Bi-directional Learning of Logical Rules with Type Constraints for Knowledge Graph Completion (Supplementary materials)

Kunxun Qi
School of Computer Science and Engineering, Sun Yat-sen University
Guangzhou, China
qikx@mail2.sysu.edu.cn

Jianfeng Du*
Guangdong University of Foreign Studies, Guangzhou, China
Bigmath Technology, Shenzhen
jfdu@gdufs.edu.cn

Hai Wan*
School of Computer Science and Engineering, Sun Yat-sen University
Guangzhou, China
wanhai@mail.sysu.edu.cn

ACM Reference Format:
Kunxun Qi, Jianfeng Du, and Hai Wan. 2024. Bi-directional Learning of Logical Rules with Type Constraints for Knowledge Graph Completion (Supplementary materials). In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3627673.3679695

A PROOF OF THEOREM 1

We first recall the basic notations and definitions.

A.1 Preliminaries

Knowledge Graph. Let $\mathcal{E}$ be a set of entities, $\mathcal{R}$ a set of relations and C a set of types. A *knowledge graph* $\mathcal{G}$ can be separated into two parts, i.e., $\mathcal{G} = \mathcal{G}_{\text{rel}} \cup \mathcal{G}_{\text{type}}$, where $\mathcal{G}_{\text{rel}} = \{(h_i, r_i, t_i)\}_{1\le i\le N_{\text{rel}}}$, $\mathcal{G}_{\text{type}} = \{(e_i, \mathsf{Type}, c_i)\}_{1\le i\le N_{\text{type}}}$, N_{rel} denotes the number of triples in $\mathcal{G}_{\text{rel}}$, N_{type} the number of triples in $\mathcal{G}_{\text{type}}$, $h_i \in \mathcal{E}$ (resp. $t_i \in \mathcal{E}$ or $r_i \in \mathcal{R}$) is the *head* entity (resp. *tail* entity or *relation*) for the i^{th} triple in $\mathcal{G}_{\text{rel}}$, and $e_i \in \mathcal{E}$ (resp. $c_i \in C$) is the entity (resp. type) for the i^{th} triple in $\mathcal{G}_{\text{type}}$. By r^- we denote the inverse relation of $r \in \mathcal{R}$. The set of inverse relations for $\mathcal{R}$, namely $\{r^- \mid r \in \mathcal{R}\}$, is denoted by $\mathcal{R}^-$. Accordingly, the equivalent set of triples for $\mathcal{G}_{\text{rel}}$ composed by inverse relations, namely $\{(t, r^-, h) \mid (h, r, t) \in \mathcal{G}_{\text{rel}}\}$, is denoted by $\mathcal{G}^-_{\text{rel}}$.

Inference Rule. An *atom* is a basic first-order logic formula of the form $p(t_1, \dots, t_n)$, where p is a *predicate* and $t_1, \dots, t_n$ are terms that denote either constants or variables. An r-specific inference rule R for the target relation r can be written of the form $r^{\text{new}}(x, y) \longleftarrow \exists \vec{z} : \varphi(x, y, \vec{z})$, where $\varphi(x, y, \vec{z})$ is a conjunction of atoms on variables x, y and $\vec{z}$, and r^{new} denotes the predicate of a new fact that inferred by a r-specific inference rule. The part of R at the left (resp. right) of $\longleftarrow$ is called the *head* (resp. *body*) of R. By H_R and B_R we denote the atom in the head of R and the set of atoms in the body of R, respectively. Note that we distinguish r^{new} from r to avoid recursive inference of new facts on r. An atom or a rule is *ground* if it does not contain any variable. An r-specific inference rule R is called a fact if B_R is empty and H_R is ground. To uniformly represent r-specific inference rules using fixed-length bodies, we introduce the *identity relation* (denoted by I) to rule bodies. For example, $r(x, y) \longleftarrow p(x, y)$ can be converted into a rule with two body atoms, namely $r(x, y) \longleftarrow p(x, z) \wedge I(z, y)$. We also allow to use both relations and inverse relations as predicates in inference rules. Throughout this paper, a triple (h_i, r_i, t_i) in $\mathcal{G}_{\text{rel}}$ and a binary atom $r_i(h_i, t_i)$, as well as a triple $(e_i, \mathsf{Type}, c_i)$ in $\mathcal{G}_{\text{type}}$ and a unary atom $c_i(e_i)$, are used interchangeably.

