Embedding Tarskian Semantics in Vector Spaces

Taisuke Sato
AI research center AIST / National Institute of Informatics, Japan
email:satou.taisuke@aist.go.jp

Abstract

We propose a new linear algebraic approach to the computation of Tarskian semantics in logic. We embed a finite model \( M \) in first-order logic with \( N \) entities in \( N \)-dimensional Euclidean space \( \mathbb{R}^N \) by mapping entities of \( M \) to \( N \) dimensional one-hot vectors and \( k \)-ary relations to order-\( k \) adjacency tensors (multi-way arrays). Second given a logical formula \( F \) in prenex normal form, we compile \( F \) into a set \( \Sigma_F \) of algebraic formulas in multi-linear algebra with a nonlinear operation. In this compilation, existential quantifiers are compiled into a specific type of tensors, e.g., identity matrices in the case of quantifying two occurrences of a variable. It is shown that a systematic evaluation of \( \Sigma_F \) in \( \mathbb{R}^N \) gives the truth value, \[ \langle F \rangle = 1 \text{ if } M \models F \text{ else } \langle F \rangle = 0. \]

Our proposal is motivated by recent work on logical inference concerning knowledge graphs(KGs). KGs are graphs encoding RDF triples of the form \((\text{subject} : s, \text{predicate} : p, \text{object} : o)\) and can be considered as a set of ground atoms of the form \( p(s, o) \). There are huge KGs available such as Freebase(Bollacker et al. 2008). The problem is that although they are good resources of the real world information and logically simple, they are huge, containing tens of millions of nodes and furthermore incomplete; there are lots of inconsistent data and also lots of missing data. To carry out various KG tasks such as computing the truth value (or more generally probability) of \( p(s, o) \) while coping with the sheer amount of data and incompleteness, three major approaches are developed (Nickel et al. 2015); one that is based on probabilistic models, one that uses explicit features sampled from the graph and one that learns latent feature vectors from the graph. The last approach, latent feature approach, compiles entities and predicates in the domain into vectors and tensors(Kolda and Bader 2009) respectively and apply various linear algebraic operations, with dimension reduction, to compute the probability of \( p(s, o) \).

In the development of these approaches, formulas beyond ground atoms are introduced and investigated such as existentially quantified conjunctions as queries and definite clauses as constraints on KGs (Grefenstette 2013; Rocktäschel, Singh, and Riedel 2015; Krompaß, Nickel, and Tresp 2014; Guu, Miller, and Liang 2015; Yang et al. 2015). However, from a logical point of view, their treatment was confined to propositional logic level and the evaluation of general first-order formulas is left untouched except for the work done by Grefenstette (Grefenstette 2013). Regrettably, while he succeeded in completely embedding the fragment of model theory, model theory of quantifier-free first-order logic, in tensor spaces, quantified formulas were excluded and had to be treated separately by another framework. Nested quantification was not allowed either. So how to evaluate arbitrarily quantified formulas in a vector space still remains open.

We solve this problem by introducing specific tensors for existential quantifiers together with a nonlinear operation. Our contribution is two-fold. First we introduce a single framework for the evaluation of quantified first-order for-
Formulas in a vector space, assuming the domain is finite, thus solving the remaining problem.

The second contribution is to present a concrete method, based on our framework, to compute the least model of Dialog programs in a vector space, which opens up a completely new way of evaluating recursive programs, though we have to skip details due to page limitations and only sketch experimental result.

At this point it would be beneficial to ask why evaluating logical formulas in a vector space is an interesting idea. First, there are a rich family of algebraic operations available in a vector space such as inner product, outer product, projection, PCA, SVD and so on that helps analyzing and manipulating vector data. Second, basically they are of polynomial time complexities, so we can expect efficient computation. Last but not least, approximation through various matrix and tensor decomposition potentially leads to logical inference for Web scale symbolic data.

We assume the reader is familiar with basics of logic and linear algebra including matrices and tensors(Kolda and Bader 2009; Cichocki et al. 2009).

## Preliminaries

We first review some terminology in logic. We assume our first order language $\mathcal{L} = (\mathcal{D}, \mathcal{F})$ contains $N$ constants \{e$_1$, ..., e$_N$\} and no function symbols.

A model $\mathcal{M} = (\mathcal{D}, I)$ is a pair of domain, a nonempty set $\mathcal{D}$ and an interpretation $I$ that maps constants e$_i$ to elements (entities, individuals) $I(e_i) \in \mathcal{D}$ and $k$-ary predicate symbols $k$ times $r$ to $k$-ary relations $I(r) \subseteq \mathcal{D} \times \cdots \times \mathcal{D}$. An assignment $a$ is a mapping from variables $x$ to an element $a(x) \in \mathcal{D}$. It provides a way of evaluating formulas containing free variables.

Syntactically terms mean variables and/or constants and atomic formulas or atoms $r(t_1, \ldots, t_k)$ are comprised of a $k$-ary predicate symbol $r$ and $k$ terms $t_1, \ldots, t_k$ some of which may be variables. Formulas $F$ in $\mathcal{L}$ are inductively constructed as usual from atoms using logical connectives (negation $\neg$, conjunction $\wedge$, disjunction $\vee$) and quantifiers ($\exists, \forall$).

