A Generative Adversarial Strategy for Modeling Relation Paths in Knowledge Base Representation Learning

Anonymous Author(s)
Affiliation
Address
email

Abstract

Enabling neural networks to perform multi-hop (mh) reasoning over knowledge bases (KBs) is vital for tasks such as question-answering and query expansion. Typically, recurrent neural networks (RNNs) trained with explicit objectives are used to model mh relation paths (mh-RPs). In this work, we hypothesize that explicit objectives are not the most effective strategy effective for learning mh-RNN reasoning models, proposing instead a generative adversarial network (GAN) based approach. The proposed model – mh Relation GAN (mh-RGAN) – consists of two networks; a generator G, and discriminator D. G is tasked with composing a mh-RP and D with discriminating between real and fake paths. During training, G and D contest each other adversarially as follows: G attempts to fool D by composing an indistinguishably invalid mh-RP given a head entity and a relation, while D attempts to discriminate between valid and invalid reasoning chains until convergence. The resulting model is tested on benchmarks WordNet and FreeBase datasets and evaluated on the link prediction task using MRR and HIT@ 10, achieving best-in-class performance in all cases.

1 Introduction

There is an increasing interest in enabling neural networks (NNs) to accomplish the type of complex reasoning previously only performed by logical and symbolic reasoning systems. Equipped with this capability, NNs become able to reason over large-scale knowledge bases (KBs) such as WordNet [1], Freebase [2], which is essential for natural language understanding tasks such as question answering. NN-based reasoning is currently most successfully performed by modeling KBs via the embedding approach [3] in which KB elements (entities and relations) are embedded into vector spaces of appropriate dimensionality. The dimensionality of the vector space is controlled by forcing the model to generalize to novel facts of the kind generated by iterative symbolic reasoning. Nonetheless, most existing methods use only a one-hop reasoning path to learn the models [3, 4, 5] and ignore mh-RP, which is crucial for capturing complex inference patterns such as, e.g., inference of a person's nationality given her city of birth (the city belonging to a state, and the state belonging to a country). Typically, mh-RPs are modeled via recurrent neural networks (RNNs) trained with explicit objectives to progressively utilize entities and relations in order to produce a novel entity at each step [6, 7, 8]. The models are trained with objectives that either rely on hidden negative samples [7] or perform step-wise optimization of a loss function [8]. While the former can lead to the learning of unsound models due to its reliance on unrealistic negative samples, the later, although it seeks to generate the best candidate entity at each step, may yet introduce a bad component with respect to the logical chain over the long run. For example, consider two mh-RPs *(Donald Trump, child, Ivanka Trump, mother, Ivana Trump)* and *(Donald Trump, child, Barron Trump, mother, Melania Trump)*. Given

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

(Donald Trump, child) as input, the output *Ivanka Trump* is a good single step decision but a bad decision if we are reasoning generally about the spouse of Donald Trump.

In this work, we address the above mentioned issues in the training objective of learning mh relation model (mh-RM). We take inspiration from generative adversarial networks (GANs) to propose the direct training of mh-RM for producing a relation path which is indistinguishable from real paths. The proposed GAN model consists of two networks; a mh-RP generator G, and a mh-RP discriminator D. We enable G to compose a mh-RP and D to discriminate between real and fake paths. During training, G and D contest each other as follows: G attempts to fool D by composing an indistinguishable invalid mh-RP given a head entity and a relation, and D attempts to discriminate between valid and invalid relation paths. This contention drives both to improve their learning as training progresses, until the convergence point in which generator paths are indistinguishable from true reasoning paths and the discriminator is effective at distinguishing between valid and invalid relation paths.

Related work on mh path modeling approaches and GANs is described in Section 2. The proposed method (mh-RGAN) is presented in Section 3. Experiments and results are reported in Section 4. Finally, the paper is concluded in Section 5.

2 Related Work

2.1 Multi-hop Reasoning over KBs

In this study, we focus on the modeling of mh-RP for reasoning over KBs. In this context, the majority of extant work proposes to model the relation path $p = (r_1, \ldots, r_t)$ by using addition as $(p = r_1 + \cdots + r_t)$ or multiplication as $(p = r_1 \ldots r_t)$ [9]: Neelakantan et al. (2015) use RNNs to model the path as $p_i = RNN(p_{i-1}, r_i)$, where p_i is cumulative path information up to point i [6]; Das et al. (2017) enable RNNs to invoke within-path entities while modeling the paths [7] (however, in this case, the RNN takes a fixed set of entities as input); Yin et al. (2018) equip RNNs with the capability of predicting the entity as output and use this prediction to update the path, rather than using fixed set of entities as in [8].