A *substitution* σ is a function that maps a set T of variables to a set T' of variables or constants. It is called ground if it maps all variables to constants. By $t\sigma = t'$ we denote that $t \in T$ is mapped to $t' \in T'$ by σ. Given an atom $A(t_1, \dots, t_n)$, we have $A(t_1, \dots, t_n)\sigma = A(t_1\sigma, \dots, t_n\sigma)$, and this naturally extends to a set of atoms. Given a knowledge graph $\mathcal{G}$, an r-specific inference rule R and a new fact $r^{\text{new}}(a, b)$, let $\mathcal{K} = \mathcal{G}_{\text{rel}} \cup \mathcal{G}^-_{\text{rel}} \cup \mathcal{G}_{\text{type}} \cup \{I(e, e) \mid e \in \mathcal{E}\}$, we say $\mathcal{K} \models_R r^{\text{new}}(a, b)$ if there exists a ground substitution σ such that $H_R\sigma = r^{\text{new}}(a, b)$ and $B_R\sigma \subseteq \mathcal{K}$. Given a set Σ of r-specific inference rules, we say $\mathcal{K} \vdash_\Sigma r^{\text{new}}(a, b)$ if there exists an r-specific inference rule $R \in \Sigma$ such that $\mathcal{K} \models_R r^{\text{new}}(a, b)$. We say $r^{\text{new}}(a, b)$ is *plausible* in $\mathcal{G}$ if there exists a set of possibly correct r-specific inference rules Σ such that $\mathcal{K} \vdash_\Sigma r^{\text{new}}(a, b)$.

Chain-like Rule. An r-specific inference rule is said to be *chain-like* if every body atom shares one variable with the previous body atom and the other variable with the next body atom. Formally, an r-specific chain-like rule with L body atoms, simply called an r-specific L-CR, is of the form:

$$r^{\text{new}}(x, y) \longleftarrow p_1(x, z_1) \wedge p_2(z_1, z_2) \wedge \cdots \wedge p_L(z_{L-1}, y)$$

where $p_1, \dots, p_L$ are relations in $\mathcal{R} \cup \mathcal{R}^- \cup \{I\}$.

Typed Rule. A typed rule [1] extends an r-specific L-CR with unary atoms. Let C be the set of types. An r-specific L-typed rule R is of the form:

$$r^{\text{new}}(x, y) \longleftarrow c_1(x) \wedge c_2(z_1) \wedge p_1(x, z_1) \wedge \cdots \wedge p_L(z_{L-1}, y) \wedge c_{L+1}(y)$$

where $c_1, c_2, \dots, c_{L+1}$ are types in C.

Link Prediction. Given a knowledge graph $\mathcal{G}$, a head query $(?, r^{\text{new}}, t)$ or a tail query $(h, r^{\text{new}}, ?)$, link prediction aims to find all entities $e \in \mathcal{E}$ such that (e, r^{new}, t) for $(?, r^{\text{new}}, t)$ or (h, r^{new}, e) for $(h, r^{\text{new}}, ?)$ is plausible in $\mathcal{G}$.

Triple Classification. Given a knowledge graph $\mathcal{G}$ and a triple (h, r^{new}, t) where $h \in \mathcal{E}$, $t \in \mathcal{E}$ and $r \in \mathcal{R}$, triple classification aims to estimate whether (h, r^{new}, t) is plausible in $\mathcal{G}$.

*Both authors are corresponding authors.

CIKM '24, October 21–25, 2024, Boise, ID, USA

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/10.1145/3627673.3679695

We also recall the definition of TC-rules.

Definition 1. *An r-specific L-TC-rule (simply a TC-rule if r and L are clear from the context) R is of the form: $r^{\text{new}}(x,y) \leftarrow p_1(x,z_1) \wedge p_2(z_1,z_2) \wedge ... \wedge p_L(z_{L-1},y) \wedge C_1(x) \wedge C_2(z_1) \wedge ... \wedge C_L(z_{L-1}) \wedge C_{L+1}(y)$, where $C_l(u) \in \{E_l(u), H_l(u), E_l(u) \vee H_l(u)\}$, $E_l(u) = \bigvee_{i=1}^{m_l} g_{l,i}(u)$ with $g_{l,i}$ being different predicates in C and $0 \le m_l \le |C|$ is called an* explicit type constraint *on u, and $H_l(u) = \bigvee_{i=1}^{n_l} q_{l,i}(u, v_{l,i})$ with $v_{l,i}$ being new variables, $q_{l,i}$ being different predicates in $\mathcal{R} \cup \mathcal{R}^-$ and $0 \le n_l \le |\mathcal{R} \cup \mathcal{R}^-|$ is called an* implicit type constraint *on u. Some entity variables v can have no type constraint; in this case $C_l(v)$ is empty, i.e., $m_l = 0$ and $n_l = 0$.*

A.2 Formalization of TCLM

Let $\mathcal{G}$ be a given knowledge graph, N the maximum number of rules to be learnt, L the maximum number of body atoms, $\mathcal{K} = \mathcal{G}_{\text{rel}} \cup \mathcal{G}^-_{\text{rel}} \cup \{(e, I, e) \mid e \in \mathcal{E}\}$ the background knowledge, $C = \{c_1, \cdots, c_m\}$ the set of explicit types, and $n = |\mathcal{R}|$. Suppose $\mathcal{R} = \{r_i\}_{1 \le i \le n}$, its corresponding set of inverse relations $\mathcal{R}^- = \{r_i\}_{n+1 \le i \le 2n}$, and $I = r_{2n+1}$. The goal of TCLM is to estimate a truth degree $\xi^{N,L}_{r,x,y}$ for the triple $(x,r,y) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$, where the estimated truth degree $\xi^{N,L}_{r,x,y}$ reflects the degree of whether the triple (x,r,y) can be inferred by a certain rule among N L-TC-rules. For $1 \le k \le N, 1 \le l \le L$, the intermediate estimated truth degree $s^{(k,l)}_{r,x,y}$ for the l^{th} atom in the k^{th} rule is defined as below.

