Knowledge Graph Embeddings with Ontologies: Reification for Representing Arbitrary Relations

Mena Leemhuis[1][0000−0003−1017−8921], Özgür L. Özçep[1][0000−0001−7140−2574], and Diedrich Wolter[2][0000−0001−9185−0147]

[1] University of Lübeck, Germany
{leemhuis,oezcep}@ifis.uni-luebeck.de
[2] University of Bamberg, Germany
diedrich.wolter@uni-bamberg.de

Abstract. Knowledge graph embeddings offer prospects to integrate machine learning and symbolic reasoning. Learning algorithms are designed that map constants, concepts, and relations to geometric entities in a real-valued domain $\mathbb{R}^n$. By identifying logics that feature these geometric entities as their model, one is able to achieve a tight integration of logic reasoning with machine learning. However, interesting description logics are more expressive than current knowledge graph embeddings, as description logics allow concept definitions using arbitrary relations, such as non-functional relationships and partial ones. By contrast, geometric models of relations used so far in knowledge graph embeddings such as translations, rotations, or linear functions can only represent total functional relationships. In this paper we describe a new geometric model of the description logic $\mathcal{ALC}$ based on cones that exploits reification combined with linear functions to represent arbitrary relations. While this paper primarily describes reification in context of a particular model for $\mathcal{ALC}$, the proposed reification technique is general and applicable with other ontology languages and knowledge graph embeddings.

Keywords: Knowledge-Graph Embedding· Ontologies · Explainable AI.

1 Introduction

Knowledge graph embeddings (KGEs) offer prospects of a true integration of machine learning and symbolic reasoning, given that models acquired by means of machine learning can also serve as models in a logic sense. Until now, several properties of prominent ontology languages for representing non-trivial concepts are beyond what can be grounded in machine learning models. In this paper we develop an approach using *reification* to advance the expressiveness of relations in KGEs. In knowledge graph embeddings, concepts are commonly represented as geometric entities (e.g., balls [9], boxes [14], or cones [13,18,2]), constants as points, and relations as geometric operations. TransE [3] continues to be the classical reference for a knowledge graph embedding of relations, drawing its charm from a simple geometrical representation of relations that can be learned efficiently. Indeed, TransE represents relations as vector translations, and hence embedding a triple $(s\ R\ o)$ (stating that a subject s stands in relation R to an object o, also written $R(s, o)$) into a continuous space is easily integrated with a loss function used for learning the embedding. The downside of this simple representation is TransE's limited expressivity [11]: only binary relations that are functional in their first argument can be modeled. Such considerations on expressivity have lead to many other embedding approaches that rely on more complex representations of relations. Most of them are in the tradition of TransE—some more such as TransH [17], TransR [10] that rely on representing relations as translation, and some less, relying on other, more involved geometrical operations such as SimplE [11] or Rescal [12].

Still, all these approaches share the limitation of being restricted to relations that are total and functional. Important features like partiality or non-functionality of relations cannot be modeled correctly. What we mean by correct modeling is not going beyond acceptable performance in some combinations of datasets and tasks, but to give a proper logic-like Tarskian semantics to relations. By doing so, one does not only pave a more solid theoretical foundation but also establishes the basis for KGEs associated with a background ontology which states axiomatically constraints on the entities, concepts, and relations to be embedded. For example, a background theory may state that certain concepts are mutually exclusive (e.g., `familyMovie` and `horrorMovie` are disjoint) or that some elements of a concept are related (e.g., `hasChild` as a partial, non-functional relation on the concept class `human`). In [7] this logic-like representation is expressed as the suggestion to represent (subjects and objects in triples) as vectors (this is represented as arbitrary n-ary relations as subsets of the n-cartesian product over the embedding space). An obvious downside of that approach is the high increase in dimensionality required with adverse effects on learning. In this paper we take a middle-road: We insist on representing concepts (unary relations) as sets of vectors but allow representations of arbirary binary relations in mathematically well-behaved operations.

In this paper we propose the idea of *reification* to be applied in KGEs to represent relations. We adopt the idea of relying on matrix multiplication to represent relations previously used in KGEs, but we rewrite relations into equivalent structures which allows us to model arbitrary relations, including partial and non-functional ones. The idea of reification is to represent relations as objects in the embedding space. In our case, e.g., a triple $(a\ R\ b)$ would be represented by an object $c_{R(a,b)}$ and functions stating that its "arguments" are a and b: $\pi_{1,R}(c_{R(a,b)}) = a$, $\pi_{2,R}(c_{R(a,b)}) = b$. The only relations $\pi_{i,R}$ that have to be represented and learnt, then, are functional relations, namely projections of triples to its subject and its object. In context of an approach to cone-based embeddings we are able to show that the set of triples $c_{R(a,b)}$ for a particular relation R forms a well-behaved object too, namely a cone itself. In order to achieve partiality, we develop a semantics that allows the projections π_1, π_2 to project pairs (a, b) outside the domain, thus representing non-existence.

While reification is a well-known approach in logic modeling, the technical challenge tackled in this paper is to develop a geometric model that can link machine learning (by using feasible ingredients such as convex sets and simple geometric operations) and ontologies (by defining a model for a logic). While we believe the idea of reification to be compatible with a range of approaches in KGEs, we have opted in this paper to extend an approach using linear functions as projections similar to TransR [10] as building blocks for relations and convex cones [13] which have already been shown to support full concept negation in background knowledge. Taken together, we can give a feasible geometric model for the well-known description logic $\mathcal{ALC}$ and thereby advance expressivity of KGEs.

In summary, the key contribution of this paper is to give an embedding of $\mathcal{ALC}$-ontologies based on generalized cones with a novel interpretation of relations based on reification having the ability of representing partial and non-functional relations.

