Validation of Growing Knowledge Graphs by Abductive Text Evidences

Jianfeng Du,[†] Jeff Z. Pan,[‡] Sylvia Wang,[‡] Kunxun Qi,[†] Yuming Shen,[‡] Yu Deng[§]
[†]Guangdong University of Foreign Studies, Guangzhou 510420, P.R.China
[‡]Department of Computing Science, The University of Aberdeen, Aberdeen AB24 3UE, UK
[§]IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA
jfdu@gdufs.edu.cn, jeff.z.pan@abdn.ac.uk, sylviawangfr@hotmail.com,
kunxunqi@foxmail.com, ymshen2002@163.com, dengy@us.ibm.com

Abstract

This paper proposes a validation mechanism for newly added triples in a growing knowledge graph. Given a logical theory, a knowledge graph, a text corpus, and a new triple to be validated, this mechanism computes a sorted list of explanations for the new triple to facilitate the validation of it, where an explanation, called an *abductive text evidence*, is a set of pairs of the form (triple, window) where appending the set of triples on the left to the knowledge graph enforces entailment of the new triple under the logical theory, while every sentence window on the right which is contained in the text corpus explains to some degree why the triple on the left is true. From the angle of practice, a special class of abductive text evidences called *TEP-based abductive text evidence* is proposed, which is constructed from explanation patterns seen before in the knowledge graph. Accordingly, a method for computing the complete set of TEP-based abductive text evidences is proposed. Moreover, a method for sorting abductive text evidences based on distantly supervised learning is proposed. To evaluate the proposed validation mechanism, four knowledge graphs with logical theories are constructed from the four great classical masterpieces of Chinese literature. Experimental results on these datasets demonstrate the efficiency and effectiveness of the proposed mechanism.

Introduction

Knowledge graphs have been widely used in many applications recently. A knowledge graph is a directed graph with vertices labeled by entities and edges labeled by relations. It can be treated as a set of triples of the form $\langle h, r, t\rangle$, where h is the *head entity* (simply *head*), r the *relation* and t the *tail entity* (simply *tail*). In order to widen the applicability, knowledge graphs need to grow with the increasing data nowadays. There exist some common approaches to enlarging knowledge graphs such as knowledge graph completion. Knowledge graph completion aims to enlarge a knowledge graph with those new triples that are predicted by a learnt model and pass the manual validation. However, manual validation of triples is laborious and error prone. Thus it calls for automatic methods for explaining the existence of new triples to assist the validation of new triples.

There are two categories of traditional explanations that can be used to prove the existence of new triples.

One is the logic-based category. Given a logical theory formalizing the schema of the knowledge graph, a new triple can be naturally explained by a set of triples that are already in the knowledge graph under the logical theory. But the coverage of this mechanism is limited. To explain more triples, we can employ logic-based abduction (Eiter and Gottlob 1995), i.e. explain a new triple by a set of triples that are possibly not in the knowledge graph. However, in this mechanism we need to prove the existence of the absent triples. For example, given a logical theory made up of a single rule $\forall x, y, z : \mathsf{Father}(x, y) \wedge \mathsf{Father}(y, z) \rightarrow \mathsf{Grandfather}(x, z)$, the new triple $\langle \mathsf{Tom}, \mathsf{Grandfather}, \mathsf{Tim}\rangle$ can be explained by $\{\langle \mathsf{Tom}, \mathsf{Father}, e\rangle, \langle e, \mathsf{Father}, \mathsf{Tim}\rangle\}$ for some entity e under the logical theory. But to get a complete proof, we need to explain why $\langle \mathsf{Tom}, \mathsf{Father}, e\rangle$ and $\langle e, \mathsf{Father}, \mathsf{Tim}\rangle$ exist. It shows that logic-based abduction improves the coverage for new triples but may require extra validations.

The other category of explanations, called *text evidences*, are text-based. A text evidence for a triple $\langle h, r, t\rangle$ is a *sentence window* composed by several consecutive sentences that explain why h and t have the directed relation r. However, this explanation mechanism is only applied to new triples explainable by consecutive sentences. Consider the aforementioned example. If there is no sentence window mentioning both Tom and Tim, $\langle \mathsf{Tom}, \mathsf{Grandfather}, \mathsf{Tim}\rangle$ cannot be explained. It shows that purely text-based explanations have a rather low coverage for new triples.

We consider combining the above two categories to resolve limitations therein. There are two research problems about this combination — one on the way to combine logical information and text information, and the other on the way to present the explanations.

For the first problem, we propose a new category of explanations, called *abductive text evidences*, to combine logical information and text information. Given a logical theory, a knowledge graph and a text corpus, an abductive text evidence for a new triple τ is a set of pairs $\{(\tau_i, w_i)\}_{1\leq i\leq n}$ for τ_i a triple and w_i a sentence window in the text corpus, such that appending $\{\tau_i\}_{1\leq i\leq n}$ to the knowledge graph makes τ entailed under the logical theory, while w_i explains to some degree the existence of τ_i for all $1 \leq i \leq n$. The set $\{\tau_i\}_{1\leq i\leq n}$ is actually an *abductive explanation* for τ under the logical theory in logic-based abduction (Eiter and Gottlob 1995). The usage of abductive text evidences gives more chances to explain a new triple. Consider the aforementioned example again. Even when none of the sentence windows mentions both Tom and Tim, $\langle \mathsf{Tom}, \mathsf{Grandfather}, \mathsf{Tim} \rangle$ can still be explained by an abductive text evidence $\{(\langle \mathsf{Tom}, \mathsf{Father}, e \rangle, w_1), (\langle e, \mathsf{Father}, \mathsf{Tim} \rangle, w_2)\}$ for e an entity, and w_1 and w_2 sentence windows respectively explaining to some degree why $\langle \mathsf{Tom}, \mathsf{Father}, e \rangle$ and $\langle e, \mathsf{Father}, \mathsf{Tim} \rangle$ exist.

To make explanations as concise as possible, we consider only subset-minimal abductive explanations, simply called *mina-explanations*. Computing the complete set of mina-explanations is still impractical because the number of mina-explanations can be easily exponential in the number of entities. Take for example a knowledge graph with N entities $\{v_i\}_{1 \le i \le N}$ but without triples of the form $\mathsf{edge}(v_i, v_j)$, a new triple $\mathsf{path}(v_1, v_N)$, and a logical theory composed by two rules $\forall x, y, z : \mathsf{path}(x, y) \wedge \mathsf{path}(y, z) \to \mathsf{path}(x, z)$ and $\forall x, y : \mathsf{edge}(x, y) \to \mathsf{path}(x, y)$, then there will be $(N-2)!$ mina-explanations of the form $\{\mathsf{edge}(v_1, v_{k_1}), \ldots, \mathsf{edge}(v_{k_i}, v_{k_{i+1}}), \ldots, \mathsf{edge}(v_{k_{N-2}}, v_N)\}$ for $\mathsf{path}(v_1, v_N)$, where $(k_1, \ldots, k_{N-2})$ is a permutation of $(2, \ldots, N-1)$.