For $l = 1$, the truth degree is calculated by:

$$s^{(k,1)}_{r,x,y} = \phi^{(k,1)}_r(x)\phi^{(k,l+1)}_r(y) \sum_{i=1}^{2n+1} w^{(r,k,1)}_i \mathbb{I}((x, r_i, y) \in \mathcal{K}) \quad (1)$$

For $2 \le l \le L$, the truth degree is calculated by:

$$s^{(k,l)}_{r,x,y} = \phi^{(k,l+1)}_r(y) \sum_{i=1}^{2n+1} w^{(r,k,l)}_i \sum_{z:(z,r_i,y)\in\mathcal{K}} s^{(k,l-1)}_{r,x,z} \quad (2)$$

where $w^{(r,k,l)} \in [0,1]^{2n+1}$ denotes the trainable relational selection weights for the l^{th} atoms in the k^{th} rule for the head relation r. $\mathbb{I}(\psi)$ is an indicator function that returns 1 if ψ is true or 0 otherwise. $w^{(r,k,l)}$ is confined to $[0,1]^{2n+1}$ by a softmax layer. $\phi^{(k,l)}_r(u)$ is a scoring function for the type constraints on u. Formally, $\phi^{(k,l)}_r(u)$ is defined as:

$$\begin{aligned}\phi^{(k,l)}_r(u) = \sigma_{01}(&\alpha^{(r,k,l)} \sum_{i=1}^{m} h^{(r,k,l)}_i \mathbb{I}((u, \mathsf{Type}, c_i) \in \mathcal{G}_{\text{type}}) \\ &+ \beta^{(r,k,l)} \sum_{i=1}^{2n} h^{(r,k,l)}_{i+m} \mathbb{I}(\exists z : (u, r_i, z) \in \mathcal{G}_{\text{rel}} \cup \mathcal{G}^-_{\text{rel}}) \\ &+ (1 - \sigma_{01}(\alpha^{(r,k,l)} + \beta^{(r,k,l)})))\end{aligned} \quad (3)$$

where $\sigma_{01}(x) = \max(\min(x,1),0)$, $h^{(r,k,l)} \in [0,1]^{m+2n}$ denotes the trainable type selection weights of the l^{th} type constraint in the l^{th} rule for relation r. $h^{(r,k,l)}$ is confined to $[0,1]$ by σ_{01}. We use $\alpha^{(r,k,l)}$ (resp. $\beta^{(r,k,l)}$) to control whether the entity has an explicit (resp. implicit) type constraint. For example, $\alpha^{(r,k,l)} = 1$ (resp. $\beta^{(r,k,l)} = 1$) implies that there is an explicit (resp. implicit) type constraint for the given entity. Note that $\alpha^{(r,k,l)} = 0$ and $\beta^{(r,k,l)} = 0$ imply that there is no type constraint for the given entity.

Intuitively, Equation (1-3) simulates the inference of TC-rules, where the part on the right side of $\phi^{(k,l)}_r(u)$ in Equation (1-2) simulates the inference of chain-like rules, while $\phi^{(k,l)}_r(u)$ captures the type constraints of entities.

Then the ultimate estimated truth degree is calculated by weight-summing the estimated truth degrees for N rules:

$$\xi^{N,L}_{r,x,y} = \sum_{k=1}^{N} \mu^{(r)}_k s^{(k,L)}_{r,x,y} \quad (4)$$

where $\mu^{(r)}_l \in [-1,1]$ is a trainable weight that represents the weight of the k^{th} rule. $\mu^{(r)}_l$ is confined to $[-1,1]$ by a tanh layer. By assigning different weights to each rule, TCLM can learn different numbers of rules for different head relations, as the rules with weights close to 0 can be omitted. The model is trained by minimizing the following objective function

$$\mathcal{L} = -\sum_{(x,r,y)\in\mathcal{G}} \log \frac{\exp(\xi^{N,L}_{r,x,y})}{\exp(\xi^{N,L}_{r,x,y}) + \sum_{e\in\mathcal{E},(x,r,e)\notin\mathcal{G}} \exp(\xi^{N,L}_{r,x,e})} \quad (5)$$

The intuition of the above objective is to distinguish a true triple $(x,r,y) \in \mathcal{G}$ from its corrupted, probably false triples $(x,r,e) \notin \mathcal{G}$. By introducing the following notion of induced parameter assignment, we show in Theorem 1 that the formalization of TCLM is faithful to a certain set of TC-rules.