The remainder of this paper is structured as follows. In Section 2 we introduce notation and summarize relevant properties of ontologies. Section 3 presents the proposed reification approach for cones. In Section 4 we show how a geometric model for the description logic $\mathcal{ALC}$ can be constructed and discuss its properties. Section 5 presents related works. The paper concludes with a brief résumé and outlook.

2 Preliminaries

In order to define embeddings of knowledge graph triples with respect to background knowledge, we first introduce a suitable language to model such background knowledge, usually referred to as ontology. Description logics (DLs) [1] are formal languages tailored towards representing ontologies and they thus present themselves as a basis. DLs provide a clear distinction between factual knowledge (expressed in the so-called abox) and terminological knowledge (expressed in the so-called tbox). In context of KGEs, the abox provides specific data instances and the tbox provides background knowledge. We note that one may be interested in semantics beyond classical DLs and as we will discuss later this is indeed possible. For example, one may be interested in some settings to account for partial information, say there may be elements that are neither known to be members of a concept C nor of its negation $C^{\perp}$. This may be accomplished by choosing the appropriate semantics. To keep our approach general, we first describe semantics for a very general orthologic and then refine it to the classic semantics of the well-known and widely used DL $\mathcal{ALC}$.

2.1 Ortholattice and Orthologic

In short, ortholattices are structures similar to Boolean algebras but with fewer properties, e.g., no distributivity.

An *(algebraic) ortholattice* is a partially ordered set L with functions defined on it, namely a structure $(L, \wedge, \vee, \cdot^{\perp}, 0, 1)$ fulfilling the following properties:

- $a \vee a = a$, $a \wedge a = a$. (idempotence)
- $a \vee b = b \vee a$, $a \wedge b = b \wedge a$. (commutativity)
- $(a \vee b) \vee c = a \vee (b \vee c)$, $(a \wedge b) \wedge c = a \wedge (b \wedge c)$ (associativity)
- $a \vee (a \wedge b) = a$, $a \wedge (a \vee b) = a$ (absorption)
- $a \wedge 0 = 0$, $a \vee 0 = a$, $a \vee 1 = 1$, $a \wedge 1 = a$
- $a^{\perp\perp} = a$ (double negation elimination)
- $0 = a \wedge a^{\perp}$ (intuitionistic absurdity)
- $(a \vee b)^{\perp} = a^{\perp} \wedge b^{\perp}$, $(a \wedge b)^{\perp} = a^{\perp} \vee b^{\perp}$ (De Morgan)

Intuitively, $\cdot^{\perp}$ represents negation, more precisely called orthocomplement in orthologics, and partial order $\leq$ on L corresponds to concept inclusion $\sqsubseteq$ in the ontology language. Logics defined on ortholattices (as opposed to Boolean algebras) are called *orthologics*. Classical logics like propositional logics are also orthologics, albeit ones that satisfy additional, stronger properties.

2.2 Background Logic

As the basic logical syntax we consider that of $\mathcal{ALC}$ [1]. The syntax of $\mathcal{ALC}$ promises to provide operators to express non-trivial background knowledge. This language is neither trivial nor too complex to distract from developing the main points in this paper. The $\mathcal{ALC}$ syntax rests on a DL vocabulary $\mathcal{V}$ given by a set of constants N_c, a set of role names (binary relation symbols) N_R, and a set of concept names N_C. The set $conc(\mathcal{V})$ of $\mathcal{ALC}$ concepts (concept descriptions) over $N_C \cup N_R$ is described by the grammar

$$C \longrightarrow A \mid \bot \mid \top \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \exists R.C \mid \forall R.C$$

where $A \in N_C$ is an atomic concept, $R \in N_R$ is a role symbol, and C stands for arbitrary concepts. An *ontology* $\mathcal{O}$ is defined as a pair $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ of a *terminological box (tbox)* $\mathcal{T}$ and an *assertional box (abox)* $\mathcal{A}$. A tbox consists of *general inclusion axioms* $C \sqsubseteq D$ ("C is subsumed by D") with concept descriptions C, D. For ease of notation, we write $C = D$ instead of $C \sqsubseteq D$ and $D \sqsubseteq C$. An abox consists of a finite set of *assertions*, i.e., facts of the form $C(a)$ or of the form $R(a, b)$ for $a, b \in N_c$. We define the notion of an interpretation as usual:

Definition 1. *A structure* $(\Delta, \cdot^{\mathcal{I}})$ *is called an* interpretation $\mathcal{I}$ *for a given* $\mathcal{ALC}$ *vocabulary of constants, concept and role symbols* $\mathcal{V} = N_c \cup N_C \cup N_R$ *iff* Δ*, the so-called* domain*, is a set and* $\cdot^{\mathcal{I}}$ *is the* denotation function *defined for all* $b \in N_c, A \in N_C, R \in N_R$ *and concepts* C, D *over* $\mathcal{V}$ *such that the following conditions are fulfilled:*

$b^{\mathcal{I}} \in \Delta$,	$(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$,
$A^{\mathcal{I}} \subseteq \Delta$,	$(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$,
$R^{\mathcal{I}} \subseteq \Delta \times \Delta$,	$(\neg C)^{\mathcal{I}} = \Delta \setminus C^{\mathcal{I}}$,
$\top^{\mathcal{I}} = \Delta$,	$(\exists R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid$ *There is* $y \in \Delta^{\mathcal{I}}$ *s.t.* $(x, y) \in R^{\mathcal{I}}$ *and* $y \in C^{\mathcal{I}}\}$,
$\bot^{\mathcal{I}} = \emptyset$,	$(\forall R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid$ *For all* $y \in \Delta^{\mathcal{I}}$*: If* $(x, y) \in R^{\mathcal{I}}$, $y \in C^{\mathcal{I}}\}$

An interpretation $\mathcal{I}$ *models* a GCI $C \sqsubseteq D$, for short $\mathcal{I} \models C \sqsubseteq D$, iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. An interpretation $\mathcal{I}$ *models* an ABox axiom $C(a)$, for short $\mathcal{I} \models C(a)$, iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and it models an ABox axiom of the form $R(a, b)$ iff $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$. An interpretation is a *model of an ontology* $(\mathcal{T}, \mathcal{A})$ iff it models all axioms appearing in $\mathcal{T} \cup \mathcal{A}$.