To reduce the number of mina-explanations to be computed, we assume that a pragmatic explanation should be instantiated from an *explanation pattern* seen before and be supported by the text corpus. An explanation pattern P is a set of *lifted triples* which are almost the same as triples except that all entities therein are replaced with variables. P is said to be *seen before* in a knowledge graph for a lifted triple $\langle x, r, y \rangle$ if there is a *differentiated substitution* θ for P making $P\theta$ a minimal subset of the knowledge graph that entails $\langle x\theta, r, y\theta \rangle$ under the logical theory, where a differentiated substitution for P replaces different variables in P with different entities. Moreover, given a text corpus composed of sentence windows, an explanation is required to consist of only triples $\langle h, r, t \rangle$ such that a mention of h and a mention of t co-occur in at least one sentence window, since this kind of explanations can directly be extended to abductive text evidences. Roughly speaking, a mina-explanation for a new triple is called a *text and explanation pattern based explanation* (simply *TEP-based explanation*) if it is derived from an explanation pattern seen before in the given knowledge graph and consists of only triples $\langle h, r, t \rangle$ such that a mention of h and a mention of t co-occur in a sentence window in the given text corpus. Accordingly, we call an abductive text evidence extended from a TEP-based explanation under the same text corpus a *TEP-based abductive text evidence*. We then propose a method for computing the complete set of TEP-based abductive text evidences.

For the second problem, we aim to present TEP-based abductive text evidences in an order that prefers more explicable sentence windows so that anyone who wants to validate a new triple can merely focus on high-rank explanations. To avoid laborious annotation, we adapt a distantly supervised method for relation extraction (Lin et al. 2016) to estimate, for every pair (τ, w) in an abductive text evidence, to what degree the sentence window w explains the existence of the triple τ. We then propose a total order for sorting all TEP-based abductive text evidences that are extended from the same TEP-based explanation.

To evaluate the proposed mechanism for validating new triples, we construct four knowledge graphs and their corresponding logical theories from the text corpora that we are familiar with, which are the four great classical masterpieces of Chinese literature. We divide each knowledge graph into a training set and a test set, where triples in the test set are all treated as new triples. To validate every triple in the test set, we compute all TEP-based abductive text evidences for it and sort them in a variant of the proposed total order. Experimental results show that the computation of TEP-based abductive text evidences is efficient while the proposed total order is more effective than other variants (including the random order) in ranking true explanations at top places.

Preliminaries

This work considers only logical theories that are expressed in first-order logic. Such a logical theory is a set of rules R of the form $\forall \vec{x} : \phi(\vec{x}) \to \exists \vec{y}\ \varphi(\vec{x}, \vec{y})$, where $\phi(\vec{x})$ is a conjunction of atoms on the universally quantified variables $\vec{x}$, and $\varphi(\vec{x}, \vec{y})$ is a disjunction of atoms on both $\vec{x}$ and the existentially quantified variables $\vec{y}$. The part of R at the left (resp. right) of $\to$ is called the *body* (resp. *head*) of R. By $\mathrm{body}(R)$ (resp. $\mathrm{head}(R)$) we denote the set of atoms in the body (resp. head) of R. If the head of R has no atoms, R is also called a *constraint* while the empty head is written as $\perp$. If the head of R has a single atom without existentially quantified variables, R is also called a *datalog* rule. For brevity, in this paper we present our work with the Horn fragment of first-order logic, although the proposed methods can be applied to other fragments such as description logics (DLs) (Baader et al. 2003). We simply call a set of datalog rules and constraints a *Horn theory*, which may contain constants. A knowledge graph can be treated as a set of ground rules with empty bodies, or simply a set of ground atoms, by rewriting triples $\langle h, r, t \rangle$ to ground atoms $r(h, t)$, where entity-type triples of the form $\langle e, \mathsf{type}, t \rangle$ (meaning e has a type t) are also rewritten to $\mathsf{type}(e, t)$ as other triples.

A *model* of a Horn theory $\mathcal{T}$ is a set S of ground atoms such that (1) $\mathrm{body}(R)\,\theta \subseteq S$ implies $\mathrm{head}(R)\,\theta \cap S \neq \emptyset$ for any datalog rule $R \in \mathcal{T}$ and any ground substitution θ for $\mathrm{var}(R)$, and (2) $\mathrm{body}(R)\,\theta \nsubseteq S$ for any constraint $R \in \mathcal{T}$ and any ground substitution θ for $\mathrm{var}(R)$, where a *ground substitution* for a symbol maps all variables in the symbol to constants, and $\mathrm{var}(R)$ denotes the set of variables in R. A model of the union $\mathcal{K}$ of a knowledge graph $\mathcal{G}$ and a Horn theory $\mathcal{T}$ is a model of $\mathcal{T}$ that includes $\mathcal{G}$. The union $\mathcal{K}$ is said to be *consistent* if it has a model, which means that $\mathcal{K}$ has a unique least model since $\mathcal{T}$ is Horn. We say a triple is *entailed* by a knowledge graph $\mathcal{G}$ under a Horn theory $\mathcal{T}$, or *entailed* by $\mathcal{G} \cup \mathcal{T}$, if the triple is in the unique least model of $\mathcal{G} \cup \mathcal{T}$. Given an *observation* which is a triple τ, a knowledge graph $\mathcal{G}$ and a Horn theory $\mathcal{T}$ such that $\mathcal{G} \cup \mathcal{T}$ is consistent, an *abductive explanation* $\mathcal{E}$ for τ in $\mathcal{G}$ under $\mathcal{T}$ is a set of triples constructed from entities in $\mathcal{G}$ such that (1) τ is entailed by $\mathcal{G} \cup \mathcal{E}$ under $\mathcal{T}$, and (2) $\mathcal{G} \cup \mathcal{E} \cup \mathcal{T}$ is consistent. $\mathcal{E}$ is further called a *subset-minimal abductive explanation* (simply *mina-explanation*) for τ in $\mathcal{G}$ under $\mathcal{T}$ if there is no proper subset of $\mathcal{E}$ that is also an abductive explanation for τ in $\mathcal{G}$ under $\mathcal{T}$.

A Validation Mechanism for New Triples

To maintain high quality for a growing knowledge graph, it is crucial to find explanations to prove the existence of any new triple to be added. As discussed in Introduction, although abductive explanations can be used to prove a large number of new triples through a background logical theory, the absent triples in an abductive explanation need to be further validated. Fortunately, text-based explanations can be used to prove these absent triples. Thus we propose below a new kind of explanations which integrates text-based explanations into abductive explanations, where a *sentence window* is a text composed by several consecutive sentences.