Definition 2. *Given a set of N r-specific L-TC-rules $\Sigma = \{R_k\}_{1 \le k \le N}$, where R_k is of the form $r^{\text{new}}(x,y) \leftarrow p_{k,1}(x,z_1) \wedge ... \wedge p_{k,L}(z_{L-1},y) \wedge C_{k,1}(x) \wedge ... \wedge C_{k,L+1}(y)$, where $p_{k,l} \in \mathcal{R} \cup \mathcal{R}^- \cup \{I\}$, $C_{k,l}(u) \in \{E_{k,l}(u), H_{k,l}(u), E_{k,l}(u) \vee H_{k,l}(u)\}$, $E_{k,l}(u) = \bigvee_{i=1}^{m_{k,l}} g_{k,l,i}(u)$ with $g_{k,l,i}$ being different predicates in C and $0 \le m_{k,l} \le |C|$, $H_{k,l}(u) = \bigvee_{i=1}^{n_{k,l}} q_{k,l,i}(u, v_{k,l,i})$ with $v_{k,l,i}$ being new variables, $q_{k,l,i}$ being different predicates in $\mathcal{R} \cup \mathcal{R}^-$ and $0 \le n_{k,l} \le |\mathcal{R} \cup \mathcal{R}^-|$, we call a parameter assignment of TCLM $\theta^{N,L}_r = \{w^{(r,k,l)}_i\}_{1\le k\le N, 1\le l\le L, 1\le i\le 2n+1} \cup \{h^{(r,k,l)}_i\}_{1\le k\le N, 1\le l\le L+1, 1\le i\le 2n+m} \cup \{\alpha^{(r,k,l)}, \beta^{(r,k,l)}\}_{1\le k\le N, 1\le l\le L+1} \cup \{\mu^{(r)}_k\}_{1\le l\le L}$* Σ-induced *if it satisfies the following conditions for all $1 \le k \le N, 1 \le l \le L$:*

(1) $\forall 1 \le i \le 2n+1 : w^{(r,k,l)}_i = 1$ if $p_{k,l} = r_i$, otherwise $w^{(r,k,l)}_i = 0$.

(2) $\forall 1 \le i \le m : h^{(r,k,l)}_i = 1$ if $g_{k,l,j} = c_i$ for some $j \in \{1, \cdots, m_{k,l}\}$, otherwise $h^{(r,k,l)}_i = 0$.

(3) $\forall 1 \le i \le 2n : h^{(r,k,l)}_{i+m} = 1$ if $q_{k,l,j} = r_i$ for some $j \in \{1, \cdots, n_{k,l}\}$, otherwise $h^{(r,k,l)}_{i+m} = 0$.

(4) $\alpha^{(r,k,l)} = 1$ if there is some $g_{k,l,j}$ appearing in $C_{k,l}$, otherwise $\alpha^{(r,k,l)} = 0$.

(5) $\beta^{(r,k,l)} = 1$ if there is some $q_{k,l,j}$ appearing in $C_{k,l}$, otherwise $\beta^{(r,k,l)} = 0$.

(6) $u^{(k)}_r = 1$.

B PROOF OF THEOREM 1

To prove Theorem1, we first introduce Lemma 1.

Lemma 1. *Let $\mathcal{G}$ be a knowledge graph, $\mathcal{K} = \mathcal{G}_{\text{rel}} \cup \mathcal{G}^-_{\text{rel}} \cup \mathcal{G}_{\text{type}} \cup \{I(e,e) \mid e \in \mathcal{E}\}$, R an r-specific L-TC-rule, and $\theta_r^{(1,L)}$ the $\{R\}$-induced parameter assignment of TCLM. Given an arbitrary triple (a, r^{new}, b), then either $\xi^{1,L}_{r,a,b} \geq 1$ or $\xi^{1,L}_{r,a,b} = 0$, and meanwhile $\xi^{1,L}_{r,a,b} \geq 1$ if $\mathcal{K} \models_R r^{\text{new}}(a,b)$, $\xi^{1,L}_{r,a,b} = 0$ if $\mathcal{K} \not\models_R r^{\text{new}}(a,b)$.*

Proof. Suppose that R is of the form: $r^{\text{new}}(x,y) \leftarrow p_1(x,z_1) \wedge p_2(z_1,z_2) \wedge \ldots \wedge p_L(z_{L-1},y) \wedge C_1(x) \wedge C_2(z_1) \wedge \ldots \wedge C_L(z_{L-1}) \wedge C_{L+1}(y)$, where $C_l(u) \in \{E_l(u), H_l(u), E_l(u) \vee H_l(u)\}$, $E_l(u) = \bigvee_{i=1}^{m_l} g_{l,i}(u)$ with $g_{l,i}$ being different predicates in C and $0 \leq m_l \leq |C|$, and $H_l(u) = \bigvee_{i=1}^{n_l} q_{l,i}(u, v_{l,i})$ with $v_{l,i}$ being new variables, $q_{l,i}$ being different predicates in $\mathcal{R} \cup \mathcal{R}^-$ and $0 \leq n_l \leq |\mathcal{R} \cup \mathcal{R}^-|$.