We now discuss embeddings of cones in $\mathbb{R}^n$ before turning our attention to $\mathcal{ALC}$.

3 Cone Embedding

Our approach is based on a geometric interpretation function $\mathcal{I}$ that represents concepts and relations as *convex cones* in $\mathbb{R}^n$. Cones satisfy the property that if x, y are inside a cone, then $\lambda x + \mu x$, $\lambda, \mu \geq 0$ is also inside that cone. We make use out of this property to construct the reification. We adopt the approach of [13] in using *polarity* $(\cdot)^{\circ}$ derived from the scalar product in $\mathbb{R}^n$ to construct a negation of concepts. The polar of a cone C, written C°, is defined as the set $\{x \in \mathbb{R}^n | \forall y \in C. x^T \cdot y \leq 0\}$, i.e., the set of all vectors being rotated at least 90 degrees away from any element of C. For all points neither belonging to a cone C nor to its polar C°, no statement about membership to concepts C, $\neg C$ can be made – the model is thus capable of representing uncertainty and thus is able to cope with the open world assumption.

Definition 2. *A* convex cone *is a set $C \subseteq \mathbb{R}^n$ with the property $\forall x, y \in C. \forall \lambda, \mu \in \mathbb{R}. (\lambda \geq 0 \land \mu \geq 0) \rightarrow \lambda x + \mu y \in C$. For readability, we refer to convex cones simply as cones. We define H_m as the m-dimensional hyperoctant cone $H_m \subset \mathbb{R}^m$ generated by m vectors $\{(1\,0\,\cdots 0)^T, (0\,1\,0\,\cdots 0)^T, (0\,0\,1\,0\,\cdots 0)^T, \ldots, (0 \cdots 0\,1)^T\}$.*

Closed convex cones are closed under set intersection, so $\cap$ is a meet operator $\wedge$ wrt. $\leq$ but not closed under set union. Instead they have to be closed up by the conic hull operator. The *conic hull* of a set b, for short $ch(b)$, is the smallest convex cone containing b. So, we can define the join operation $\vee$ by $a \vee b = ch(a \cup b)$. Considering $\mathbb{R}^n$ as the largest lattice element 1 and the empty set as the smallest lattice element 0 makes the resulting structure a bounded lattice.

The polarity operator for closed convex cones fulfills properties of orthocomplement. Hence the set of all closed convex cones (over $\mathbb{R}^n$) forms an ortholattice. As de Morgan's laws hold in any ortholattice, one gets in particular the following characterization of the conic hull: $ch(a \cup b) = (a^{\circ} \cap b^{\circ})^{\circ}$. We denote the set of all closed convex cones in $\mathbb{R}^n$ by $\mathcal{C}_n$. Then the following fact holds: For any $n \geq 1$, $\mathcal{C}_n$ is an ortholattice.

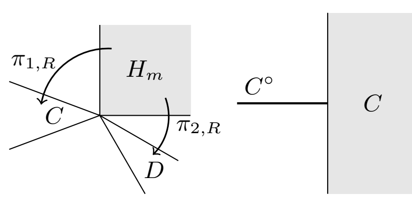

Fig. 1: Left: Reification of relation R is based on linear functions $\pi_{1,R}$, $\pi_{2,R}$ that project relation concept H_m to its arguments. Right: Illustration for reification requiring H_m with $m > n$.

We now define reification as illustrated in Figure 1 left to relate two concepts C, D. Relations are represented like concepts, i.e., by convex sets of specific geometric shape, and projections π_1, π_2 are introduced that link the embedded relations with the corresponding concepts. The main advantage of this over previous attempts is that the use of projections allows non-functional and partial relations to be represented. Approaches representing relations by geometric transformations in the embedding space such as TransE [3] are attractive as they also do not require further dimensions to be introduced, yet at the severe cost of being only able to represent functional relations, i.e., any x is always related to exactly one y. Several work aimed to remedy this severe limitation, but no general cure is possible when relying on a single geometric transformation function.

Definition 3. *We say a* reification *of a binary relation R between two cones C, D is given by a hyperoctant H_m and two linear functions $\pi_{1,R}$ and $\pi_{2,R}$ given as matrices $M \in \mathbb{R}^{n\times m}$.*

Let us now complement the geometric model. A cone interpretation for $\mathcal{ALC}$ maps symbols and formulae to cones in $\mathbb{R}^n$. The definition is as usual for interpretations, except that we exclude the origin $\vec{0}$ from the domain. This creates a convenient way for projections in the reification of relations to map subspaces to a well-defined 'nirvana' $\vec{0}$ whenever a mapping to the empty set is required. For example, the formula $\forall R.\top \sqsubseteq \bot$ saying that nothing is reachable by role R can elegantly be represented by setting $\pi_{2,R}$ to $\mathbf{0} \in \mathbb{R}^{n\times n}$, the projection to $\vec{0}$.