Definition 1 (abductive text evidence) Given a triple τ, a knowledge graph $\mathcal{G}$, a Horn theory $\mathcal{T}$ and a text corpus $\mathcal{C}$ composed of sentence windows, such that $\mathcal{G} \cup \{\tau\} \cup \mathcal{T}$ is consistent, an *abductive text evidence* for τ is a set of pairs $\{(\tau_i, w_i)\}_{1 \le i \le n}$ such that $\{\tau_i\}_{1 \le i \le n}$ is an abductive explanation for τ in $\mathcal{G}$ under $\mathcal{T}$, while for all $1 \le i \le n$, $\tau_i \notin \mathcal{G}$, $w_i \in \mathcal{C}$ and w_i explains to some degree the existence of τ_i.

The above definition is not given rigorously as "to some degree" is not well-defined. More importantly, computing all abductive text evidences for a triple is often impractical since it requires to compute all abductive explanations. It has been shown in Introduction that the number of mina-explanations can be up to exponential in the number of entities that appear in the knowledge graph. Hence, we need to introduce a special class of abductive text evidences which is defined rigorously and is easy to compute.

By observing that practical logical reasoning often follows fixed patterns, we assume that practical abduction can be a mimic of deductive reasoning in the same knowledge graph; i.e., every explanation-observation pair used in practice, say $(\mathcal{E}, \tau)$, is copied from some $(\mathcal{E}', \tau')$ by substituting entities and removing entailed triples, where τ' is a triple and $\mathcal{E}'$ is a minimal subset of the knowledge graph that entails τ'. To define such mina-explanations, we introduce a notion of pattern defined below, where a *differentiated substitution* is a ground substitution that maps different variables to different entities, and a *lifted triple* is a triple whose head and tail are both variables.

Definition 2 (SBE-pattern) Given a knowledge graph $\mathcal{G}$, a Horn theory $\mathcal{T}$ and a lifted triple ϱ, a set P of lifted triples is called an *explanation pattern seen before* (simply *SBE-pattern*) in $\mathcal{G}$ for ϱ under $\mathcal{T}$ if there exists a differentiated substitution θ for P such that $P\theta$ is a minimal subset of $\mathcal{G}$ that entails $\varrho\theta$ under $\mathcal{T}$.

Example 1 Let lt be short for less_then. Consider a knowledge graph $\mathcal{G} = \{\langle a, \mathsf{lt}, b\rangle, \langle a, \mathsf{lt}, c\rangle, \langle b, \mathsf{lt}, d\rangle, \langle c, \mathsf{lt}, e\rangle, \langle d, \mathsf{lt}, f\rangle\}$ and a Horn theory $\mathcal{T}$ consisting of two rules $\forall x, y, z : \mathsf{lt}(x, y) \wedge \mathsf{lt}(y, z) \rightarrow \mathsf{lt}(x, z)$ and $\forall x, y : \mathsf{lt}(x, y) \wedge \mathsf{lt}(y, x) \rightarrow \perp$. There are exactly three SBE-patterns in $\mathcal{G}$ for $\langle x, \mathsf{lt}, y\rangle$ under $\mathcal{T}$. They are $P_1 = \{\langle x, \mathsf{lt}, y\rangle\}$, $P_2 = \{\langle x, \mathsf{lt}, v_1\rangle, \langle v_1, \mathsf{lt}, y\rangle\}$ and $P_3 = \{\langle x, \mathsf{lt}, v_1\rangle, \langle v_1, \mathsf{lt}, v_2\rangle, \langle v_2, \mathsf{lt}, y\rangle\}$.

Under our assumption, every desirable mina-explanation should be obtained by applying a differentiated substitution to an SBE-pattern and removing entailed triples from the substituted result. Moreover, every triple τ in a desirable mina-explanation should be explained to some degree by a sentence window. A condition probably fulfilling this requirement is that a mention of the head of τ and a mention of the tail of τ co-occur in a sentence window. This condition may not work well for entity-type triples of the form $\langle e, \mathsf{type}, t\rangle$ since a sentence window explaining $\langle e, \mathsf{type}, t\rangle$ may often not mention any surface names about t. We will leave this problem in our future work. In the current work we introduce a class of desirable mina-explanations defined below according to the above condition, where $\mathrm{lift}(\tau)$ denotes a lifted triple obtained from a triple τ by replacing different entities with different variables, and $\mathrm{h}(\tau)$ and $\mathrm{t}(\tau)$ respectively denote the head and tail of τ.

Definition 3 (TEP-based explanation) Given a triple τ, a knowledge graph $\mathcal{G}$, a Horn theory $\mathcal{T}$, a set $\mathcal{C}$ of sentence windows, and a function f mapping entities in $\mathcal{G}$ to mention sets, such that $\mathcal{G} \cup \{\tau\} \cup \mathcal{T}$ is consistent, a mina-explanation $\mathcal{E}$ for τ in $\mathcal{G}$ under $\mathcal{T}$ is called a *text and explanation pattern based explanation* (simply *TEP-based explanation*) for τ wrt $(\mathcal{G}, \mathcal{T}, \mathcal{C}, f)$ if there is an SBE-pattern P in $\mathcal{G}$ for $\mathrm{lift}(\tau)$ under $\mathcal{T}$ and a differentiated substitution θ for P such that $\mathcal{E} \subseteq P\theta$ and every triple in $P\theta \setminus \mathcal{E}$ is entailed by $\mathcal{G} \cup \mathcal{T}$, while for every triple $\tau' \in \mathcal{E}$, there is a mention $m_h \in f(\mathrm{h}(\tau'))$, a mention $m_t \in f(\mathrm{t}(\tau'))$ and a sentence window $w \in \mathcal{C}$ such that m_h and m_t co-occur in w.

Example 2 Continue with Example 1. Let $\mathcal{C} = \{$"c and d are comparable", "e is smaller than f"$\}$ be a text corpus, and $f(x) = \{x\}$ be a mapping function from entities to mention sets. For the observation $\langle c, \mathsf{lt}, f\rangle$ there are exactly two TEP-based explanations wrt $(\mathcal{G}, \mathcal{T}, \mathcal{C}, f)$. They are $\mathcal{E}_1 = \{\langle c, \mathsf{lt}, d\rangle\}$ and $\mathcal{E}_2 = \{\langle e, \mathsf{lt}, f\rangle\}$. Both of them can be derived from P_2. In fact, there are also other three mina-explanations derivable from the SBE-patterns, namely $\mathcal{E}_3 = \{\langle c, \mathsf{lt}, f\rangle\}$, $\mathcal{E}_4 = \{\langle c, \mathsf{lt}, b\rangle\}$ and $\mathcal{E}_5 = \{\langle e, \mathsf{lt}, d\rangle\}$. But they are not supported by sentence windows in $\mathcal{C}$.