(I) Consider the case where $\mathcal{K} \models_R r^{\text{new}}(a,b)$. There exists at least one ground substitution σ such that $H_R\sigma = r^{\text{new}}(a,b)$ and $B_R\sigma \subseteq \mathcal{K}$. There will be a sequence of entities $e_1, \ldots, e_{L-1} \in \mathcal{E}$ such that $(a,p_1,e_1), (e_1,p_2,e_2), \ldots, (e_{L-1},p_L,b) \in \mathcal{K}$ and $g_{1,1}(a)$, $g_{1,2}(a), \ldots, g_{2,1}(e_1), \ldots, g_{L+1,m_{L=1}}(b) \in \mathcal{K}$, and a sequence of entities $e'_{1,1}, e'_{1,2}, \ldots, e'_{L+1,n_{L+1}} \in \mathcal{E}$ such that $q_{1,1}(a, e'_{1,1})$, $q_{1,2}(a, e'_{1,2})$, $\ldots$, $q_{L+1,n_{L+1}(b,e'_{L+1,n_{L+1}})} \in \mathcal{K}$. Suppose r_1 is the k^{th} relation in $\mathcal{R} \cup \mathcal{R}^- \cup \{I\}$, then by Condition 1 in Definition 2, we have $w_k^{(r,1,1)} = 1$ for some k. By $(a,p_1,e_1) \in \mathcal{K}$ and Equation (1), we have $\sum_{i=1}^{2n+1} w_i^{(r,1,1)} \mathbb{I}((a,r_i,e_1) \in \mathcal{K}) \geq 1$. For $\phi_r^{1,1}(a)$, there are three cases. In the first case, we have $m_1 > 0$. Suppose $g_{1,j}$ for all $j \in \{1,\ldots,m_1\}$ is the k^{th} type in C. By Condition 2 in Definition 2, we have $h_k^{(r,1,1)} = 1$ for some k. By $m_1 > 0$ and Condition 4 in Definition 2, we have $\alpha^{(r,1,1)} = 1$. By Equation (3), we have $\phi_r^{1,1}(a) = 1$. In the second case, we have $n_1 > 0$. Suppose $q_{1,j}$ for all $j \in \{1,\ldots,n_1\}$ is the k^{th} relation in $\mathcal{R} \cup \mathcal{R}^-$. By Condition 3 in Definition 2, we have $h_{m+k}^{(r,1,1)} = 1$ for some k. By $n_1 > 0$ and Condition 5 in Definition 2, we have $\beta^{(r,1,1)} = 1$. By Equation (3), we have $\phi_r^{1,1}(a) = 1$. In the third case, we have $m_1 = 0$ and $n_1 = 0$. By Condition 4 and 5 in Definition 2, we have $\alpha^{(r,1,1)} = 1$ and $\beta^{(r,1,1)} = 0$, respectively. By Equation (3), we have $\phi_r^{1,1}(a) = 1$. Therefore, we have $\phi_r^{1,1}(a) = 1$ for all three cases. Likewise, we can prove that $\phi_r^{1,1}(e_1) = 1$. By Equation (1), we further have $s^{(1,1)}_{r,a,e_1} \geq 1$.

Likewise, suppose p_2 is the k^{th} relation in $\mathcal{R} \cup \mathcal{R}^- \cup \{I\}$, then by Condition 1 in Definition 1, we have $w_k^{(r,1,2)} = 1$. By $(e_1,p_2,e_2) \in \mathcal{K}$ and Equation (2), we have $\sum_{i=1}^{2n+1} w_i^{(r,1,1)} \sum_{z:(z,r_i,e_2)} s^{(1,1)}_{r,x,z} \geq 1$. For $\phi_r^{1,1}(e_2)$, there are three cases. In the first case, we have $m_3 > 0$. Suppose $g_{3,j}$ for all $j \in \{1,\ldots,m_3\}$ is the k^{th} type in C. By Condition 2 in Definition 2, we have $h_k^{(r,1,3)} = 1$ for some k. By $m_3 > 0$ and Condition 4 in Definition 2, we have $\alpha^{(r,1,3)} = 1$. By Equation (3), we have $\phi_r^{1,3}(e_2) = 1$. In the second case, we have $n_3 > 0$. Suppose $q_{3,j}$ for all $j \in \{1,\ldots,n_3\}$ is the k^{th} relation in $\mathcal{R} \cup \mathcal{R}^-$. By Condition 3 in Definition 2, we have $h_{m+k}^{(r,1,3)} = 1$ for some k. By $n_3 > 0$ and Condition 5 in Definition 2, we have $\beta^{(r,1,3)} = 1$. By Equation (3), we have $\phi_r^{1,3}(e_2) = 1$. In the third case, we have $m_3 = 0$ and $n_3 = 0$. By Condition 4 and 5 in Definition 2, we have $\alpha^{(r,1,3)} = 1$ and $\beta^{(r,1,3)} = 0$, respectively. By Equation (3), we have $\phi_r^{1,3}(e_2) = 1$. Therefore, we have $\phi_r^{1,3}(e_2) = 1$ for all three cases. Likewise, we can prove that $\phi_r^{1,3}(e_2) = 1$. By Equation (1), we further have $s^{(1,2)}_{r,a,e_2} \geq 1$.