Now we define free/bound occurrences of variables in $F$. When $F$ is a term, all variables in $F$ occur free in $F$. When $F$ is a negation $\neg F_1$, disjunction $F_1 \lor F_2$ or conjunction $F_1 \land F_2$, free variables in $F_1$ and those in $F_2$ both occur free in $F$ and vice versa. When $F$ is an existentially quantified formula $\exists x F_1$, free variables in $F_1$ except $x$ occur free in $F$ and vice versa. Variables in $F$ that do not occur free in $F$ are said to be bound. A formula is closed if it has no free variable whereas it is open if it has no quantification.

Given a model $\mathcal{M} = (\mathcal{D}, I)$ and an assignment $a$, the denotation $[F]_I^a$ in $\mathcal{M}$ of an expression $E$ is inductively defined for terms and formulas as follows. $[r]_I^a = I(r)$ if $r$ is a constant, else $a(r)$. When $F$ is a $k$-ary relation symbol, $[r(t_1, \ldots, t_k)]_I^a = 1$ (true) if $(I(t_1)_I^a, \ldots, I(t_k)_I^a) \in I(r)$, else 0 (false). Let $F_1, F_2$ be formulas. For a negation $\neg F_1$, we define $[\neg F_1]_I^a = 1 - [F_1]_I^a$. $[F_1]_I^a \lor [F_2]_I^a = 1$ if $[F_1]_I^a = 1$ or $[F_2]_I^a = 1$, else 0. $[\exists x F_1]_I^a = 1$ if there exists some $d \in \mathcal{D}$ such that $[F_1]_I^a[x \leftarrow d] = 1$, else 0. Here $a[x \leftarrow d]$ is a new assignment that is the same as $a$ except that it assigns $d$ to the variable $x$. Finally $[A \land B]_I^a = [\neg (\neg A \lor \neg B)]_I^a$ and $[\forall x F_1]_I^a = [\neg \exists x \neg F_1]_I^a$. For any formula $F$, $[F]_I^a \in \{1, 0\}$ and when $[F]_I^a = 1$, we write $\mathcal{M} \models a F$. However when $F$ is closed, since $[F]_I^a$ does not depend on the assignment $a$, we just write $[F]$ and $\mathcal{M} \models F$ if $F$ is true in $\mathcal{M}$. For formulas $F, G$, we say $F$ and $G$ are equivalent and write $F \equiv G$ if $[F]_I^a = [G]_I^a$ for any model $\mathcal{M}$, any interpretation $I$ and any assignment $a$. In what follows, $I, a$ are omitted when they are clear from the context.

Then recall that a literal is an atom (positive literal) or its negation (negative literal). Suppose $F$ is an open formula. $F$ has an equivalent formula in disjunctive normal form (DNF) $A_1 \lor \cdots \lor A_k$ such that each disjunct $A_i$ is a monomial, i.e., conjunction of literals. Dually $F$ has an equivalent formula in conjunctive normal form (CNF) $A_1 \land \cdots \land A_k$ such that each conjunct $A_i$ is a clause, i.e., disjunction of literals.

It is known that every formula has an equivalent formula in prenex normal form $Q_1 t_1 \cdots Q_m t_m G$ where $Q_1, \ldots, Q_m$ are quantifiers $\exists, \forall$ and $G$ is open. So to evaluate the truth value $[F]$ of a given $F$ in $\mathcal{M}$, since $G$ is equivalent to DNF or CNF, we have only to evaluate a prenex normal form $Q_1 t_1 \cdots Q_m t_m G$ where $G$ is an open DNF or CNF.

Note the subformula $Q_m t_m G$. Since DNF and CNF are convertible to each other, it is equivalent to $\exists$-DNF or $\forall$-CNF where $\exists$-DNF is a class of formulas $F$ such that $F$ is a disjunction of disjuncts $\exists x (L_1 \land \cdots \land L_M)$ comprised of literals $L_1, \ldots, L_M$, whereas $\forall$-CNF is a class of formulas $F$ such that $F$ is a conjunction of conjunctions $\forall x (L_1 \lor \cdots \lor L_M)$. Suppose $F = Q_1 t_1 \cdots Q_m t_m G$ is given. We may assume, without loss of generality, that if $Q_m = \exists$, $Q_m t_m G$ is in $\exists$-DNF. Otherwise $Q_m t_m G$ is in $\forall$-CNF.

Now we turn to vector spaces. We consider tensors as multi-linear maps as mathematical objects and multi-way arrays as data structure depending on the context(Kolda and Bader 2009; Cichocki et al. 2009). Although tensors are a generalization of vectors and matrices, we specifically say vectors and matrices when their shape needs to be distinguished.