Based on Definition 3, we accordingly define a special class of abductive text evidences below.

Definition 4 (TEP-based abductive text evidence) Given a triple τ, a knowledge graph $\mathcal{G}$, a Horn theory $\mathcal{T}$, a set $\mathcal{C}$ of sentence windows, and a function f mapping entities in $\mathcal{G}$ to mention sets, such that $\mathcal{G} \cup \{\tau\} \cup \mathcal{T}$ is consistent, a *TEP-based abductive text evidence* for τ wrt $(\mathcal{G}, \mathcal{T}, \mathcal{C}, f)$ is a set of pairs $\{(\tau_i, w_i)\}_{1 \le i \le n}$ such that $\{\tau_i\}_{1 \le i \le n}$ is a TEP-based explanation for τ wrt $(\mathcal{G}, \mathcal{T}, \mathcal{C}, f)$, while for all $1 \le i \le n$, $w_i \in \mathcal{C}$ and there is $m_h \in f(\mathrm{h}(\tau_i))$ and $m_t \in f(\mathrm{t}(\tau_i))$ such that m_h and m_t co-occur in w_i.

TEP-based abductive text evidence can be treated as a special class of abductive text evidences, because a sentence window mentioning both the head and the tail of a triple explains to some degree the existence of the triple. This special class is defined rigorously and can be computed in a practical way described by the algorithm in Figure 1.

The undefined notations in Figure 1 are explained as follows. `ComputeEntailments`$(\mathcal{G}, \mathcal{T})$ returns the set of triples entailed by $\mathcal{G}$ under $\mathcal{T}$, which can be rewritten from the unique least model M of $\mathcal{G} \cup \mathcal{T}$ by treating ground atoms as triples since $\mathcal{T}$ is Horn, where M is the least fix-point of

Algorithm. `ComputeExplanations`$(\langle h,r,t\rangle, \mathcal{G}, \mathcal{T}, \mathcal{C}, f)$

Input: An observation $\langle h,r,t\rangle$, a knowledge graph $\mathcal{G}$, a Horn theory $\mathcal{T}$, a set $\mathcal{C}$ of sentence windows, and a function f mapping entities in $\mathcal{G}$ to mention sets.

Output: The complete set of TEP-based abductive text evidences for $\langle h,r,t\rangle$ wrt $(\mathcal{G},\mathcal{T},\mathcal{C},f)$.

1: $\mathcal{G}^+ \leftarrow$ `ComputeEntailments`$(\mathcal{G},\mathcal{T})$;
2: **if** $\langle h,r,t\rangle \in \mathcal{G}^+$ **then return** $\{\emptyset\}$;
3: $\Xi \leftarrow \emptyset$; $\Delta \leftarrow \emptyset$;
4: **for** each $\langle h',r,t'\rangle \in \mathcal{G}^+$ and each minimal subset S of $\mathcal{G}$ that entails $\langle h',r,t'\rangle$ under $\mathcal{T}$ **do** $\Xi \leftarrow \Xi \cup \{\mathrm{lift}(S, h' \mapsto x, t' \mapsto y)\}$;
5: **for** each $P \in \Xi$, each bipartition (P_1,P_2) of P s.t. $P_2 \neq \emptyset$, each $\{x \mapsto h, y \mapsto t\}$-compatible differentiated substitution θ for P_1 s.t. $P_1\theta \subseteq \mathcal{G}^+$, and each θ-compatible differentiated substitution σ for P_2 s.t. $P_2\sigma \cap \mathcal{G}^+ = \emptyset$ **do**
6: **if** $\mathcal{G} \cup P_2\sigma \cup \mathcal{T}$ is consistent
and $\langle h,r,t\rangle \in$ `ComputeEntailments`$(\mathcal{G} \cup P_2\sigma, \mathcal{T})$
and $\langle h,r,t\rangle \notin$ `ComputeEntailments`$(\mathcal{G} \cup S, \mathcal{T})$ for all $S \in \mathrm{sub}_1(P_2\sigma)$ **then**
7: $\Delta \leftarrow \Delta \cup \{\{(\tau_1,w_1),\dots,(\tau_n,w_n)\} \mid \{\tau_1,\dots,\tau_n\} = P_2\sigma, w_1 \in \mathcal{C},\dots,w_n \in \mathcal{C}, \forall 1 \le i \le n\ \exists m_h \in f(\mathrm{h}(\tau_i)), m_t \in f(\mathrm{t}(\tau_i)) : m_h$ and m_t co-occur in $w_i\}$;
8: **return** Δ;

Figure 1: Computing all TEP-based abductive text evidences

$M^{(t)}$ defined by $M^{(0)} = \mathcal{G}$ and for $t > 0$,

$$M^{(t)} = M^{(t-1)} \cup \bigcup_{R \in \mathcal{T}, \mathrm{body}(R)\theta \subseteq M^{(t-1)}} \mathrm{head}(R)\theta.$$

For a set S of triples, $\mathrm{lift}(S, h' \mapsto x, t' \mapsto y)$ returns a set of lifted triples obtained from S by replacing h' with x, t' with y, and other different entities with different variables, whereas $\mathrm{sub}_1(S)$ returns the set of all subsets of S that are obtained from S by removing one triple. Given a differentiated substitution θ, a *θ-compatible* differentiated substitution is a differentiated substitution extended from θ; i.e., it is a superset of θ when θ is treated as a set of replacements. A pair (P_1,P_2) of sets of lifted triples is called a *bipartition* of a set P of lifted triples if $P_1 \cup P_2 = P$ and $P_1 \cap P_2 = \emptyset$.

The algorithm in Figure 1 works as follows. If $\langle h,r,t\rangle$ is entailed by $\mathcal{G}$ under $\mathcal{T}$, there is only an empty TEP-based abductive text evidence for $\langle h,r,t\rangle$ (line 2). Otherwise, the set Ξ of SBE-patterns in $\mathcal{G}$ for $\langle x,r,y\rangle$ under $\mathcal{T}$ is computed (line 4). The SBE-patterns are derived from all minimal subsets of $\mathcal{G}$ entailing $\langle h',r,t'\rangle$ under $\mathcal{T}$ (simply called *justifications* for $\langle h',r,t'\rangle$) for every triple $\langle h',r,t'\rangle$ entailed by $\mathcal{G}$ under $\mathcal{T}$, according to Definition 2. The set of justifications for an entailment can be efficiently computed by the method proposed in (Kalyanpur et al. 2007) and can be further optimized by justification-preserving module extraction (Du, Qi, and Ji 2009) or by the MapReduce technology (Wu, Qi, and Du 2011). Afterwards, the set of TEP-based abductive text evidences for $\langle h,r,t\rangle$ wrt $(\mathcal{G},\mathcal{T},\mathcal{C},f)$ is computed from the set of SBE-patterns according to Definition 3 and Definition 4 (lines 5–7), where the consistency checking in line 6 can also be done by least fix-point computation. The following theorem shows the correctness of the algorithm. Due to the space limitation, the proof of this theorem is moved to our technical report (Du et al. 2018).