In the same way, we can prove that $s^{(1,3)}_{r,a,e_3} \geq 1, \ldots, s^{(1,L-1)}_{r,a,e_{L-1}} \geq 1$ and $s^{(1,L)}_{r,a,b} \geq 1$ in turn. By Equation (4) and Condition 6 in Definition 2, we have $\xi^{1,L}_{r,a,b} \geq 1$ if $\mathcal{K} \models_R r^{\text{new}}(a,b)$.

(II) Consider the case where $\mathcal{K} \not\models_R r^{\text{new}}(a,b)$. Suppose $\xi^{1,L}_{r,a,b} \geq 1$, then by Equation (1) and (4), there must be some $k \in \{1,\ldots,2n+1\}$ such that $p_1 = r_k$ and $w_k^{(r,1,1)} = 1$, and there exists an entity e_1 such that $(a,p_1,e_1) \in \mathcal{K}$ fulfilling $s^{(1,1)}_{r,a,e_1} \geq 1$. Since $s^{(1,1)}_{r,a,e_1} \geq 1$, we have $\phi_r^{(1,1)}(a) = 1$. There are three cases for $\phi_r^{(1,1)}(a)$. In the first case, we have $m_1 > 0$. By Equation (3) and Condition 2 in Definition 2, for all $j \in \{1,\ldots,m_1\}$, there must be some $k \in \{1,\ldots,m\}$ such that $g_{1,j} = c_k$, $h_k^{(r,1,1)} = 1$, $\alpha^{(r,1,1)} = 1$ and $g_{1,j}(a) \in \mathcal{K}$. In the second case, we have $n_1 > 0$. By Equation (3) and Condition 3 in Definition 2, for all $j \in \{1,\ldots,n_1\}$, there must be some $k \in \{1,\ldots,2n\}$ and some entity $e'_{1,j}$ such that $q_{1,j} = r_k$, $h_{m+k}^{(r,1,1)} = 1$, $\beta^{(r,1,1)} = 1$ and $(a, g_{1,j}, e'_{1,j}) \in \mathcal{K}$. In the third case, we have $m_1 = 0$ and $n_1 = 0$. Likewise, we prove that for all $j \in \{1,\ldots,m_2\}$ when $m_2 > 0$, there must be some $k \in \{1,\ldots,m\}$ such that $g_{2,j} = c_k$, $h_k^{(r,1,2)} = 1$, $\alpha^{(r,1,2)} = 1$ and $g_{2,j}(e_1) \in \mathcal{K}$. Meanwhile, for all $j \in \{1,\ldots,n_2\}$ when $n_2 > 0$, there must be some $k \in \{1,\ldots,2n\}$ and some entity $e'_{2,j}$ such that $q_{2,j} = r_k$, $h_{m+k}^{(r,1,2)} = 1$, $\beta^{(r,1,2)} = 1$ and $(e_1, q_{2,j}, e'_{2,j}) \in \mathcal{K}$.

Since $s^{(1,1)}_{r,a,e_1} \geq 1$, by Equation (2), there must be also some $k \in \{1,\ldots,2n+1\}$ such that $p_2 = r_k$ and $w_k^{(r,1,2)} = 1$, and there exists an entity e_2 such that $(e_1,p_2,e_2) \in \mathcal{K}$ fulfilling $s^{(1,2)}_{r,a,e_2} \geq 1$. Since $s^{(1,2)}_{r,a,e_2} \geq 1$, we have $\phi_r^{(1,3)}(e_2) = 1$. There are three cases for $\phi_r^{(1,3)}(e_2)$. In the first case, we have $m_3 > 0$. By Equation (3) and Condition 2 in Definition 2, for all $j \in \{1,\ldots,m_3\}$, there must be some $k \in \{1,\ldots,m\}$ such that $g_{3,j} = c_k$, $h_k^{(r,1,3)} = 1$, $\alpha^{(r,1,3)} = 1$ and $g_{3,j}(a) \in \mathcal{K}$. In the second case, we have $n_3 > 0$. By Equation (3) and Condition 3 in Definition 2, for all $j \in \{1,\ldots,n_3\}$, there must be some $k \in \{1,\ldots,2n\}$ and some entity $e'_{3,j}$ such that $q_{3,j} = r_k$, $h_{m+k}^{(r,1,3)} = 1$, $\beta^{(r,1,3)} = 1$ and $(e_2, q_{3,j}, e'_{3,j}) \in \mathcal{K}$. In the third case, we have $m_3 = 0$ and $n_3 = 0$.