Definition 4. *A* cone interpretation *for a given $\mathcal{ALC}$ vocabulary $\mathcal{V} = N_c \cup N_C \cup N_R$ of constants, concept and role symbols is a structure $(\Delta, \cdot^{\mathcal{I}})$ where $\Delta = \mathbb{R}^n \setminus \{\vec{0}\}$ for some $n \in \mathbb{N}$ and $\cdot^{\mathcal{I}}$ is the* denotation function *defined for all $b \in N_c, A \in N_C, R \in N_R$ and concepts C, D over $\mathcal{V}$ such that the following conditions are fulfilled:*

$b^{\mathcal{I}} \in \Delta$,	$(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$,
$A^{\mathcal{I}} \in \mathcal{C}_n$,	$(\neg C)^{\mathcal{I}} = C^{\circ}$,
$R^{\mathcal{I}} \subseteq \Delta \times \Delta$,	$(C \sqcup D)^{\mathcal{I}} = (\neg(\neg C \sqcap D))^{\mathcal{I}}$,
$\top^{\mathcal{I}} = \Delta$,	$(\exists R.C)^{\mathcal{I}} = \pi_{1,R}\left(\pi_{2,R}^{-1}(C^{\mathcal{I}}) \cap H_m\right)$
$\bot^{\mathcal{I}} = \emptyset$,	*where H_m, $\pi_{1,R}$, $\pi_{2,R}$ are a reification of relation R and $\pi_{1,R}\left(\pi_{2,R}^{-1}(C^{\mathcal{I}}) \cap H_m\right)$ is a convex cone.*
	$(\forall R.C)^{\mathcal{I}} = (\neg\exists R.\neg C)^{\mathcal{I}}$

The notion of a cone interpretation being a model *(of an abox, tbox, ontology) is defined in the same way as for classical interpretations according to Def. 1.*

We now show that arbitrary $\mathcal{ALC}$ knowledge bases $\mathrm{KB} = (A, T)$ consisting out of abox A and tbox T are representable by a cone interpretation. First, we define how relations on the abox level are modeled.

Definition 5. *Given an $\mathcal{ALC}$ vocabulary with concept symbols $\mathcal{C}$, constant symbols $\mathcal{A}$, and role symbols $\mathcal{R}$, and $\mathcal{ALC}$ knowledge base* KB*, we say that the roles in* KB *are representable if there is a geometric interpretation $(\Delta, \mathcal{I})$ that is a model of* KB *and* $\mathrm{KB} \models R(a, b)$ *if and only if $b \in \pi_{1,R}(\pi_{2,R}^{-1}(a) \cap H_m)$ for some hyperoctant H_m.*

Proposition 1. *A cone interpretation of concepts maps all concept descriptions to closed convex cones.*

Proof. This is clear for atomic concepts, intersection, and for the polar operator. Disjunction is defined by de Morgan via intersection and polarity. But this is the conic hull, hence a mapping to a closed convex cone. Linear mappings also preserve cones, as they distribute over arbitrary linear combinations (not only those with positive scalars). For the existential, being a convex cone is enforced directly by the definition.

Note that enforcing closed convex cones for the embedding of existentials is not a strong constraint. Taking the null vector into account one can show that the inverse preserves cones: Let X be a cone and M be a linear mapping as used in reification. Let $v \in$

$M^{-1}[X]$, then $M(v) = w \in X$. Then for $\lambda > 0$ due to linearity $M(\lambda v) = \lambda M(v) = \lambda w$ and $\lambda w \in X$ due to the fact that X is a cone. Let $v, v' \in M^{-1}[X]$, then $M(v) = w \in X, M(v') = w' \in X$ (for some w, w'). Now $M(v + v') = M(v) + M(v') = w + w' \in X$, so $v + v' \in M^{-1}[X]$.

Reification employs matrix multiplications like several previous approaches, but it employs an 'in-between stop' at H_m which is the central trick to represent 1-to-k relations by making π_1 a k-to-1 mapping. Let us consider a simple example shown in Figure 1 to see that a stop H_m is necessary and that even $m > n$ may be necessary.

Example 1. We consider the cone C generated by vectors $\{(0\,1)^T, (0\,-1)^T, (1\,0)^T\}$ in $\mathbb{R}^2$ shown in Figure 1 right. Its negation given by the polarity operator C° is the cone generated by $\{(-1\,0)^T\}$. Now consider background knowledge $\exists R.C = \top$ saying that any entity is reachable from C by means of relation R. $\top$ is interpreted as $\mathbb{R}^2 \setminus \{\vec{0}\}$ and it requires four independent rays $\lambda_i c_i$, $\lambda_i > 0$, $c_i \in C$ to span $\mathbb{R}^2$, more than offered by C. It requires at least H_4 to achieve the desired mapping:

$$\pi_{1,R} := \begin{pmatrix} 0 & 0 & 1 & -1 \\ 1 & -1 & 0 & 0 \end{pmatrix}, \pi_{2,R} := \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix}$$

For any point $\vec{x} = (x_1 \cdots x_4)^T$ in H_4 we have $\pi_2(\vec{x}) = x_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and since $x_1, \ldots, x_4 \geq 0$ we have $\pi_2(\vec{x}) \in C$ and $\pi_1(\vec{x})$ covers $\mathbb{R}^2$ for $\vec{x} \in H_4$. In general, H_m with $m > n$ is required when concept C is a sub-space of lower dimensionality than the concept it is related to.

Let us discuss a more involved example showcasing the ability to model complex relationships.