In what follows, scalars are denoted by lower case letters like $a$. Vectors mean column vectors and we denote them by boldface lower case letters like $a$ and $a$’s components by $a_i$. $\mathcal{D} = \{e_1, \ldots, e_N\}$ stands for the standard basis of $N$-dimensional Euclidean space $\mathbb{R}^N$ where $e_i = (0, \ldots, 1, 0)^T$ is a vector that has one at the $i$-th position and zeros elsewhere. Such vectors are called one-hot vectors. $1$ is a vector of all ones. $(a \bullet b) = a^T b$ is the inner product of $a$ and $b$ whereas $a \circ b = ab^T$ is their outer product. Matrices are assumed to be square and written by boldface upper case letters like $A$. In particular $I$ is an identity matrix. $I = 1 \circ I$ is a matrix of all ones. $\text{tr}(A)$ stands for the trace of $A$. Order-$p$ tensors in $\mathbb{R}^{N \times \cdots \times N}$ are denoted by underlined matrices like $\underline{A}$ or $\{a_{i_1, \ldots, i_p}\}$ (1 $\leq i_1, \ldots, i_p \leq N$). $\underline{A}$’s component $a_{i_1, \ldots, i_p}$ is also written as $a_{i_1, \ldots, i_p}^A$. Let $\underline{A} = \{a_{i_1, \ldots, i_p}\}$ and $\underline{B} = \{b_{i_1, \ldots, i_p}\}$ be tensors. The mode-$(n,m)$ contracted product $\underline{A} \times_{n,m} \underline{B}$ of $\underline{A}$ and $\underline{B}$ is defined by $(\underline{A} \times_{n,m} \underline{B})_{k_1, \ldots, k_n, i_1, \ldots, i_p, k_{n+1}, \ldots, k_{n+m}}$. 


= \sum_i a_{i_1, \ldots, i_{n-1}, 1, \ldots, p} b_{i_1, \ldots, i_{n-1}, 1, \ldots, k} \quad \text{with the convention that the association is to the left, i.e., } A \times_{n,m} B \times_{p,q} C = (A \times_{n,m} B) \times_{p,q} C. \quad \text{So } A \bullet u, \text{ the contracted product of } A \text{ and vector } u, \text{ which is computed by } \langle A \bullet u \rangle_{i_1, \ldots, i_{n-1}, j_1, \ldots, j_{n-1}, i_p, \ldots, i_{n-1}} = \sum_i a_{i_1, \ldots, i_{n-1}, 1, \ldots, p} u_j \text{ is equal to } A \times_{n,m} u \text{ and the usual n-mode product } A \times_n U \text{ of } A \text{ and matrix } U \text{ is equal to } A \times_n U.}

Tensors can be constructed by outer products; \((a \circ b \circ c)_{ijk} = a_{i} b_{j} c_{k}\) is an order-3 tensor and \((A \circ B)_{i_1, \ldots, i_{n-1}, j_1, \ldots, j_{n-1}, k_1, \ldots, k_{n-1}} = a_{i_1, \ldots, i_{n-1}} b_{j_1, \ldots, j_{n-1}} k_{k_1, \ldots, k_{n-1}}\) is the outer product of \(A = \{a_{i_1, \ldots, i_{n-1}}\}\) and \(B = \{b_{j_1, \ldots, j_{n-1}}\}\).

**Embedding a model into a vector space**

Let \(\{e_1, \ldots, e_N\}\) be the set of constants in \(\mathcal{D}\) and \(M = (\mathcal{D}, I)\) a model where \(\mathcal{D} = \{e_1, \ldots, e_N\}\) (we here identify \(I(e_i)\) and \(I(r)\) in \(M\) with \(e_i\) and \(r\) respectively to avoid notational complications). We show how to replace the evaluation \([F]_I\) of a prenex formula \(F\) in \(M\) with the evaluation of \(\Sigma_F\), a set of tensors compiled from \(F\), in \(N\)-dimensional Euclidean space \(\mathbb{R}^N\). The compilation of \(F\) into \(\Sigma_F\) starts from literals then proceeds to compound formulas and quantifications.

**Entities, literals, logical connectives and existential quantifier**

First we isomorphically map \(M\) to a model \(M'\) in \(\mathbb{R}^N\). We map entities \(e_i\in\mathcal{D}\) to one-hot vectors \(e_i\). So \(\mathcal{D}\) is mapped to \(\mathcal{D}' = \{e_1, \ldots, e_N\}\), the basis of \(\mathbb{R}^N\). We next map a k-ary relation \(r\) in \(M\) to a k-ary relation \(r'\) over \(\mathcal{D}'\) which is computed by an order-k tensor \(R = \{r_{i_1, \ldots, i_k}\}\). \(R\) is designed to retain the truth value \([p(e_{i_1}, \ldots, e_{i_k})]_I\) in \(M\) and given by the equation

\[
[p(e_{i_1}, \ldots, e_{i_k})]_1 = R(e_{i_1}, \ldots, e_{i_k}) \quad \text{as multi-linear map}
\]

\[
R = \bigotimes_{i=1}^{k} e_{i} \quad \text{as \(k\)-ary relation}
\]

We identify \(r'\) with \(R\) for simplicity and say \(R\) encodes the \(M\)-relation \(r\). Let \(M'\) be a model \((\mathcal{D}', I')\) in \(\mathbb{R}^N\) such that \(I'\) interprets entities by \(I'(e_i) = e_i\) and \(I'\) the inner most quantified subformula \(p\) in \(M'\) as follows.