In the same way, we can show that there exists an entity e_i such that $(e_{i-1},p_i,e_i) \in \mathcal{K}$ and $s^{(1,i)}_{r,a,e_i} \geq 1$ for $i = 3,\ldots,L-1$ in turn, while we have $(e_{L-1},p_L,b) \in \mathcal{K}$. Meanwhile, for all $j \in \{1,\ldots,m_i\}$ when $m_i > 0$, we have $g_{i+1,j}(e_i) \in \mathcal{K}$. For all $j \in \{1,\ldots,n_i\}$ when $n_i > 0$, there must be an entity $e'_{i+1,j}$ such that $(e_i, q_{i+1,j}, e'_{i+1,j}) \in \mathcal{K}$. Hence there exists a sequence of entities $e_1, \ldots, e_{L-1}$ such that $(a,p_1,e_1), (e_1,p_2,e_2) \ldots, (e_{L-1},p_L,b) \in \mathcal{K}$ and $g_{1,1}(a), \ldots, g_{2,1}(e_1), \ldots, g_{L+1,m_{L+1}}(b) \in \mathcal{K}$, and a sequence of entities $e'_{1,1}, e'_{1,2}, \ldots, e'_{1,n_1}, \ldots, e'_{L+1,1}, \ldots, e'_{L+1,n_{L+1}}$ such that $(a, q_{1,1}, e'_{1,1})$, $\ldots$, $(a, q_{1,n_1}, e'_{1,n_1})$, $\ldots$, $(b, q_{L+1,1}, e'_{L+1,1})$,$\ldots$, $(b, q_{L+1,n_{L+1}}, e'_{L+1,n_{L+1}}) \in \mathcal{K}$. These two sequences constitute a ground substitution σ such that $H_R\sigma = r^{\text{new}}(a,b)$ and $B_R\sigma \subseteq \mathcal{K}$, contradicting $\mathcal{G}_d \not\models_R r^{\text{new}}(a,b)$. Thus, we have $\xi^{1,L}_{r,a,b} < 1$. By Equation (1-4) we have $\xi^{1,L}_{r,a,b} \geq 0$. Therefore, we have $\xi^{1,L}_{r,a,b} = 0$ if $\mathcal{K} \not\models_R r^{\text{new}}(a,b)$. □

Table 1: Hyper-parameter settings on different datasets.

Hyper-parameter	Datasets with explicit types		Datasets without explicit types					
	AirGraph	YAGO26K906	Family	Kinship	UMLS	WN18RR	FB15K237	YAGO3-10
Number of rules N for each relation	50	50	70	70	70	100	70	50
Maximum length L of each rule	3	3	3	3	3	4	3	3
Maximum number of training epoch	100	100	50	50	50	100	50	50
Learning rate	1e-1	1e-1	1e-1	1e-1	1e-1	1e-1	1e-1	1e-1
Dropout rate	0.1	0.1	0.3	0.3	0.3	0.1	0.1	0.1
Batch size	4	4	32	32	32	32	4	4

Theorem 1. *Let $\mathcal{G}$ be a knowledge graph, $\mathcal{K} = \mathcal{G}_{\text{rel}} \cup \mathcal{G}^-_{\text{rel}} \cup \mathcal{G}_{\text{type}} \cup \{I(e,e) \mid e \in \mathcal{E}\}$, $\Sigma = \{R_k\}_{1\le k\le N}$ a set of r-specific L-TC-rules and $\theta_r^{N,L}$ the Σ-induced parameter assignment of TCLM. Given an arbitrary triple (a, r^{new}, b), $\xi^{N,L}_{r,a,b} \ge 1$ if and only if $\mathcal{K} \vdash_\Sigma r^{\text{new}}(a,b)$.*

Proof. From Lemma 1 we know that for all $R_k \in \Sigma$, $\mathcal{K} \models_{R_k} r^{\text{new}}(a,b)$ if $\xi^{1,L}_{r,a,b} \ge 1$, and $\mathcal{K} \not\models_{\{R_k\}} r^{\text{new}}(a,b)$ if $\xi^{1,L}_{r,a,b} = 0$.

($\Rightarrow$) Suppose $\xi^{N,L}_{r,a,b} \ge 1$. Then by Condition 6 in Definition 2, there exists at least one TC-rule $R_k \in \Sigma$ such that $s^{(k,L)}_{r,a,b} \ge 1$. By Lemma 1, we have $\mathcal{K} \models_{R_k} r^{\text{new}}(a,b)$. By $\mathcal{K} \models_{R_k} r^{\text{new}}(a,b)$ and $R_k \in \Sigma$, we have $\mathcal{K} \vdash_\Sigma r^{\text{new}}(a,b)$.

($\Leftarrow$) Suppose $\mathcal{K} \vdash_\Sigma (a, r^{\text{new}}, b)$. Then we have $\mathcal{K} \models_{R_k} r^{\text{new}}(a,b)$ for some $R_k \in \Sigma$. By Lemma 1 and Condition 6 in Definition 2, we have $s^{(k,L)}_{r,a,b} \ge 1$ and for all $k' \ne k$, $s^{(k',L)}_{r,a,b} \ge 0$. By Equation (4) and Condition 6 in Definition 2, we have $\xi^{N,L}_{r,a,b} \ge 1$. □

C PROOF OF THEOREM 2

We first recall the formalization of bi-directional learning.

To explain why a given triple (h, r^{new}, t) is plausible in $\mathcal{G}$, we should avoid confusing explanations, i.e., the explanations for answering both $(?, r^{\text{new}}, t)$ and $(h, r^{\text{new}}, ?)$ should be the same. Therefore, we propose a bi-directional learning mechanism that enforces the model to yield the same set of logical rules by learning shared parameters for answering both $(?, r^{\text{new}}, t)$ and $(h, r^{\text{new}}, ?)$. Formally, the estimation of the truth degree of (h, r^{new}, t) from the angle of answering head queries is defined below.