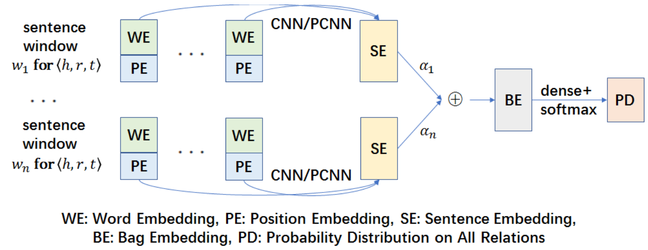

Figure 2: The neural model for relation extraction

Theorem 1 *`ComputeExplanations`$(\langle h,r,t\rangle, \mathcal{G}, \mathcal{T}, \mathcal{C}, f)$ computes the complete set of TEP-based abductive text evidences for $\langle h,r,t\rangle$ wrt $(\mathcal{G},\mathcal{T},\mathcal{C},f)$.*

All TEP-based abductive text evidences can be grouped by the TEP-based explanations that they are extended from. We can present all the TEP-based abductive text evidences group by group, where TEP-based abductive text evidences in the same group are sorted in an order preferring more explicable sentence windows. To do this, we rank all sentence windows that explain the same triple and extend this ranking to sort TEP-based abductive text evidences in the same group. The problem to be solved is: given a triple $\langle h,r,t\rangle$ and a set $\mathcal{C}$ of sentence windows in each of which both a mention of h and a mention of t are marked, how to rank sentence windows in $\mathcal{C}$ by support degrees for $\langle h,r,t\rangle$?

We consider the probability $P(r|\langle h,r,t\rangle, w)$ as an estimation of the support degree of a sentence window w for $\langle h,r,t\rangle$, and adapt a distantly supervised method for relation extraction (Lin et al. 2016) to compute it. Without annotation on whether a sentence window truly prove the existence of $\langle h,r,t\rangle$, this method creates training data under the assumption that all sentence windows mentioning both h and t contribute to the proof of $\langle h,r,t\rangle$ more or less. Then the method builds a neural model shown in Figure 2 to compute $P(r|\langle h,r,t\rangle, w)$. It treats all sentence windows mentioning both h and t as a *bag* for every training triple $\langle h,r,t\rangle$.

The model consists of three layers. The first layer, called *sentence encoder*, builds an *embedding* (i.e. a real-value vector) for every sentence window. It concatenates the word embedding and position embedding for every word in the given sentence window, where the position embedding has two parts — one corresponds to the distance to the marked mention of h and the other corresponds to the distance to the marked mention of t, and then aggregates all concatenated embeddings to form a sentence embedding with a fixed dimension d through a convolutional neural network (CNN) or a piecewise convolutional neural network (PCNN). From now on, let $\overrightarrow{s_{h,t,w}} \in \mathbb{R}^d$ denote the sentence embedding constructed from a sentence window w wrt h and t. We refer the interested reader to (Lin et al. 2016) for more details on computing $\overrightarrow{s_{h,t,w}}$. The second layer, called *attention layer*, aggregates all sentence embeddings in the same bag to form a bag embedding by applying an attention mechanism. Let $\overrightarrow{a_r} \in \mathbb{R}^d$ denote the attention vector for relation r, which needs to be learnt also. The bag embedding $\vec{b}$ for a bag of sentence windows $w_1, \ldots, w_n$ can be defined as

$$\vec{b} = \sum_{i=1}^{n} \alpha_i \overrightarrow{s_{h,t,w_i}}, \tag{1}$$

$$\alpha_i = \exp(c_i) / \sum_{j=1}^{n} \exp(c_j), \tag{2}$$

$$c_i = \overrightarrow{s_{h,t,w_i}} \cdot \overrightarrow{a_r}, \tag{3}$$

where $\vec{x} \cdot \vec{y}$ is the dot product of two vectors $\vec{x}$ and $\vec{y}$. The last layer, called *classification layer*, converts a bag embedding into a probability distribution on all relations through a dense sublayer and a softmax sublayer. Formally, by introducing a matrix $\mathbf{W} \in \mathbb{R}^{m \times d}$ and a vector $\vec{v} \in \mathbb{R}^m$ for m the number of different relations, the probability for the i^{th} relation r_i, written $P(r_i|h, t, w_1, \ldots, w_n)$, can be defined as

$$P(r_i|h, t, w_1, \ldots, w_n) = \exp([\vec{u}]_i) / \sum_{j=1}^{m} \exp([\vec{u}]_j), \tag{4}$$

$$\vec{u} = \mathbf{W}\vec{b} + \vec{v}. \tag{5}$$

In the training phase, all training triples $\langle h_i, r_i, t_i \rangle$ ($1 \leq i \leq N$) are organized as N bags $\{(\langle h_i, r_i, t_i \rangle, w_1, \ldots, w_{n_i})\}_{1 \leq i \leq N}$, where $w_1, \ldots, w_{n_i}$ are all sentence windows mentioning both h_i and t_i, and then the target function $\sum_{i=1}^{N} P(r_i|h_i, t_i, w_1, \ldots, w_{n_i})$ is maximized. In the prediction phase, given a test triple $\langle h, r, t \rangle$ and a sentence window w in which both a mention of h and a mention of t are marked, the probability $P(r|\langle h, r, t \rangle, w)$ is computed from the learnt model by ignoring the attention layer; i.e.,

$$P(r|\langle h, r, t \rangle, w) = \exp([\overrightarrow{u'}]_{i_r}) / \sum_{j=1}^{m} \exp([\overrightarrow{u'}]_j), \tag{6}$$

$$\overrightarrow{u'} = \mathbf{W}\overrightarrow{s_{h,t,w}} + \vec{v}, \tag{7}$$

where i_r denotes the index of r among all relations.

It is inappropriate to rank sentence windows for $\langle h, r, t \rangle$ directly by $P(r|\langle h, r, t \rangle, w)$, because $P(r|\langle h, r, t \rangle, w)$ has been normalized without considering other sentence windows. Instead, the support degree of a sentence window w for $\langle h, r, t \rangle$ can be better estimated by its contribution to the proof of $\langle h, r, t \rangle$ (i.e. $\overrightarrow{s_{h,t,w}} \cdot \overrightarrow{a_r}$) as defined by Formula (3).