In the approach to be outlined, although we shall follow the broad modeling approach of [8], rather than training RNN on single-hop paths, we will train the RNN on mh-RPs by using GANs. This development restricts the likelihood of the RNN generating candidate entities that are optimal at a given instant but sub-optimal with respect to the path as a whole. It also enables the RNN to produce natural or human-like reasoning paths. Relevant to our study is the research area concerned with learning rules from KBs. However, contrary to the modeling of reasoning pathways, rule learning seeks to emulate an inference procedure. Commonly, these methods represent facts using vectors or tensors and employ RNNs, such as memory networks, to model transitivity relations between facts in order to perform inferences [10, 11]; other studies approach the learning of representations of entities and relations via a transitivity structure [12]. Neuro-symbolic computing is another relevant domain dealing with integrating and reasoning over symbolic knowledge utilizing neural networks; some notable recent efforts include [13, 14].

2.2 Generative Adversarial Networks

Primarily, GANs were conceived for producing data samples from a continuous space such as images [15]. In the original setting, the generator is used to generate an image from random noise and discriminator is employed to classify between real and generated (a.k.a. fake) images [16]. Latterly, it has been demonstrated that GANs can also produce images conditioned on specific inputs [17]. However, an undesirable drawback of the GAN approach, in its original form, is that it cannot generate discrete samples such as the entity of a relation, since discrete sampling restricts gradients from being transmitting back to the generator [18]. To address this problem, one solution is to use reinforcement learning (RL) to learn a generator with policy gradient [18]. This approach, using single-step policy gradient, is employed in [19] for generating negative samples for training KB completion models. A similar approach is followed for graph representation learning where the generator is used to predict edges between vertices and the discriminator is employed to distinguish well-connected vertex pairs from ill-connected pairs [20].

3 The Method

A broad outline of the proposed mh-RGAN architecture is given in Figure 1 ; given a dataset of N relation paths, $P(h,t) = \{p^{(1)}, \ldots, p^{(N)}\}$, connecting pairs of entities h and t, we wish to train a generator G to compose a mh-RP $p_{1:L} = (h_1, r_1, \ldots, t_L)$ of length L, and use a discriminator D to process the generated paths and provide supervision to the generator. In the typical GAN setting, the generator cannot predict discrete samples. To deal with this issue, we adopt a policy gradient based reinforcement learning (RL) method for training G alongside D based on the approach of [18].

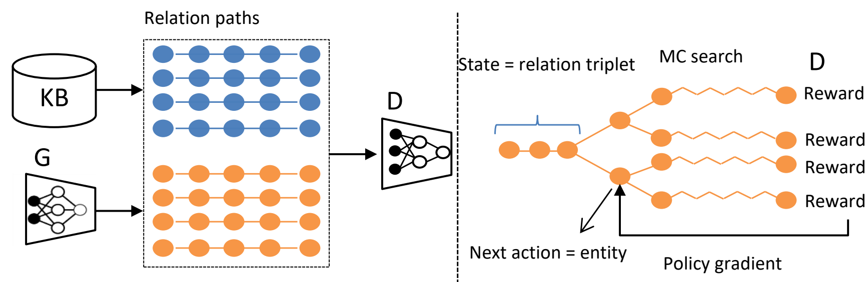

Figure 1: Proposed mh-RGAN architecture. **Left:** D is trained over the real and generated relation paths. **Right:** G is trained via policy gradient for which the final reward signal is provided by D and is passed back to the intermediate action value via Monte Carlo search.

To model G, we require a method that can sequentially compose a reasoning path. Given the demonstrable efficacy of RNNs in this domain previously alluded to, we shall employ an RNN to map the input sequence of relations $(r_1, r_2, \ldots, r_L)$ along with a head entity h_1 to an output reasoning path: $(h_1, r_1, \ldots, t_L)$. Specifically, the RNN greedily selects an entity and relation at each time step and produces an output entity; relations are kept within the input space and entities are embedded in the hidden latent space. The broad modeling approach is depicted in Figure 2 . This model is realized by modifying the RNN's recursive function as follows:

$$\hat{v}_{e_l} = f(W[\hat{v}_{e_{l-1}}; v_{r_l}]) \tag{1}$$

$$v_{e_l} = Softmax(\hat{v}_{e_l}) \tag{2}$$

where v_{e_l} and v_{r_l} are vector representations of entity and relation at respective positions l. To initialize the model, we set $\hat{v}_{e_0} = v_h$.