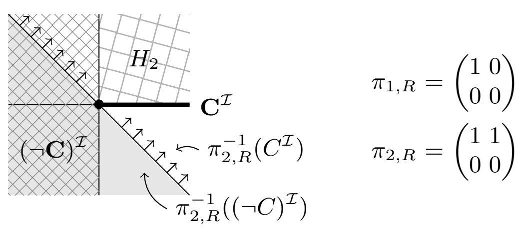

Fig. 2: Example for the construction of a geometric model of a tbox consisting of $\exists R.C = C$ and $\exists R.\neg C = \bot$

Example 2. Consider concept $C^\mathcal{I}$ represented as the positive x-axis, the complement $(\neg C)^\mathcal{I}$ is the negative halfspace in $\mathbb{R}^2$ shown in Fig 2. This model fulfills the two tbox axioms $\exists R.C = C$ and $\exists R.\neg C = \bot$. The first axiom is fulfilled, as $\pi_{2,R}^{-1}(C^\mathcal{I})$ is the region marked with arrows in the figure. The intersection with the region of possible relational facts H_2 results in H_2, the upper right quadrant. This is mapped to $C^\mathcal{I}$ by $\pi_{1,R}$. As $\pi_{2,R}^{-1}(\neg C)^\mathcal{I}$ does not intersect with H_2, $\exists R.\neg C = \bot$ is valid.

Therefore, partiality is obviously given. To show non-functionality, the instances need to be considered. Assume $c^{\mathcal{I}} = (\lambda\, 0)^T$ for an arbitrary $\lambda > 0$. $\pi_{1,R}^{-1}(c^{\mathcal{I}}) = \{(\lambda\, \mu)^T \mid \mu \geq 0\}$ and $\pi_{2,R}\pi_{1,R}^{-1}(c^{\mathcal{I}}) = \{(\lambda + \mu\, 0)^T \mid \mu \geq 0\}$. Thus, a $c^{\mathcal{I}} = (1\, 0)^T$ has a relation to all $b^{\mathcal{I}} = (\gamma\, 0)^T$ for $\gamma \geq 1$.

A property of our approach, besides its non-distributivity of $\cap$ over $\cup$, is its non-distributivity of $\exists$ over $\vee$, meaning $(\exists R.(C \sqcup D))^{\mathcal{I}} \neq (\exists R.C^{\mathcal{I}} \sqcup \exists R.D)^{\mathcal{I}}$. Despite this is not classical, e.g. different from $\mathcal{ALC}$-semantics, it may be quite useful in modeling: Assume a binary relation E is introduced to model whether a person is examined to have a specific disease. Thus, by asserting $E(person, disease)$ medical knowledge about a person at a specific point in time is reported. Now, it might be the case that an examination is not exact and thus results in the knowledge that the person could have disease A or disease B. However, assuming $\exists E.(A \sqcup B) = \exists E.A \sqcup \exists E.B$ would result in the conclusion that at this stage of examination it is already known which exact disease the person has. However, this exact specification was presumed not to be possible, and therefore, the instance representing the person should be placed in the embedding of $\exists E.(A \sqcup B)$ but neither in the embedding of $\exists E.A$ nor in the embedding of $\exists E.B$ because for both of them there is no justification in the examination. Thus, a lack of distributivity can be helpful to bridge gaps in semantics.

4 Distributive Embedding

The approach described so far leads to a possibility of expressing relational knowledge in general orthologics, which may be relevant for some applications as we have argued above. However, many knowledge bases consider stronger orthologics, i.e., expect distributivity to hold, which include all classical logics.

One prominent example is $\mathcal{ALC}$ with classical semantics. Here one requires the ortholattice also to be a Boolean algebra. In fact, classical $\mathcal{ALC}$-tboxes are characterized according to [15] by the fact that the existential is a strong operator, i.e., the existential quantifier with classical $\mathcal{ALC}$ semantics satisfies the following two properties: $(\exists R.\bot)^{\mathcal{I}} = (\bot)^{\mathcal{I}}$ and $(\exists R.(C \sqcup D))^{\mathcal{I}} = (\exists R.C \sqcup \exists R.D)^{\mathcal{I}}$.

Therefore, to adapt our approach to $\mathcal{ALC}$, distributivity of $\sqcap$ over $\sqcup$ and distributivity of $\exists$ over $\sqcup$ must be achieved. The first property can be met by restricting cones to so-called axis-aligned cones (al-cones) as introduced in [13] since geometric models based on al-cones are distributive. Al-cones have a finite basis and their generating vectors only consists out of components $+1$, -1, and 0. The second property can be met by restricting the modeling of relations in form of the role distributivity property.

Definition 6. *Role distributivity property RDP: if $x \in (C \sqcup D)^{\mathcal{I}}$ and $\pi_{2,R}^{-1}(x) \cap H_m \neq \bot^{\mathcal{I}}$, then it exists $x_c \in (C)^{\mathcal{I}}$ and $x_d \in (D)^{\mathcal{I}}$ with $x = x_c + x_d$ and $\pi_{2,R}^{-1}(x_c) \subseteq H_m$ and $\pi_{2,R}^{-1}(x_d) \subseteq H_m$.*

Each two concepts C and D must fulfill the RDP to regain distributivity of the $\exists$-operator.

Proposition 2. *$\exists R.(C \sqcup D) = \exists R.C \sqcup \exists R.D$ is valid if RDP is fulfilled.*

Proof. $\exists R.C \sqcup \exists R.D \sqsubseteq \exists R.(C \sqcup D)$ holds in any case. Therefore, it is sufficient to show $\exists R.(C \sqcup D) \sqsubseteq \exists R.C \sqcup \exists R.D$. Therefore, for all $y \in (\exists R.(C \sqcup D))^{\mathcal{I}}$ it needs to follow that $y \in (\exists R.C \sqcup \exists R.D)^{\mathcal{I}}$. Let $y \notin (\exists R.C)^{\mathcal{I}}$, $y \notin (\exists R.D)^{\mathcal{I}}$, as trivial in the other cases. Therefore, $y = \pi_{1,R}(\pi_{2,R}^{-1}(x_c + x_d) \cap H_m)$ for a $x_c \in C$ and a $x_d \in D$. With linearity of $\pi_{2,R}$ it follows that $y = \pi_{1,R}((\pi_{2,R}^{-1}(x_c) + \pi_{2,R}^{-1}(x_d)) \cap H_m)$ and $y \in \exists R.C \sqcup \exists R.D$ means $y = \pi_{1,R}((\pi_{2,R}^{-1} x_c \cap H_m) + (\pi_{2,R}^{-1} x_d \cap H_m))$. With RDP follows equality.