Let $\gamma(\langle h, r, t \rangle, w, \mathcal{C})$ denote the position of w in the sorted list of all sentence windows in $\mathcal{C}$ mentioning both h and t, which is sorted in the descending order of $\overrightarrow{s_{h,t,w}} \cdot \overrightarrow{a_r}$. We define a precedence order $\preceq_{\mathsf{TEP}}$ on two TEP-based abductive text evidences $\{(\tau_i, w_i)\}_{1 \leq i \leq n}$ and $\{(\tau_i, w'_i)\}_{1 \leq i \leq n}$ as

$$\begin{aligned} & \{(\tau_i, w_i)\}_{1 \leq i \leq n} \preceq_{\mathsf{TEP}} \{(\tau_i, w'_i)\}_{1 \leq i \leq n} \\ \iff & \sum_{i=1}^{n} \gamma(\tau_i, w_i, \mathcal{C}) < \sum_{i=1}^{n} \gamma(\tau_i, w'_i, \mathcal{C}) \text{ or} \\ & (\sum_{i=1}^{n} \gamma(\tau_i, w_i, \mathcal{C}) = \sum_{i=1}^{n} \gamma(\tau_i, w'_i, \mathcal{C}) \text{ and} \\ & \sum_{i=1}^{n} P(r_i|\tau_i, w_i) \geq \sum_{i=1}^{n} P(r_i|\tau_i, w'_i)), \end{aligned} \tag{8}$$

where r_i is the relation of τ_i. Since $\preceq_{\mathsf{TEP}}$ is antisymmetric, transitive and has the connex property, it is a total order.

Experimental Evaluation

We implemented the proposed methods for computing and ranking TEP-based abductive text evidences in Java, where the implementation of the method proposed in Figure 1 was optimized by module extraction (Du, Qi, and Ji 2009) and by caching sentence windows for every head-tail pair handled in line 7. There are two proposed ranking methods based on the neural model shown in Figure 2, where one (denoted *CNNRP*) uses CNN to encode sentence windows and the other uses PCNN (denoted *PCNNRP*) to encode sentence windows. Both ranking methods employ Adam (Kingma and Ba 2014) as the stochastic optimization algorithm in the training course. To compare different ranking strategies, we also implemented three other ranking methods. The first two use only probability information to rank TEP-based abductive text evidences; i.e., in contrast to Formula (8) they define

$$\begin{aligned} & \{(\tau_i, w_i)\}_{1 \leq i \leq n} \preceq_{\mathsf{TEP}} \{(\tau_i, w'_i)\}_{1 \leq i \leq n} \\ \iff & \sum_{i=1}^{n} P(r_i|\tau_i, w_i) \geq \sum_{i=1}^{n} P(r_i|\tau_i, w'_i)). \end{aligned} \tag{9}$$

Among these two methods, one (denoted *CNNP*) uses CNN to encode sentence windows and the other uses PCNN (denoted *PCNNP*). The last method (denoted *Rand*) simply ranks TEP-based abductive text evidences in a random order.

Data Construction

Existing benchmark knowledge graphs do not have corresponding logical theories or text corpora from which they are extracted. It is hard to add adequate rules to these datasets and seek related text corpora for them. Thus we constructed new knowledge graphs and the corresponding logical theories from existing text corpora in a domain that we are familiar with. The domain is about character relationships in the four great classical masterpieces of Chinese literature, namely *Dream of the Red Chamber* (*DRC*), *Journey to the West* (*JW*), *Outlaws of the Marsh* (*OM*), and *Romance of the Three Kingdoms* (*RTK*). We collected triples on character relationships from e-books for these masterpieces, yielding four knowledge graphs each of which corresponds to one masterpiece. For every knowledge graph, we collected all human entities in it and sought surface names for every entity by checking the Web page of that entity in Baidu Wikipedia[1]. Thus we got a set of mentions (i.e. surface names) for every entity in the knowledge graphs. For each e-book, we separated it into sentences and composed a sentence window for every three consecutive sentences since we statistically found that most collected triples can be explained by consecutive three sentences. Finally, we manually built a logical theory for modeling character relationships in one masterpiece by Protege[2], a well-known ontology editor. Every logical theory is originally expressed in OWL 2 RL (Grau et al. 2008), a tractable profile of OWL 2 for modeling ontologies, and then translated to a Horn theory by standard transformation. Every constructed OWL 2

[1]https://baike.baidu.com/

[2]https://protege.stanford.edu/

RL ontology is rather complex. It contains transitivity axioms, such as one axiom declaring that relatives are transitive, as well as property chain axioms, such as another axiom declaring that daughters in law are wives of sons.

We divided each knowledge graph into a training set and a test set, where every triple in the test set is treated as an observation whose TEP-based abductive text evidences are to be computed. Initially the training set is the complete knowledge graph and the test set is empty. Afterwards the test set is enlarged by randomly picking out triples from the training set in turn, until the cardinality of the test set reaches one-tenth of the cardinality of the knowledge graph or there is no triple that can be picked out, where a triple picked out should be constructed from entities and relations in the remainder of the training set and cannot be entailed by the remainder of the training set under the corresponding logical theory.

To evaluate ranking methods, the training set of a knowledge graph was extended by joining sentence windows that contain a mention of the head and a mention of the tail for every triple in it. Afterwards the training set was modified by deleting records in which the relations are not used in the test set, and then by adding records with unknown relations for every head-tail pair that has no relation in the training set and every sentence window containing a mention of the head and a mention of the tail. We call the resulting set the *TE-training set*, where TE is short for "text enhanced". We treated the training set as the input knowledge graph and computed the complete set of TEP-based abductive text evidences for every triple in the test set. We call the set of triple-window pairs each of which is contained in at least one computed TEP-based abductive text evidence the *TW-test set*. We carefully checked every triple-window pair in the TW-test set and marked it with a Boolean flag to indicate if the sentence window on the right truly explains the existence of the triple on the left. Then a computed TEP-based abductive text evidence is defined as *true* if all triple-window pairs in it are marked true in the TW-test set. Table 1 reports the statistics about all of our constructed datasets.[3]

Table 1: The statistics about the constructed datasets

Dataset	#rel	#ent	#train	#test	#win	#TE-train
DRC	45	388	333	38	34,530	211,169
JW	21	104	106	3	27,670	19,006
OM	38	156	178	19	34,010	298,403
RTK	30	123	132	14	29,817	77,876

Dataset	#dlog	#cons	#expl	#ATE	#true(nt)	#TW-test
DRC	110	107	63	3,140	57(15)	1,343
JW	94	30	15	1,809	59(0)	1,245
OM	111	47	122	6,989	49(4)	2,200
RTK	142	63	97	25,973	184(55)	4,013

Note: #rel/#ent are respectively the number of relations/entities in the knowledge graph, #train/#test are respectively the number of triples in the training/test set, #win is the number of sentence windows, #TE-train/ #TW-test are respectively the number of records in the TE-training/TW-test set, #dlog/#const are respectively the number of datalog rules/constraints in the Horn theory, #expl/#ATE/#true(nt) are respectively the number of TEP-based explanations/TEP-based abductive text evidences/true TEP-based abductive text evidences (true ones extended from nontrivial TEP-based explanations).