We say \(\neg R\) encodes an \(M\)-relation \(\neg r^1\). Negation other than negative literals and conjunction and disjunction are evaluated in \(M'\) as follows.

\[
\neg [F]_{I',a'1} = 1 - [F]_{I',a'1}
\]

\[
[F_1 \wedge \cdots \wedge F_n]_{I',a'1} = [F_1]_{I',a'1} \cdots [F_n]_{I',a'1}
\]

\[
[F_1 \vee \cdots \vee F_n]_{I',a'1} = \min ([F_1]_{I',a'1} \cdots [F_n]_{I',a'1})
\]

\[
[\exists y F]_{I',a'1} = \min (\sum_{i=1}^{m} [F_{I',a'1 \cdot i}]_{I',a'1})
\]

Here \(\min (x) = \min (x, 1) = x\) if \(x < 1\) else \(1\) and when applied to tensors, it means componentwise application. \(F_{y \cdot e_i}\) is a formula obtained from \(F\) by replacing every free occurrence of \(y\) in \(F\) with \(e_i\). Universal quantification is treated as \(\forall x F = \neg \exists x \neg F\).

It is straightforward to check that the evaluation \([F]_{I',a'1}\) of a formula \(F\) in \(M'\) by (2), (3), (4), (5), (6) and (7) coincides with \([F]_{I,a}\) in \(M\). However, although this evaluation is carried out in a vector space, i.e. \(\mathbb{R}^N\), it is based on the reduction of quantification to the ground level as (7) indicates and contains a lot of redundancy. We next show how to do the same thing without grounding quantifications.

\[\exists\text{-DNF}\] and \[\forall\text{-CNF}\] as tensors

Now we come to the crucial point of our proposal, evaluating quantified formulas without grounding. Consider a prenex formula \(F = Q_1 x_1 \cdots Q_m x_m G\) for the moment we assume the inner most quantified subformula \(Q_m x_m G\) is in \(\exists\text{-DNF}\).

Let \(\exists y (L_1 \wedge \cdots \wedge L_M)\) be an arbitrary disjunct of \(Q_m x_m G\) where \(L_1, \ldots, L_M\) are literals. We further assume the variable condition that \(y\) occurs once in each literal \(L_m = r^n_m (x_1^{m}, \ldots, x_{N_m}^{m})\) (\(1 \leq m \leq M\)). Here \(r^n_m = r_m\) if \(L_m\) is a positive literal else \(r^n_m = \neg r_m\) in \(M\) is called the \(M\)-relation contained in \(L_m\). Let \(R_m(1 \leq m \leq M)\) be a tensor encoding the \(M\)-relation \(r^n_m\) defined respectively by (1) or (3). So \(R_m(x_1^{m}, \ldots, x_N^{m}) = [L_m(x_1^{m}, \ldots, x_N^{m})]\) holds where \(x_1^{m}\) \((1 \leq i \leq N_m)\) range over the domain of constants \(\mathcal{D} = \{e_1, \ldots, e_N\}\) while \(x_N^{m}\) correspondingly range over the domain of the standard basis \(\mathcal{D}' = \{e_1, \ldots, e_N\}\). The notation \(L_m(x_1^{m}, \ldots, x_N^{m})\) emphasizes that \(x_1^{m}, \ldots, x_N^{m}\) occur in \(L_m\).

Suppose \(L_m = r^n_m (x_1^{m}, \ldots, x_{N_m}^{m})\) (\(1 \leq m \leq M\)) has \(y\) as the \(j_m\)-th argument. Remove \(y\), the \(j_m\)-th argument, from \((x_1^{m}, \ldots, x_{N_m}^{m})\). Write the remaining arguments (with order preserved) collectively as \(x_1^{m} \wedge \cdots \wedge x_{N_m-1}^{m}\) and consider \([L_m] = [L_m(x_1^{m} \wedge \cdots \wedge x_{N_m-1}^{m})]_{R_{N_m-1}}\) as a function of \(x_{N_m-1}^{m}\) or a function of the corresponding arguments \(x_1^{m} \wedge \cdots \wedge x_{N_m-1}^{m}\) over \(\mathbb{R}^{N_m-1}\) parameterized with \(y\). Then consider \([\exists y (L_1 \wedge \cdots \wedge L_M)]_{I,10}\) as a relation combined with arguments (free variables, possibly duplicate) \((x_1^{1}, \ldots, x_{N-1}^{1})\) over \(\mathbb{R}^{N-1}\), or equivalently, a function applied to \((x_1^{1}, \ldots, x_{N-1}^{1})\) over \(\mathbb{R}^{N-1}\). We seek a tensor \(R_{10}\) that encodes this function, i.e., \(R_{10}\) such

\[1 \circ \cdots \circ 1\] is an order-k tensor. \(1 \circ \cdots \circ 1(e_{i_1}, \ldots, e_{i_k}) = (1 \circ e_{i_1}) \cdots (1 \circ e_{i_k}) = 1 \circ e_{i_1}\) holds.
that \(\exists y(L_1 \land \cdots \land L_M)\) = \(R^{\text{new}}(x^{(1)}, \ldots, x^{(M)})\) holds. Look at
\[
\exists y(L_1 \land \cdots \land L_M) = \exists y(r_1(x^{(1)}) \land \cdots \land r_d(x^{(M)}))
\]
\[
= \min_j \left( \sum_{k=1}^M R_k(x^{(m)}) \right)
\]
\[
= \min_j \left( \sum_{k=1}^M \left( \left( R_k \circ_j e_k \right)(x^{(m)}) \right) \right)
\]
\[
= \min_j \left( \sum_{k=1}^M \left( (R_k \circ_j e_k)(x^{(1)}, \ldots, x^{(M)}) \right) \right)
\]
\[
= \min_j \left( \sum_{k=1}^M \left( (R_k \circ_j e_k)(x^{(1)}, \ldots, x^{(M)}) \right) \right)
\]

Here
\[
Q^{M}_{M} = \sum_{k=1}^N \sum_{1}^{M} e_k \circ \cdots \circ e_k
\]

is a tensor representing the existential quantifier \(\exists\).