Having this property, it is possible to show that all $\mathcal{ALC}$ knowledge bases are representable by a geometric interpretation based on al-cones.

Proposition 3. *All $\mathcal{ALC}$ knowledge bases are representable by a geometric interpretation.*

Proof. We show that an al-cone interpretation of a $\mathcal{ALC}$ knowledge base without roles, i.e., only considering the Boolean part, can be extended to a model for roles as well.

Since $\mathcal{ALC}$ features the finite model property we may assume that the geometric model is finite and represents all facts following from a given knowledge base KB. Hence assume all concepts to be represented by cones and all constants by vectors in $\mathbb{R}^n$. We write $a^{\mathcal{I}_B}$ to refer to the vector obtained for constant a in the Boolean embedding and we write $a^{\mathcal{I}}$ for its embedding we seek to construct.

We iteratively construct a geometric model from a Boolean geometric model based on al-cones and a corresponding $\mathcal{ALC}$ model by processing role after role in a two-step process. Initially, we initialize $\cdot^{\mathcal{I}}$ by setting $\cdot^{\mathcal{I}}$ to the Boolean-only model $\cdot^{\mathcal{I}_B}$. In the first step, we consider a role R with $|\{(a,b)|\mathrm{KB} \models R(a,b)\}| = m$ and assume $R = \{(a_1,b_1),\ldots,(a_m,b_m)\}$ in the finite model. We extend the dimensions of our model from n to $n(m+1)$ by cloning the components of all vectors $\vec{x}$ that generate some concept. Let 0_k denote k consecutive zero components in a vector, then we can describe the modification of the embedding $c^{\mathcal{I}}$ for any constant c as follows:

$$\phi(c) = \sum_{i=1,\ldots,m, c=a_i \vee c=b_i} (0_{n \cdot i}\, (c^{\mathcal{I}})^T\, 0_{m-i})^T \tag{1}$$

$$c^{\mathcal{I}} \leftarrow \begin{cases} \phi(c) & \phi(c) \neq 0_{n(m+1)} \\ ((c^{\mathcal{I}})^T\, 0_{n \cdot m})^T & \text{otherwise} \end{cases} \tag{2}$$

Note that $\phi(c) \neq 0_{n(m+1)}$ occurs exactly if there is at least one a_i or b_i with $c = a_i$ or $c = b_i$. Doing so, we separate all entities in $\mathrm{dom}(R) \cup \mathrm{Img}(R)$ that occur in the model. In particular, we achieve that $\lambda a_i^{\mathcal{I}} \in \mathrm{dom}(R)$, $\lambda > 0$ if and only if $a_i^{\mathcal{I}} \in \mathrm{dom}(R)$ and likewise for $\mathrm{Img}(R)$. We repeat the process for all roles.

In the second step, we need to construct the reification of any role R which can be done as follows. Assume again $R = \{(a_1,b_1),\ldots,(a_p,b_p)\}$ and then define a reification based on hyperoctant H_p embedded in the model using projections

$$\pi_1 = \left[a_1^{\mathcal{I}} \cdots a_p^{\mathcal{I}}\right], \pi_2 = \left[b_1^{\mathcal{I}} \cdots b_p^{\mathcal{I}}\right],$$

where $[\cdots]$ represents a matrix composed out of column vectors. By construction of $a_i^{\mathcal{I}}$, $b_i^{\mathcal{I}}$, we have $c_{R(a,b)} = (0_{i-1}\, 1\, 0_{n-1})^T \in H_l$ which corresponds to $R(a_i, b_i)$ since

$\pi_1(c_{R(a,b)}) = a_i^{\mathcal{I}}$ and $\pi_2(c_{R(a,b)}) = b_i^{\mathcal{I}}$. It thus follows $\pi_1(H_p) \supseteq \mathrm{dom}(R)$ and $\pi_2(H_p) \supseteq \mathrm{Img}(R)$, respectively. Also by construction, for any $c \notin \mathrm{dom}(R)$ we have $c^{\mathcal{I}} \notin \pi_1(H_p)$ since $a_i^{\mathcal{I}}$ and $c^{\mathcal{I}}$ reside in mutually exclusive sub-spaces according to (2).

We note that this proof, albeit constructive, is of theoretical nature since it exploits a large amount of dimensions for H_p to ease the construction.

It is not only possible to represent each $\mathcal{ALC}$ knowledge base with such a geometric interpretation, it is also possible to interpret each geometric model based on the axis-aligned cones introduced in the above proposition as an $\mathcal{ALC}$-ontology.

Proposition 4. *A geometric interpretation based on al-cones fulfilling RDP, where $\pi_{1,R}(\pi_{2,R}(C) \cap H_m)$ maps to an al-cone for each half-axis C, represents an $\mathcal{ALC}$ knowledge base.*

Proof. A geometric interpretation without considering roles is shown in [13]. Therefore, it is sufficient to show that the relational part also fulfills the restrictions of $\mathcal{ALC}$. $\exists R.\bot = \bot$ is fulfilled by construction. The distributivity of $\exists$ over $\vee$ is ensured by RDP, as shown in Proposition 2.