Experimental Results

The computation of TEP-based abductive text evidences was conducted in a laptop with 8GB memory and 2-core 2.5GHz CPU. The execution time for computing all TEP-based abductive text evidences for all observations ranges from 4 seconds (for JW) to 28 seconds (for OM). It can be seen from Table 1 that, for all datasets, the set of TEP-based explanations contains *nontrivial* ones that are not the given observation itself, while for all datasets but JW, there exist true TEP-based abductive text evidences that are extended from nontrivial TEP-based explanations. This means that the proposed mechanism is able to find true explanations combing both logical information and text information as expected.

The ranking of TEP-based abductive text evidences by our proposed method (CNNRP/PCNNRP) or a variant (CNNP/PCNNP) needs to learn a neural model in Figure 2 from the TE-training set and then apply the learnt model to compute certain values used in Formula (8) or Formula (9). These experiments were conducted in a workstation with 64GB memory and 28-core 2.2GHz CPU. In order to learn and apply a neural model, we split a sentence window into a set of mentions of human entities and singleton characters in the remaining part. We did not split a sentence window into words because existing word segmentation tools do not work well for ancient Chinese sentences in the masterpieces. We treated every split unit as a word, where the word embedding is randomly instantiated and learnt during training of the neural model. For training the model we uniformly set the dimension for word embeddings as 100, the dimension for position embeddings as 10, the dimension for sentence embeddings as 100 for CNN and 150 for PCNN, the window size in CNN/PCNN as 2, the initial learning rate as 0.001 for Adam (Kingma and Ba 2014), the probability for applying dropout (Srivastava et al. 2014) as 0.1, and the learning epochs as 20. The execution time for learning a neural model ranges from 7 minutes (for JW) to 18 hours (for OM). On the contrary, applying the learnt model to compute ranks for all TEP-based abductive text evidences is done in 2 seconds for every dataset.

To compare the five methods for ranking each group of TEP-based abductive text evidences that have the same TEP-based explanation, we introduce two metrics from the field of Information Retrieval. The first metric, written $\mathrm{Rank}_{\min}$, is defined as the minimum rank of true TEP-based abductive text evidences in the sorted list of a group of TEP-based abductive text evidences. The smaller $\mathrm{Rank}_{\min}$ is, the earlier people can see true explanations, thus the better the sorted list is. The second metric, written NDCG (short for Normalize Discounted Cumulative Gain), is defined as the ratio of

[3]All datasets mentioned in Table 1 and our constructed OWL 2 RL ontologies are available at http://dataminingcenter.net/papers/AAAI-19-data.zip.

Table 2: The performance measured by average $\text{Rank}_{\min}$

Dataset	CNNRP	PCNNRP	CNNP	PCNNP	Rand
DRC	12.714	**6.524**	25.095	20.095	34.331
JW	**4.000**	4.500	19.500	22.500	10.333
OM	**3.825**	4.875	6.173	6.753	6.361
RTK	**4.786**	6.071	8.286	6.571	9.333

Table 3: The performance measured by average NDCG

Dataset	CNNRP	PCNNRP	CNNP	PCNNP	Rand
DRC	0.553	**0.595**	0.535	0.512	0.327
JW	**0.543**	0.541	0.472	0.480	0.504
OM	**0.596**	0.571	0.529	0.491	0.543
RTK	**0.664**	0.614	0.607	0.602	0.417

the discounted cumulative gain (DCG) of the current sorted list to the DCG of the ideal sorted list where all true TEP-based abductive text evidences are ranked at top places. The value of NDCG is ranged from 0 to 1. The larger NDCG is, the closer the sorted list is to the ideal one, thus the better the sorted list is. Formally, NDCG is defined as

$$\mathsf{NDCG} = \frac{\mathsf{DCG}}{\mathsf{IDCG}} = \frac{\sum_{i=1}^{n} \frac{\mathsf{rel}_i}{\log_2 i+1}}{\sum_{i=1}^{m} \frac{1}{\log_2 i+1}}, \tag{10}$$

where n is the number of TEP-based abductive text evidences, m is the number of true TEP-based abductive text evidences, and $\mathsf{rel}_i = 1$ if the i^{th} TEP-based abductive text evidence in the sorted list is true or $\mathsf{rel}_i = 0$ otherwise.

Table 2 and Table 3 report the average $\text{Rank}_{\min}$ and the average NDCG, respectively, for all groups of TEP-based abductive text evidences, where the best value achieved by all compared methods is displayed in bold and the value for the random method (Rand) has been averaged by 10 random runs. It can be seen that, ranking sentence windows by their contributions to the proof of the corresponding relation which are computed by an attention mechanism is much effective than random ranking or ranking by their probabilities on inferencing the corresponding relation, in giving true TEP-based abductive text evidences high-ranks. Ranking by probabilities is sometimes worse than random ranking, because the computed probabilities have been normalized locally and the comparison result between two normalized probabilities for a pair of sentence windows may not truly reflect the comparison result between two support degrees for the same pair of sentence windows.

Related Work

This work is most related to those studies addressing the automatic seeking of external evidences to validate triples in a knowledge graph. The DeFacto (Deep Fact Validation) approach (Gerber et al. 2015) validates triples by finding trustworthy sources for them on the Web. It transforms triples into natural language sentences and retrieves web pages mentioning these sentences by a web search engine. The TTA (Triples Accuracy Assessment) approach (Liu, d'Aquin, and Motta 2017) validates triples by finding consensus of matched triples from other knowledge graphs. Either of the above approaches exploits evidences of a single type and does not consider logical information. As far as we know, our proposed approach is the first one that combines logical information and textual information to validate triples.