Summing up, the \(M\)-relation extracted from \(\exists y(L_1 \land \cdots \land L_M)\), which solely depends on the free variables in it, is encoded by
\[
R^{\text{new}} = \min_j \left( R^{M}_{M} \times x^{(1)}, x^{(1)}, \ldots, x^{(M)} \right)
\]

where \(R^{0}_{M}\) encodes the \(M\)-relation contained in \(L_m(1 \leq m \leq M)\) and the existential quantifier \(\exists y\) that quantifies \(M\) free occurrences of \(y\) in \(L_1 \land \cdots \land L_M\) is encoded by an order-\(M\) tensor \(Q^{M}_{M}\) introduced by (8). We call the equation (9) a definition for \(R^{\text{new}}\).

Similarly, if \(Q_{m}^{x} G\) is a \(\forall\)-CNF formula \(\forall y(L_1 \lor \cdots \lor L_M)\), the relation in \(M\) extracted from \(\forall y(L_1 \lor \cdots \lor L_M)\) is encoded by
\[
R^{\text{new}} = \min_j \left( Q^{M}_{M} \times x^{(1)}, x^{(1)}, \ldots, x^{(M)} \right)
\]

where \(R^{0}_{m}\) encodes the \(M\)-relation contained in \(\neg L_m(1 \leq m \leq M)\) (details omitted).

**Compiling prenex formulas**

We now compile a prenex formula \(F = Q_1 x_1 \cdots Q_m x_m G\), using (9) and (10), into an associated set \(\Sigma_F\) of tensor definitions which computes \(\llbracket F \rrbracket\) without grounding. However there is one problem to solve before compilation: (9), for example, is derived from \(\exists y(L_1 \land \cdots \land L_M)\) under the variable condition. When this condition is violated, we need to somehow recover it.

There are two cases where the condition is violated. The first case is that some atom \(r_m(x^{(m)})\) in \(G\) has duplicate occurrences of variables in the arguments \(x^{(m)}\). In this case, let \(R_m\) be a tensor encoding the \(M\)-relation \(r_m\) which is given by (1). Let \(r^{\text{new}}_m(x^{(m)})\) be a new atom defined by

\[
r^{\text{new}}_m(x^{(m)}) \leftrightarrow r_m(x^{(m)})\]

where \(x^{(m)}\) is an enumeration of \(x^{(m)}\) without duplication. It is apparent that a new relation \(r^{\text{new}}_m\) stands for in \(M\) is encoded by a tensor \(R^{\text{new}}_m\) such that
\[
R^{\text{new}}_m(x^{(m)}) = R^{M}_{M}(x^{(m)})
\]
where variables \(x^{(m)}\) and \(x^{(m)}\) run over \(\mathcal{G}'\). We replace every atom in \(G\) that violates the variable condition with a new atom \(r^{\text{new}}_m(x^{(m)})\) described above so that \(R^{\text{new}}_m\) encodes the new \(M\)-relation \(r^{\text{new}}_m\). Let the result be \(G'\) and consider \(F' = Q_1 x_1 \cdots Q_m x_m G'\). Obviously when evaluated in \(M'\), \(F'\) and \(F\) give the same result, i.e.,
\[
\llbracket F' \rrbracket = \llbracket F \rrbracket = \llbracket F \rrbracket
\]
holds. So in the first case, we compile \(F'\) instead of \(F\).

The second case is that, for example, some \(L_i\)s in \(D = \exists y(L_1 \land \cdots \land L_M)\) have no occurrence of \(y\). In this case, we
just shrink the scope of \( \exists y \) and rewrite \( D \) like \( D = \exists y (L_1 \wedge \cdots \wedge L_k) \wedge L_{k+1} \wedge \cdots \wedge L_M \).

Taking these modifications into account, we summarize our compilation procedure in Figure 1. When a model \( M \) and a closed prenex formula \( F \) are given, the compilation procedure returns an algebraic formula \( F_{\text{tensor}} \) and a set \( \Sigma_F \) of tensor definitions. Evaluating \( F_{\text{tensor}} \) using \( \Sigma_F \) gives \( \llbracket F \rrbracket \), the truth value of \( F \) in \( M \).

**A compilation example**

Let \( F_{\text{ABCD}} = \forall x \exists y ((A(x,y) \wedge B(x)) \lor (C(x,y) \wedge D(y))) \). We compile \( F_{\text{ABCD}} \) into a set \( \Sigma_{F_{\text{ABCD}}} \) of tensor definitions along the compilation procedure in Figure 1. Let \( A, B, C \) and \( D \) respectively be tensors encoding \( M \)-relations \( A, B, C \) and \( D \).