As each half-axis is mapped to an al-cone, because of linearity of π, each concept is mapped to a union of al-cones, which is still an al-cone. Also because of linearity, it is ensured that the properties needed for roles, e.g. $\exists R.C \sqsubseteq \exists R.\top$ are fulfilled. The negation of $\exists R.C$ is given by polarity (as it would not be a geometric model otherwise). Therefore, the resulting geometric model represents a $\mathcal{ALC}$ knowledge base.

5 Related work

The approach presented in this paper is a contribution to recent efforts on combining knowledge representation (KR) and machine learning (ML). Roughly, those approaches use ML algorithms to learn an ontology or to exploit the ontologies as constraint specifications in order to get more accurate models or in order to optimize statistical models. Our work and many of the recent KGE approaches (see below) tackle the problem of building accurate models in the sense that these are compatible with the background knowledge expressed in an ontology. But there is also relevant work outside of the KGE community which incorporates ontologies into standard statistical models. An example is the approach of Deng and colleagues [5] in which pairwise conditional random fields are optimized by incorparting knwoledge of the background as additional factors.

Earlier approaches to knowledge graph embedding—including TransE [3]—were motivated by efficient learning algorithms, hence resolving the expressivity vs. feasibility dilemma strictly in favor of feasibility. For example, consider the notion of "full expressivity" in [11] which only states that an approach is able to differentiate between all class members and non-members of a concept. In those approaches—including the well-known TransE [3]—heads and tails of KGE triples are represented as real-valued vectors and relations are represented as vectors, matrices or tensors, i.e., simple geometric operations. In many occasions, the geometric operations lead to relations that are functional, total or are constrained by other means. But the resulting simple mathematical operations (for representing relations) provide not much expressivity from a KR

Geometrical Structure Concepts	Relations	Logic	Concept lattice	Negation	Approach/ Reference
Convex sets	pairs	Quasi-chained Datalog$^{\pm}$	distr	atomic	[7]
Hyperspheres	translation	$\mathcal{EL}^{++}$	distr	atomic	[9]
Axis-aligned Cones	pairs	rank-restricted $\mathcal{ALC}$	distr	full Boolean	[13]
Axis-aligned Cones	**cones**	$\boldsymbol{\mathcal{ALC}}$	**distr**	**full Boolean**	**this paper**
Closed Subspaces in Hilbert Space	pairs	Minimal Quantum Logic	orthomodular	orthonegation	[6]
Hyperbolic cones	rotation	logics for taxonomies	distrib	atomic (?)	[2]
Cartesian Products of 2D-cones	rotation + volume change	FOL queries (?) (without $\forall$)	distrib	negation as failure	[18]

Table 1: Comparison of approaches for embedding with the approach of this paper in bold font

point of view. Even in later approaches, e.g. [16], functionality is not dealt logically but rather by relying on thresholds. In order to illustrate our point, consider an object represented by a vector x. A relation R is represented in [16] by a rotation M_R. The vector $y := M_R x$ gives only some "prototypical" object to which x stands in R-relation. Other objects y' to which x might stand in R-relation are given by $\|M_R x - y'\| \leq \lambda$ for some threshold λ. In particular, this means that all objects to which x is R-related are close to $M_R x$. In consequence, x can not be related to some objects y' and y'' that are quite different in that they belong to complementary concepts, $y' \in C^{\mathcal{I}}$ and a $y'' \in (\neg C)^{\mathcal{I}}$.

In the following we discuss only those KGE approaches that explicitly mention the kind of geometries used for embedding and the logic that characterizes them (see Tab. 1). Table 1 considers in particular the question how concepts and roles are embedded, whether distributivity of $\sqcap$ over $\sqcup$ is fulfilled, and what kind of negation is expressed. We note that there are good reasons for considering non-distributive logics for the investigation of concept hierarchies as discussed in [4]. Non-distributive logics are investigated thoroughly by Hartonas in [8].

[7] identify a fragment of existential Datalog (fulfilling the quasi-chainedness property) as an appropriate logic for arbitrary convex regions in euclidean spaces. [9] finds a correspondence for hyperspheres and the lightweight description logic $\mathcal{EL}$. [13] identifies axis-aligned cones as an appropriate geometrical class for embedding concepts of the semi-descriptive logic $\mathcal{ALC}$. While [7,9] do not allow for full negation of concepts to be represented, [13] define negation for the model of axis-aligned cones that uses polarity, which possibly gives rise to some interesting logic structure. On the other hand, in [13] binary relations are allowed to be arbitrary pairs of vectors, whereas [7] models also relations (of any arity) by convex regions. The approach of this paper shares the property with [13] of providing full (Boolean) negation. But our approach deviates from [13] in the interpretation of roles—with consequences at three columns of the table: Our approach does not consider arbitrary set of pairs as possible embeddings of roles. In [13], this generality is possible by restricting the quantifier rank of

the concepts in the ontology. In contrast, our approach interprets roles by reifying concepts, that are allowed to be (arbitrary) cones. This also allows handling arbitrary (non quantifier-restricted) $\mathcal{ALC}$ ontologies as background knowledge.

In all three approaches the expressible concept hierarchy fulfills distributivity of conjunction over disjunction. The approach of [6] considers minimal quantum logic which does not fulfill distributivity but (only) a weakening: orthomodularity. Relations are handled in [6] by doubling the dimension of space where the concepts are embedded and by treating $R(a, b)$ as a vector $[a\ b]^T$ in this higher-dimensional space.

The approach of [2] uses hyperbolic cones for modelling relation hierarchy graphs and to grasp properties that follow by traversing the edges. The exact logic captured by this approach is not clear (to us), as the authors allow next to subclass relations also part-of relations. Quantifiers, negation (and other Boolean operators) are not handled explicitly in this approach. One can think of antinomies being used in the hierarchy—but this rather correspond to atomic negation. Hence we describe the logic as taxonomic.