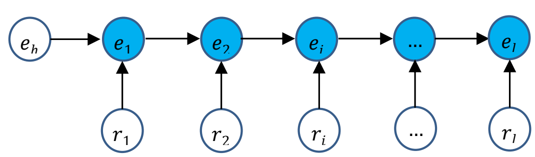

Figure 2: Generator modeling approach (latent states in blue, entities denoted e_i, relations, r_i)

In order to model D, we require a method for classifying relation paths. We select a Convolution Neural Network (CNN) given the performance of CNN in text classification. To apply the CNN, we extend the single-hop path representation approach proposed in [21] in order to represent mh-RP $p_{1:L} = (h_1, r_2, \ldots, t_L)$ as:

$$\varepsilon_{1:T} = (h_1 \oplus r_2 \oplus \ldots \oplus t_L) \tag{3}$$

where $x_i \in \mathbb{R}^k$ is a k-dimensional token (i.e. an entity or relation) and $\oplus$ is a concatenation operator applied in building a matrix $\varepsilon_{1:T}$. The convolution is performed by applying a filter $\omega \in \mathbb{R}^{l \times k}$ to a window of l tokens in order to produce a feature map v^i as:

$$v_i = g(\omega * \varepsilon_{i:i+l-1} + b) \tag{4}$$

The convolution is then followed by max-over-time pooling over feature maps as:

$$\hat{v} = max\{v_1, v_2, \dots v_{T-l+1}\} \tag{5}$$

The pooling layer is connected to a fully connected (FC) layer and finally to a sigmoid unit to produce the probability that the relation path is real.

4 Experiments and Results

4.1 Datasets and Evaluation Protocol

We employ two benchmark 'common sense reasoning' datasets to evaluate the mh-RGAN in common with the baseline studies [7, 8]. Wordnet (WN18) [1] is a collection of pairs of English dictionary and thesaurus words that are linked via relations such as: *subclass_of, type_of, part_of, has_part*, etc. Freebase (FB15k) [2] consists of relation triplets from the personal ID domain and which includes relations such as: *gender, nationality, profession, place of birth, location, religion, parents, children, ethnicity, spouse*, etc. Both datasets are separated into training, validation and test sets. The statistics of the datasets are summarized in Table 1. As we use a generator to predict the missing link of a predicate at each time step of RNN, the performance of generator is gaged via benchmark link prediction by using the standard HIT@10 metric.

Table 1: Statistics of WN18 and FB15k

Dataset	# Relation	# Entity	# Train	# Validate	# Test
WN18	18	40,943	141,442	5,000	5,000
FB15K	1,345	14,951	483,142	50,000	59,071

4.2 Results

Table 3 shows link prediction performances of the baseline, existing and proposed mh path learning models. The RNNs – trained to generate relation paths according to the method proposed in [8]– are used as baseline models and referred as PathG-RNNs. We also compare with PTransE [9] which is a variant of TransE for integrating relation paths for knowledge representation learning. Results reveal that: (1) mh-RGAN performs significantly better than baseline and existing models. (2) The relation paths provide very useful supplement for representation learning of KBs, which have been effectively modeled by mh-RGAN.

For example, since *David Cameron* and *Tony Blair* are both prime ministers of United Kingdom, they are assigned a similar embedding by single-hop representation learning methods. This may lead to confusion in single-hop methods in e.g. predicting the spouse of *Cherie Blair*. Contrarily, mh-RGAN learns relation paths between entities such as *Tony Blair* and *Cherie Blair* which helps it to perform more accurately. We also analyze the effect of path length by experimenting with a path-length of 2 (i.e. consisting of 2 triplets) and a path-length of 3 (i.e. consisting of 3 triplets). Results show that the performance of the model improves with path length. Since mh-RGAN is a generative model it can generate multi-hop relation paths. Examples of various generated relation paths are presented in Table 2. We initialize mh-RGAN with the entity given in the first column of the table. The following columns show the result of model prediction at each hop. We show a ranked list of predicted entities in the last column. In the first example, the model composes a hierarchical relation between entities. In the second, third and fourth examples, the model produces correlations such as hyponym and hypernym, and meronym and holonym between entities. In the fifth and sixth example, the model generates an elaboration of terms. It can be seen that each of the generated relation paths are plausible.

5 Conclusion

We have proposed a novel generative adversarial network based framework for reasoning over knowledge bases. The components of the framework consist in two networks: a generator for composing relation paths and a discriminator for classifying paths as real or fake. Experiments on

Table 2: Examples of relation paths generated by the path-kcgan generator on the wn18 dataset. The first column shows the entity given to the generator to initiate the generation process. The last column shows top ranked entities produced after hop 6.