Set \( \Sigma_{F_{\text{ABCD}}} = \{ \} \). First we convert \( F_{\text{ABCD}} \)’s innermost subformula \( F_2 \) into \( \exists x \)-DNF:

\[
F_2 = \exists y ((A(x,y) \wedge B(x)) \lor (C(x,y) \wedge D(y)))
\]

\[
= \exists y A(x,y) \lor \exists y C(x,y) \lor \exists y B(x) \lor \exists y D(y).
\]

Next we introduce new atoms and rewrite \( F_2 \) to \( G_2^* \):

\[
r_{A}^\text{new}(x) \iff \exists y A(x,y)
\]

\[
r_{CD}^\text{new}(x) \iff \exists y C(x,y) \lor \exists y D(x)
\]

\[
G_2^* = (r_{A}^\text{new}(x) \land B(x)) \lor r_{CD}^\text{new}(x).
\]

Correspondingly to these new atoms, we construct tensors below which encode the corresponding relations in \( M \) and add them to \( \Sigma_F \): \( R_{A}^\text{new} = \min_1 (Q^{2,1} \times L_A) \) \( R_{CD}^\text{new} = \min_1 (Q^{2,1} \times L_C \times L_D) \).

We put \( F_1 = \forall x F_2 = \forall x G_2^* \) and continue compilation. We convert \( F_1 \) to \( \forall \)-CNF:

\[
F_1 = \forall x ((r_{A}^\text{new}(x) \land B(x)) \lor r_{CD}^\text{new}(x))
\]

\[
= \forall x (r_{A}^\text{new}(x) \lor r_{CD}^\text{new}(x)) \land \forall x (B(x) \lor r_{CD}^\text{new}(x)).
\]

We introduce new atoms and rewrite \( F_1 \) to \( G_1^* \):

\[
r_{ACD}^\text{new}(x) \iff \forall x r_{A}^\text{new}(x) \lor r_{CD}^\text{new}(x)
\]

\[
r_{BCD}^\text{new}(x) \iff \exists x r_{B}^\text{new}(x) \lor r_{CD}^\text{new}(x)
\]

\[
G_1^* = r_{ACD}^\text{new} \land r_{BCD}^\text{new}.
\]

We construct tensors (scalars) for \( r_{ACD}^\text{new} \) and \( r_{BCD}^\text{new} \):

\[
R_{ACD}^\text{new} = 1 - \min_1 (Q^{2,1} \times L_A - R_{A}^\text{new} \times L_B) \ (13)
\]

\[
R_{BCD}^\text{new} = 1 - \min_1 (Q^{2,1} \times L_B - B \times L_B - R_{CD}^\text{new} \times L_B) \ (14)
\]

and add them to \( \Sigma_{F_{\text{ABCD}}} \) Now \( \Sigma_{F_{\text{ABCD}}} = \{ (11), (12), (13), (14) \} \). Finally we put

\[
F_0 = G_1^* = r_{ACD}^\text{new} \land r_{BCD}^\text{new}.
\]

\[
F_{\text{tensor}} = R_{ACD}^\text{new} \land R_{BCD}^\text{new}.
\]

So \( \llbracket F_{\text{ABCD}} \rrbracket \) in \( M \) is evaluated without grounding by computing \( F_{\text{tensor}} \) using \( \Sigma_{F_{\text{ABCD}}} = \{ (11), (12), (13), (14) \} \).

**Binary predicates: matrix compilation**

The compilation procedure in Figure 1 is general. It works for arbitrary prenex formulas \( F \) with binary predicates. However when \( r \) is a binary predicate, the corresponding tensor \( R \) is a bilinear map and represented by an adjacency matrix \( R \) as follows.

\[
\llbracket r(e_i, e_j) \rrbracket = (e_i \cdot R e_j) = e_i^T R e_j = r_{ij} \in \{ 0, 1 \} \ (15)
\]

In such binary cases, we can often “optimize” compilation by directly compiling \( F \) using matrices without introducing \( \Sigma_F \). This is quite important in processing KGs logically as they are a set of ground atoms with binary predicates. Hence we here derive some useful compilation patterns using matrices defined by (15) for formulas with binary predicates. We specifically adopt \( \llbracket F \rrbracket_{\text{Mat}} \) to denote the result of compilation using matrices that faithfully follows (2), (3), (4), (5), (6) and (7) in Subsection.

\[
\llbracket \exists x r_1(x,y) \land \exists y r_2(x,z) \rrbracket_{\text{Mat}} = \llbracket (r_1(x,e_1) \land r_2(e_1,z)) \lor \cdots \lor (r_1(x,e_N) \land r_2(e_N,z)) \rrbracket_{\text{Mat}}
\]

\[
= \min_1 (\sum_{j=1}^N r_1(x,e_j) \land r_2(e_j,z))_{\text{Mat}}
\]

\[
= \min_1 (\sum_{j=1}^N x_i r_1 e_j r_2^T r_2 z)
\]

\[
= x^T \min_1 (R_1 (e_j) R_2)^T z
\]

\[
= x^T \min_1 (R_1 R_2) z \ (16)
\]

Here \( x \) and \( z \) run over \( \mathcal{P} = \{ e_1, \ldots, e_N \} \). Hence the synthesized relation \( r_1(x,y) \overset{\text{def}}{=} \exists y r_1(x,y) \land r_2(y,z) \) is encoded by a matrix \( R_{12} = \min_1 (R_1 R_2) \). What is important with this example, or with binary predicates in general, is the fact that \( Q^{2,2} = \sum_{j=1}^N e_i e_j^T = I \), an identity matrix, holds.