For $l = 1$, the truth degree is calculated by:

$$\bar{s}^{(k,1)}_{r^-,y,x} = \phi_r^{(k,L+1)}(y)\phi_r^{(k,L)}(x) \sum_{i=1}^{2n+1} w_i^{(r,k,L)} \mathbb{I}((x, r_i, y) \in \mathcal{K}) \quad (6)$$

For $2 \le l \le L$, the truth degree is calculated by:

$$\bar{s}^{(k,l)}_{r^-,y,x} = \phi_r^{(k,L-l+1)}(x) \sum_{i=1}^{2n+1} w_i^{(r,k,L-l+1)} \sum_{z:(x,r_i,z)\in\mathcal{K}} \bar{s}^{(k,l-1)}_{r^-,y,z} \quad (7)$$

Then the ultimate estimated degree is formally defined as:

$$\bar{\xi}^{N,L}_{r^-,y,x} = \sum_{k=1}^{N} \mu_l^{(r)} \bar{s}^{(k,L)}_{r^-,y,x} \quad (8)$$

where all the trainable parameters are shared with $\xi^{N,L}_{r,x,y}$.

The following theorem shows the consistency of estimated truth degrees for answering both $(?, r^{\text{new}}, t)$ and $(h, r^{\text{new}}, ?)$.

Theorem 2. *Let $\mathcal{G}$ be a knowledge graph. For any triple $(a, r, b) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$, $\forall N \ge 1, L \ge 1 : \xi^{N,L}_{r,a,b} = \bar{\xi}^{N,L}_{r^-,b,a}$.*

Proof. Let $\mathcal{M}$ be a function that maps a relation to its index.

By $\{P_i\}_{1\le i\le N_P}$ we denote the set of paths from entity a to entity b, where N_P is the number of different paths, and P_i consists of L triples $p_{i,1}(a, e_{i,1})$, $p_{i,2}(e_{i,1}, e_{i,2})$, …, $p_{i,L}(e_{i,L-1}, b)$. Then the truth degree of (a, r, b) is calculated by $\xi^{N,L}_{r,a,b} = \sum_{k=1}^N \mu_k^{(r)} \sum_{j=1}^{N_P} \phi_r^{(k,1)}(a) \prod_{l=1}^L w^{(r,k,l)}_{\mathcal{M}(p_{k,l})} \phi_r^{(k,l+1)}(e_{j,l})$, where $e_{k,L}$ is set to b. Consider the truth degree of (b, r^-, a) namely $\bar{\xi}^{N,L}_{r^-,b,a}$. Since paths from b to a are inverse parts from a to b, we know that $\{P'_i\}_{1\le i\le N_P}$ is the set of paths from b to a, where P'_i consists of L triples $p^-_{i,L}(b, e_{i,L-1}), \ldots, p^-_{i,2}(e_{i,2}, e_{i,1}), p^-_{i,1}(e_{i,1}, a)$. Hence $\bar{\xi}^{N,L}_{r^-,b,a} = \sum_{k=1}^N \mu_k^{(r)} \sum_{j=1}^{N_P} \phi_r^{(k,L+1)}(b) \prod_{l=1}^L w^{(r,l,L-l+1)}_{\mathcal{M}(p_{k,L-l+1})} \phi_r^{(k,L-l+1)}(e_{k,L-l})$, where $e_{k,0}$ is set to a. It follows that

$$\begin{aligned} & \phi_r^{(k,1)}(a) \prod_{l=1}^{L} w^{(r,k,l)}_{\mathcal{M}(p_{k,l})} \phi_r^{(k,l+1)}(e_{j,l}) \\ = \quad & \prod_{l=1}^{L} w^{(r,k,L-l+1)}_{\mathcal{M}(p_{k,L-l+1})} \prod_{l=0}^{L} \phi_r^{(k,L-l+1)}(e_{j,L-l}) \\ = \quad & \phi_r^{(k,L+1)}(b) \prod_{l=1}^{L} w^{(r,l,L-l+1)}_{\mathcal{M}(p_{k,L-l+1})} \phi_r^{(k,L-l+1)}(e_{j,L-l}) \end{aligned}$$

Therefore, we have $\forall N \ge 1, L \ge 1 : \xi^{N,L}_{r,a,b} = \bar{\xi}^{N,L}_{r^-,b,a}$. □

D HYPER-PARAMETER DETAILS

To help reproduce our results, we provide the hyper-parameter settings used in our experiments. Table 1 reports the detailed hyper-parameter settings in regard to different baseline models and datasets. These hyper-parameters are set to maximize the MRR scores on the validation set. Note that all trainable parameters in TCLM are initialized randomly.

REFERENCES

[1] Hong Wu, Zhe Wang, Kewen Wang, and Yi-Dong Shen. 2022. Learning Typed Rules over Knowledge Graphs. In *KR*.