By treating a knowledge graph as an ABox and a logical theory as a TBox, the problem of computing abductive explanations for a triple can be viewed as a problem of ABox abduction (Elsenbroich, Kutz, and Sattler 2006) in description logics (DLs) (Baader et al. 2003) which are fragments of first-order logic. ABox abduction aims to compute *explanations* for an *observation*, where the observation is usually an ABox assertion which can be treated as a triple, and an explanation is a set of ABox assertions whose addition to a consistent DL knowledge base (KB) made up of a TBox and an ABox enforces entailment of the observation while keeping the KB consistent. Existing methods for ABox abduction are restricted to specific DLs. An early study for ABox abduction (Peraldi et al. 2007) proposes a backward inference method. It restricts the TBox axioms to specific forms and does not guarantee minimality for the output explanations. To compute all minimal explanations in an $\mathcal{ALC}$ KB, the study (Klarman, Endriss, and Schlobach 2011) proposes a method based on resolution and tableau. A subsequent study (Halland and Britz 2012) proposes a purely tableaux-based method for ABox abduction in $\mathcal{ALC}$. Another study (Ma et al. 2012) extends the method proposed in (Klarman, Endriss, and Schlobach 2011) to work for $\mathcal{ALCI}$ KBs. All the above methods do not guarantee termination in computing all minimal explanations. In (Du et al. 2011) the termination problem is tackled by restricting explanations to be constructed from a finite vocabulary. A sound and complete method is accordingly proposed to compute all minimal explanations in Horn-$\mathcal{SHIQ}$ KBs. It is later extended by (Wang et al. 2015) to handle OWL 2 EL KBs. In (Du, Wang, and Shen 2014) a tractable method for ABox abduction is proposed. It only works for KBs that are first-order rewritable. In (Del-Pinto and Schmidt 2017) a forgetting based method for ABox abduction in $\mathcal{ALC}$ is proposed. In contrast to the above methods, we propose a novel method for computing certain minimal explanations, where SBE-patterns are introduced to confine the structure of minimal explanations. It is applicable to many fragments of first-order logic including DLs as long as a reasoner for the fragment exists to perform consistency checking or entailment checking.

This work addresses not only how to compute abductive explanations but also how to rank abductive text evidences that are extended from abductive explanations. The proposed ranking methods rely on an attention mechanism for weighting different sentence windows that contribute to explaining a given triple. Hence any attention-based method for distantly supervised learning of relation extraction models, besides (Lin et al. 2016), can be employed in this work. We refer the interested reader to (Kumar 2017) for more neural models for relation extraction.

Conclusions and Future Work

To validate new triples in a growing knowledge graph, we have proposed a new kind of explanations namely abductive text evidence for new triples. Both logical information and text information are combined in an abductive text evidence to prove the existence of a new triple. From the angle of practice, we proposed a special class of abductive text evidences namely TEP-based abductive text evidence which is easy to compute. Accordingly, we proposed a method for computing the complete set of TEP-based abductive text evidences. Moreover, we proposed a method for sorting TEP-based abductive text evidences based on distantly supervised learning. The efficiency and effectiveness of our proposed methods have been demonstrated in our experiments.

As mentioned before, a TEP-based abductive text evidence is hard to contain any entity-type triple of the form $\langle e, \mathsf{type}, t\rangle$ although entity-type triples can be used as other triples in our proposed computational framework. In our future work we plan to define other classes of abductive text evidences that can contain entity-type triples as required. In addition, we plan to develop more elaborate ranking methods to give true abductive text evidences higher ranks.

Acknowledgements

This work was partly supported by National Natural Science Foundation of China (61375056 and 61876204), Guangdong Natural Science Foundation (2018A030313777), Science and Technology Program of Guangzhou (201804010496), Scientific Research Innovation Team in Department of Education of Guangdong (2017KCXTD013), IBM Faculty Award and the EU Marie Currie K-Drive project (286348).

References

Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D.; and Patel-Schneider, P. F., eds. 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.

Del-Pinto, W., and Schmidt, R. A. 2017. Forgetting-based abduction in ALC. In *Proceedings of the Workshop on Second-Order Quantifier Elimination and Related Topics (SOQE)*, 27–35.

Du, J.; Qi, G.; Shen, Y.; and Pan, J. Z. 2011. Towards practical ABox abduction in large OWL DL ontologies. In *Proceedings of the 25th National Conference on Artificial Intelligence (AAAI)*, 1160–1165.

Du, J.; Pan, J. Z.; Wang, S.; Qi, K.; Shen, Y.; and Deng, Y. 2018. Validation of growing knowledge graphs by abductive text evidences. Technical report, Guangdong University of Foreign Studies. http://dataminingcenter.net/papers/AAAI-19-TR.pdf.

Du, J.; Qi, G.; and Ji, Q. 2009. Goal-directed module extraction for explaining OWL DL entailments. In *Proceedings of the 8th International Semantic Web Conference (ISWC)*, 163–179.

Du, J.; Wang, K.; and Shen, Y. 2014. A tractable approach to ABox abduction over description logic ontologies. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, 1034–1040.

Eiter, T., and Gottlob, G. 1995. The complexity of logic-based abduction. *Journal of the ACM* 42(1):3–42.

Elsenbroich, C.; Kutz, O.; and Sattler, U. 2006. A case for abductive reasoning over ontologies. In *Proceedings of the 3rd OWLED Workshop on OWL: Experiences and Directions*.

Gerber, D.; Esteves, D.; Lehmann, J.; Bühmann, L.; Usbeck, R.; Ngomo, A. N.; and Speck, R. 2015. Defacto - temporal and multilingual deep fact validation. *Journal of Web Semantics* 35:85–101.

Grau, B. C.; Horrocks, I.; Motik, B.; Parsia, B.; Patel-Schneider, P. F.; and Sattler, U. 2008. OWL 2: The next step for OWL. *Journal of Web Semantics* 6(4):309–322.

Halland, K., and Britz, K. 2012. ABox abduction in ALC using a DL tableau. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists Conference (SAICSIT)*, 51–58.

Kalyanpur, A.; Parsia, B.; Horridge, M.; and Sirin, E. 2007. Finding all justifications of OWL DL entailments. In *Proceedings of the 6th International Semantic Web Conference (ISWC)*, 267–280.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *CoRR* abs/1412.6980.

Klarman, S.; Endriss, U.; and Schlobach, S. 2011. ABox abduction in the description logic $\mathcal{ALC}$. *Journal of Automated Reasoning* 46(1):43–80.

Kumar, S. 2017. A survey of deep learning methods for relation extraction. *CoRR* abs/1705.03645.

Lin, Y.; Shen, S.; Liu, Z.; Luan, H.; and Sun, M. 2016. Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2124–2133.

Liu, S.; d'Aquin, M.; and Motta, E. 2017. Measuring accuracy of triples in knowledge graphs. In *Proceedings of the 1st International Conference on Language, Data, and Knowledge (LDK)*, 343–357.

Ma, Y.; Gu, T.; Xu, B.; and Chang, L. 2012. An ABox abduction algorithm for the description logic ALCI. In *Proceedings of the Intelligent Information Processing VI - 7th IFIP TC International Conference (IIP)*, 125–130.

Peraldi, I.; Kaya, A.; Melzer, S.; Möller, R.; and Wessel, M. 2007. Towards a media interpretation framework for the semantic web. In *Proceedings of 2007 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 374–380.

Srivastava, N.; Hinton, G. E.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15(1):1929–1958.

Wang, Z.; Chitsaz, M.; Wang, K.; and Du, J. 2015. Towards scalable and complete query explanation with OWL 2 EL ontologies. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM)*, 743–752.

Wu, G.; Qi, G.; and Du, J. 2011. Finding all justifications of OWL entailments using TMS and mapreduce. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM)*, 1425–1434.