Similarly by applying (16), we can compile a doubly quantified formula \( \forall x \exists y r_1(x,y) \land r_2(x,y) \) as follows: \( \llbracket \exists x \exists y r_1(x,y) \land r_2(x,y) \rrbracket_{\text{Mat}} \). \( \llbracket \exists x \exists y r_1(x,y) \land r_2(x,y) \rrbracket_{\text{Mat}} \). \( \llbracket \exists x \exists y r_1(x,y) \land r_2(x,y) \rrbracket_{\text{Mat}} \).

Here, a Horn formula \( \forall x \exists y r_1(x,y) \Rightarrow r_2(x,y) \) is compiled into

\[
\llbracket \forall x \exists y r_1(x,y) \Rightarrow r_2(x,y) \rrbracket_{\text{Mat}} = \llbracket \neg \exists x \exists y r_1(x,y) \land \neg r_2(x,y) \rrbracket_{\text{Mat}}
\]

\[
= 1 - \min_1 (\text{tr}(R_1 - R_2)^T).
\]

Note that \( \text{tr}(R_1 - R_2)^T \) gives the number of pairs \( (x,y) \) that do not satisfy \( r_1(x,y) \Rightarrow r_2(x,y) \). Consequently \( \text{tr}(R_1 - R_2)^T = \)

\[2\]When \( r(x,y) \) is encoded by \( R \) as \( (x \cdot R y) \), \( r(y,x) \) is encoded by \( R^T \) because \( (y \cdot R x) = (x \cdot R^T y) \) holds.
0 implies every pair \((x, y)\) satisfies \(r_1(x, y) \Rightarrow r_2(x, y)\) and vice versa. Our compilation is thus confirmed correct.

Another, typical, Horn formula \(\exists y r_1(x, y) \land r_2(y, z) \Rightarrow r_3(x, z)\) is compiled into

\[
\begin{align*}
\llbracket \forall x \forall z (\exists y r_1(x, y) \land r_2(y, z) \Rightarrow r_3(x, z)) \rrbracket_{\text{Mat}} &= 1 - \min_1 (\text{tr}(\min_1 (R_1 R_2 - R_3^T))).
\end{align*}
\]

(20)

Again \(\text{tr}(\min_1 (R_1 R_2 - R_3^T))\) is the total number of \((x, z)\)s that do not satisfy \(\exists y r_1(x, y) \land r_2(y, z) \Rightarrow r_3(x, z)\). So our compilation is correct.

**Recursive matrix equations**

Our non-grounding linear-algebraic approach yields tensor equations from logical equivalence, and this property provides a new approach to the evaluation of Datalog programs. We sketch it using a simple example. Consider the following Datalog program that computes the transitive closure \(\exists z\) of a binary relation \(\exists 1\).

\[
x 2 (x, z) : = x 1 (x, z).
\]

(21)

\[
x 2 (x, z) : = x 2 (x, y), x 2 (y, z).
\]

This program defines the least Herbrand model \(M\) where \(x 1\) is interpreted as \(r_1\) and \(x 2\) as \(r_1\). \(r_2(x_1, x_0)\) holds true if-and-only-if there is a chain \(x_1, x_2, \ldots, x_h \in M\) \(h \geq 1\) such that \(r_1(x_1, x_2), r_1(x_2, x_3), \ldots, r_1(x_{h-1}, x_h)\) are all true in \(M\). Then we see the logical equivalence

\[
r_2(x, z) \Leftrightarrow r_1(x, z) \lor \exists y(r_1(x, y) \land r_2(y, z))
\]

(21)

holds for all \(x, z\) in \(M\). That means

\[
\llbracket r_2(x, z) \rrbracket = \llbracket r_1(x, z) \lor \exists y(r_1(x, y) \land r_2(y, z)) \rrbracket
\]

(22)

also holds for any \(x, z\). Let \(R_1\) and \(R_2\) be adjacency matrices encoding \(r_1\) and \(r_2\) in \(M\) respectively. We translate (22) in terms of \(R_1\) and \(R_2\) as follows.

\[
x^T R_2 z = \llbracket x^T r_2(x, z) \rrbracket
\]

\[
= \llbracket x^T r_1(x, z) \lor \exists y r_1(x, y) \land r_2(y, z) \rrbracket
\]

\[
= \min_1 (x^T R_1 z + x^T \min_1 (R_1 R_2) z)
\]

\[
= x^T \min_1 (R_1 + R_1 R_2) z
\]

Since \(x, z \in \mathscr{D}'\) are arbitrary, we reach a recursive equation

\[
R_2 = \min_1 (R_1 + R_1 R_2).
\]

(23)

It is to be noted that when considered an equation for unknown \(R_2\), (23) may have more than one solution\(^5\) but we can prove that the transitive closure is the “least” solution of (23) in the sense of matrix ordering\(^6\)(proof omitted).