Entity	Relation	Entity	Relation	Entity	Relation	Generated entites
asia	_has _part	Syria	_part _of	middle _east	_has _part	'lebanon', 'syria', 'turkey', 'iran', 'iraq', 'himalayas', 'nepal', 'india', 'london', 'rome'
dicot _family	_hyponym	magnoliid _dicot _family	_hypernym	dicot _family	_hyponym	'sapotaceae', 'rosid _dicot _family', 'asterid _dicot _family', 'anacardiaceae', 'magnoliid _dicot _family', 'dilleniid _dicot _family', 'myrtaceae'
diptera	_member _meronym	dipterous _insect	_member _holonym	diptera	_member _holonym	'insecta', 'animal_order', 'liliales', 'vertebrata', 'property', 'physical _condition', 'tool'
amphibian _family	_member _holonym	vertebrata	_member _meronym	Vertebrate	_hyponym	'bird', 'organization _of _american _states', 'feminist', 'coleoptera', 'blow', 'division _eubacteria', 'vertebrata', 'asterid _dicot _genus', 'dilleniid _dicot _genus', 'edible _fruit'

Table 3: Link prediction empirical results on WN18 and FB15k test sets with respect to the relation path representation method variants.

Method	WN18		FB15k	
	MRR	HIT@10	MRR	HIT@10
PathG-RNN [2 hop]	0.41	73.7	0.47	76.3
PathG-RNN [3 hop]	0.43	75.0	0.48	78.5
PTransE [2 hop]	0.49	78.2	0.50	82.2
PTransE [3 hop]	0.54	80.5	0.58	84.6
MH-RGAN [2 hop]	0.58	85.4	0.63	87.1
MH-RGAN [3 hop]	0.60	87.7	0.67	91.3

standard data sets show that proposed method outperforms both the baseline and the most relevant method in the literature. In future, we intend to explore GAN-based methods for generating discrete samples as the existing policy gradient based framework requires pretraining of generator and discriminator models.

References

[1] G.A. Miller, "WordNet: a lexical database for English", Commun. ACM 38.11 (1995): 39-41.

[2] B. Kurt et al., "Freebase: a collaboratively created graph database for structuring human knowledge", Proceedings of the 2008 ACM SIGMOD international conference on Management of data, AcM, 2008.

[3] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In Proceedings of ICML, pages 809–816.

[4] R. Socher et al., "Reasoning with neural tensor networks for knowledge base completion", Advances in neural information processing systems, 2013.

[5] A. Bordes et al., "Translating embeddings for modeling multi-relational data", Advances in neural information processing systems, 2013.

[6] Arvind Neelakantan, Benjamin Roth, and Andrew McCallum. 2015. Compositional vector space models for knowledge base completion. In Proceedings of ACL, pages 156–166.

[7] Rajarshi Das, Arvind Neelakantan, David Belanger, and Andrew McCallum. 2017. Chains of reasoning over entities, relations, and text using recurrent neural networks. In Proceedings of EACL, pages 132–141.

[8] Yin, Wenpeng, Yadollah Yaghoobzadeh, and Hinrich Schütze. "Recurrent One-Hop Predictions for Reasoning over Knowledge Graphs." arXiv preprint arXiv:1806.04523 (2018).

[9] Lin, Yankai, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. "Modeling relation paths for representation learning of knowledge bases." arXiv preprint arXiv:1506.00379 (2015).

[10] F. Yang, Y. Zhilin and W.C. William, "Differentiable learning of logical rules for knowledge base completion", CoRR, abs/1702.08367, 2017.

[11] T. Rocktäschel and R. Sebastian, "End-to-end differentiable proving", Advances in Neural Information Processing Systems, 2017.

[12] P. Minervini et al., "Towards neural theorem proving at scale", arXiv preprint arXiv:1807.08204, 2018.

[13] S. N. Tran, and S. A. G. Artur. "Deep logic networks: Inserting and extracting knowledge from deep belief networks", IEEE T NEUR NET LEAR. vol. 29, issue.2, 2016, pp: 246-258.

[14] G. Sourek et al., "Lifted relational neural networks: Efficient learning of latent relational structures", J ARTIF INTELL RES vol. 62, 2018, pp: 69-100.

[15] I. Goodfellow et al., "Generative adversarial nets", Advances in neural information processing systems, 2014.

[16] A. Radford, M. Luke and C. Soumith, "Unsupervised representation learning with deep convolutional generative adversarial networks", arXiv preprint arXiv:1511.06434, 2015.

[17] M. Mirza, and O. Simon, "Conditional generative adversarial networks", Manuscript: https://arxiv.org/abs/1709.02023, vol. 9, 2014, pp. 24.

[18] L. Yu et al. "Seqgan: Sequence generative adversarial nets with policy gradient." Thirty-First AAAI Conference on Artificial Intelligence, 2017.

[19] L. Cai, and Y.W. William, "Kbgan: Adversarial learning for knowledge graph embeddings", arXiv preprint arXiv:1711.04071, 2017.

[20] H. Wang et al., "Graphgan: Graph representation learning with generative adversarial nets", Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[21] Nguyen, Dai Quoc, et al. "A novel embedding model for knowledge base completion based on convolutional neural network." arXiv preprint arXiv:1712.02121 (2017).