The approach of [18] also uses the idea of [13] to handle negation of concepts by using cones. But they do not consider negation as polarity, but negation as set-complement in 2D and cartesian products to embed concepts. Relations in [18] are handled by rotating the support point and changing the volume of a cone. The authors claim to embed FOL queries. Interestingly, they exclude the universal quantifier $\forall$ form their considerations. Given the fact that $\forall$ is dual to $\exists$ via negation we consider this as a sign that negation is not treated in its full expressivity. In particular, they cannot fully account for de Morgan rules since negation as used there is a form of negation as failure.

6 Conclusions and Outlook

Algorithms involving computations over some declarative specification of the world have to trade-off between expressivity and feasibility. Feasibility of embeddings has been traditionally favored over expressivity, because many works are governed by practical implementations. Current investigations now try to push the expressivity envelope and to strive for a better alignment between expressivity provided by an embedding and the expressivity required for sound representation of some domain knowledge. Achieving a true alignment of the geometric structures determined in learning methods with logical models is necessary to exploit embeddings in hybrid AI approaches, in particular with reasoning beyond link prediction. This paper shows how the idea of reification can be applied to knowledge graph embeddings and presented the first geometric model of full $\mathcal{ALC}$ which is based on feasible structures previously employed in knowledge graph embedding, namely convex sets (cones) and linear functions (matrix multiplication). Our approach is not tailored to $\mathcal{ALC}$ but may be useful to a much larger family of orthologics. As this paper has been taken the second roadway of pushing forward expressivity in geometric models, future work will aim to complement these fundamental findings with a learning method to acquire an embedding with reification of roles.

Acknowledgements The research of Mena Leemhuis and Özgür L. Özçep is funded by the BMBF- funded project SmaDi. Diedrich Wolter acknowledges support by Technologieallianz Oberfranken and BMBF (Dependable Intelligent Software Lab).

References

1. Baader, F.: Description logic terminology. In: Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.) The Description Logic Handbook, pp. 485–495. Cambridge University Press (2003)
2. Bai, Y., Ying, R., Ren, H., Leskovec, J.: Modeling Heterogeneous Hierarchies with Relation-specific Hyperbolic Cones. In: Proc. 35th Annual Conference on Neural Information Processing Systems (NeurIPS 2021). p. arXiv:2110.14923 (2021)
3. Bordes, A., Usunier, N., García-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Burges, C.J.C., Bottou, L., Ghahramani, Z., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States. pp. 2787–2795 (2013), http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data
4. Conradie, W., Palmigiano, A., Robinson, C., Wijnberg, N.: Non-distributive logics: from semantics to meaning. arXiv e-prints arXiv:2002.04257 (Feb 2020)
5. Deng, J., Ding, N., Jia, Y., Frome, A., Murphy, K., Bengio, S., Li, Y., Neven, H., Adam, H.: Large-scale object classification using label relation graphs. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision — ECCV 2014, Lecture Notes in Computer Science, vol. 8689, pp. 48–64. Springer International Publishing (2014), http://dx.doi.org/10.1007/978-3-319-10590-1_4
6. Garg, D., Ikbal, S., Srivastava, S.K., Vishwakarma, H., Karanam, H., Subramaniam, L.V.: Quantum embedding of knowledge for reasoning. In: Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 32. Curran Associates, Inc. (2019)
7. Gutiérrez-Basulto, V., Schockaert, S.: From knowledge graph embedding to ontology embedding? An analysis of the compatibility between vector space representations and rules. In: Thielscher, M., Toni, F., Wolter, F. (eds.) Proc. of KR 2018. pp. 379–388. AAAI Press (2018)
8. Hartonas, C.: Reasoning with Incomplete Information in Generalized Galois Logics Without Distribution: The Case of Negation and Modal Operators, pp. 279–312. Springer International Publishing, Cham (2016)
9. Kulmanov, M., Liu-Wei, W., Yan, Y., Hoehndorf, R.: El embeddings: Geometric construction of models for the description logic EL++. In: Proceedings of IJCAI-19 (2019)
10. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. pp. 2181–2187. AAAI'15, AAAI Press (2015)
11. Mehran Kazemi, S., Poole, D.: SimplE Embedding for Link Prediction in Knowledge Graphs. arXiv e-prints arXiv:1802.04868 (Feb 2018)
12. Nickel, M., Tresp, V., Kriegel, H.P.: A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th International Conference on International Conference on Machine Learning. pp. 809–816. ICML'11, Omnipress, USA (2011), http://dl.acm.org/citation.cfm?id=3104482.3104584
13. Özçep, Ö.L., Leemhuis, M., Wolter, D.: Cone semantics for logics with negation. In: Bessiere, C. (ed.) Proc. of IJCAI 2020. pp. 1820–1826. ijcai.org (2020)
14. Ren, H., Hu, W., Leskovec, J.: Query2box: Reasoning over knowledge graphs in vector space using box embeddings. In: International Conference on Learning Representations (2020)
15. Schild, K.: A correspondence theory for terminological logics: Preliminary report. In: Proceedings of IJCAI-91. pp. 466–471 (1991)

16. Sun, Z., Deng, Z.H., Nie, J.Y., Tang, J.: RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. arXiv e-prints arXiv:1902.10197 (Feb 2019)
17. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. Proceedings of the AAAI Conference on Artificial Intelligence **28**(1) (Jun 2014), https://ojs.aaai.org/index.php/AAAI/article/view/8870
18. Zhang, Z., Wang, J., Jiajun, C., Shuiwang, J., Feng, W.: ConE: Cone embeddings for multihop reasoning over knowledge graphs. In: Proc. 35th Annual Conference on Neural Information Processing Systems (NeurIPS 2021) (2021)