7] Samuel R Bowman, Luke Vilnis, Oriol Vinyals, Andrew M Dai, Rafal Jozefowicz, and Samy Bengio. 2016. Generating Sentences from a Continuous Space. *CoNLL* (2016).
[8] Kai-Wei Chang, Scott Wen-tau Yih, Bishan Yang, and Chris Meek. 2014. Typed tensor decomposition of knowledge bases for relation extraction. (2014).
[9] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. 2015. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289* (2015).
[10] Aron Culotta, Andrew McCallum, and Jonathan Betz. 2006. Integrating probabilistic extraction models and data mining to discover relations and patterns in text. In *NAACL-HLT*.
[11] John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *JMLR* (2011).
[12] Susannah Fox and Maeve Duggan. 2013. Health online 2013. *Washington, DC: Pew Internet & American Life Project* (2013).
[13] Matt Gardner and Tom Mitchell. 2015. Efficient and expressive knowledge base completion using subgraph feature extraction. In *EMNLP*.
[14] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*.
[15] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Rezende, and Daan Wierstra. 2015. DRAW: A Recurrent Neural Network For Image Generation. In *ICML*.
[16] Meng Jiang, Jingbo Shang, Taylor Cassidy, Xiang Ren, Lance M Kaplan, Timothy P Hanratty, and Jiawei Han. 2017. MetaPAD: Meta Pattern Discovery from Massive Text Corpora. In *KDD*.
[17] Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[18] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
[19] Diederik P Kingma and Max Welling. 2014. Stochastic gradient VB and the variational auto-encoder. In *ICLR*.
[20] Yaliang Li, Chaochun Liu, Nan Du, Wei Fan, Qi Li, Jing Gao, Chenwei Zhang, and Hao Wu. 2016. Extracting medical knowledge from crowdsourced question answering website. *IEEE Transactions on Big Data* (2016).
[21] Cindy Xide Lin, Bo Zhao, Tim Weninger, Jiawei Han, and Bing Liu. 2010. Entity relation discovery from web tables and links. In *WWW*.
[22] Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2016. Neural Relation Extraction with Selective Attention over Instances.. In *ACL*.
[23] Liyuan Liu, Xiang Ren, Qi Zhu, Huan Gui, Shi Zhi, Heng Ji, and Jiawei Han. 2017. Heterogeneous Supervision for Relation Extraction: A Representation Learning Approach. In *EMNLP*.
[24] Diego Marcheggiani and Ivan Titov. 2016. Discrete-state variational autoencoders for joint discovery and factorization of relations. *TACL* (2016).
[25] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
[26] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *ACL*.
[27] Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *ICML*.
[28] Jennifer Neville, Özgür Şimşek, David Jensen, John Komoroske, Kelly Palmer, and Henry Goldberg. 2005. Using relational knowledge discovery to prevent securities fraud. In *KDD*.
[29] Mike Oaksford and Nick Chater. 2007. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.
[30] Yunchen Pu, Zhe Gan, Ricardo Henao, Xin Yuan, Chunyuan Li, Andrew Stevens, and Lawrence Carin. 2016. Variational autoencoder for deep learning of images, labels and captions. In *NIPS*.
[31] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015).
[32] Saurav Sahay, Sougata Mukherjea, Eugene Agichtein, Ernest V Garcia, Shamkant B Navathe, and Ashwin Ram. 2008. Discovering semantic biomedical relations utilizing the web. *TKDD* (2008).
[33] Adam Santoro, David Raposo, David GT Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. 2017. A simple neural network module for relational reasoning. *arXiv preprint arXiv:1706.01427* (2017).
[34] Ying Shen, Yang Deng, Min Yang, Yaliang Li, Nan Du, Wei Fan, and Kai Lei. 2018. Knowledge-aware attentive neural network for ranking question answer pairs. *SIGIR* (2018).
[35] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *NIPS*.
[36] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. 2015. Learning structured output representation using deep conditional generative models. In *NIPS*.
[37] Casper Kaae Sønderby and COM Tapani Raiko. 2016. How to Train Deep Variational Autoencoders and Probabilistic Ladder Networks. In *ICML*.
[38] Yizhou Sun, Jiawei Han, Charu C Aggarwal, and Nitesh V Chawla. 2012. When will it happen?: relationship prediction in heterogeneous information networks. In *WSDM*.
[39] Zareen Syed, Evelyne Viegas, and Savas Parastatidis. 2010. Automatic Discovery of Semantic Relations using MindNet.. In *LREC*.
[40] Tijmen Tieleman and Geoffrey Hinton. 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning* (2012).
[41] Patrick Verga, Arvind Neelakantan, and Andrew McCallum. 2016. Generalizing to unseen entities and entity pairs with row-less universal schema. *EACL* (2016).
[42] Chenguang Wang, Yangqiu Song, Dan Roth, Chi Wang, Jiawei Han, Heng Ji, and Ming Zhang. 2015. Constrained Information-Theoretic Tripartite Graph Clustering to Identify Semantically Similar Relations.. In *IJCAI*.
[43] Quan Wang, Bin Wang, and Li Guo. 2015. Knowledge Base Completion Using Embeddings and Rules.. In *IJCAI*.
[44] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph and Text Jointly Embedding.. In *EMNLP*.
[45] Weidi Xu, Haoze Sun, Chao Deng, and Ying Tan. 2017. Variational Autoencoder for Semi-Supervised Text Classification.. In *AAAI*.
[46] Limin Yao, Aria Haghighi, Sebastian Riedel, and Andrew McCallum. 2011. Structured relation discovery using generative models. In *EMNLP*.
[47] Matthew D Zeiler. 2012. ADADELTA: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701* (2012).
[48] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, Jun Zhao, et al. 2014. Relation Classification via Convolutional Deep Neural Network.. In *COLING*.
[49] Chenwei Zhang, Nan Du, Wei Fan, Yaliang Li, Chun-Ta Lu, and Philip S. Yu. 2017. Bringing semantic structures to user intent detection in online medical queries. In *IEEE Big Data*.
[50] Chenwei Zhang, Wei Fan, Nan Du, and Philip S Yu. 2016. Mining user intentions from medical queries: A neural network based heterogeneous jointly modeling approach. In *WWW*.