(23) is a nonlinear equation due to \(\min_1\) operation, it looks impossible to apply a matrix inverse to obtain \(R_2\). However we found a way to circumvent this difficulty and proved that it is possible to obtain \(R_2\) by computing (24) and (25) as follows.

\[
R_2 = (R_1^d) > 0
\]

(24)

\[
R_2^d = (I - \epsilon R_1)^{-1} \epsilon R_1
\]

(25)

Here \((R_1^d) > 0\) means to threshold all elements in \(R_2\) at 0, i.e., positive ones are set to 1, o.w. to 0\(^5\).

**Experiment with transitive closure computation**

We compared our linear algebraic approach to Datalog evaluation with state-of-the-art symbolic approaches using two tabled Prolog systems (B-Prolog (Zhou, Kameya, and Sato 2010) and XSB (Swift and Warren 2012)) and two ASP systems (DLV (Alviano et al. 2010) and Clingo (Gebser et al. 2014)). Although we conducted a number of experiments computing various programs with artificial and real data, due to space limitations, we here pick up one example that computes the transitive closure of random matrices. In the experiment\(^8\), we generate random adjacency matrices by specifying the number of dimension \(N\) and the probability \(p_e\) of each entry being 1 and compute their transitive closure matrices using (24) and (25). We set \(N = 1000\) and vary \(p_e\) from 0.0001 to 1.0 and measure the average computation time over five runs (details omitted).

<table>
<thead>
<tr>
<th>(p_e)</th>
<th>Matrix</th>
<th>B-Prolog</th>
<th>XSB</th>
<th>DLV</th>
<th>Clingo</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0001</td>
<td>0.096</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0.001</td>
<td>0.094</td>
<td>0.004</td>
<td>0.003</td>
<td>0.293</td>
<td>0.038</td>
</tr>
<tr>
<td>0.01</td>
<td>0.117</td>
<td>2.520</td>
<td>1.746</td>
<td>10.657</td>
<td>14.618</td>
</tr>
<tr>
<td>0.1</td>
<td>0.165</td>
<td>18.382</td>
<td>16.296</td>
<td>75.544</td>
<td>125.993</td>
</tr>
<tr>
<td>1.0</td>
<td>0.100</td>
<td>188.280</td>
<td>137.903</td>
<td>483.380</td>
<td>1,073.301</td>
</tr>
</tbody>
</table>

Table 1: Average computation time for transitive closure computation (sec)

Table 1 shows the result. Our approach is termed “Matrix” in the table. Two observations are clear. First the computation time of our approach, Matrix, is almost constant while others seem linear w.r.t. \(p_e\). Second, when \(p_e\) is small, \(p_e = 0.0001 \sim 0.001\) and matrices are sparse, the Matrix method takes more time than existing systems but when \(p_e\) gets bigger, it runs orders of magnitude faster than them. The same observation is made with other programs (details omitted).

**Related work**

There is not much literature concerning first-order logic embedded in vector spaces. The most related work to ours is a formalization of first-order logic in tensor spaces by Grefenstette (Grefenstette 2013). He actually proposed two formalizations. The first one represents entities by one-hot

\(^{3}\) For example, \(R_2 = I \circ I\) is a solution.

\(^{4}\) Matrices \(A = \{a_{ij}\}\) and \(B = \{b_{ij}\}\) are ordered by \(A \leq B\) if and-only-if \(a_{ij} \leq b_{ij}\) for all \(i, j\).

\(^{5}\) The proof and details are stated in an accompanying paper submitted for publication.

\(^{6}\) All experiments are carried out on a PC with Intel(R) Core(TM) i7-3770@3.40GHz CPU, 28GB memory.
vectors, predicates by adjacency tensors and truth values by two-dimensional vectors (true by $\top = [1,0]^T$ false by $\bot = [0,1]^T$). AND and OR are order-3 order tensors whereas NOT is a $2 \times 2$ matrix that maps $\top$ to $\bot$ and vice versa. The first formalization can completely formalize a quantifier-free fragment of first-order logic in finite domains. The second formalization represents a finite set by a vector of multiple fragments of first-order logic in finite domains. The second formalization remains an open problem to his tensor approach.

Krompaß et al. (Krompaß, Nickel, and Tresp 2014) proposed a way of answering existential queries of the form $\exists x Q_1 \land Q_2$ in the context of low-dimensional embeddings. Their approach however does not assign an independent representation to existential quantifiers and is limited to a narrow class of the form $\exists x Q_1 \land Q_2$.

We found no literature on computing the least model of Datalog programs via solving recursive matrix equations. So the transitive closure computation presented in this paper is possibly the first example of this kind.

Conclusion

We proposed a general approach to evaluate first-order formulas $\exists$ in prenex normal form in vector spaces. Given a finite model $M$ with $N$ entities, we compile $F$ into a set $\Sigma_F$ of hierarchical tensor definitions (equations) with a nonlinear operation. Computing $\Sigma_F$ in $R^N$ yields the truth value $[F]^M$ in $M$. In this compilation process, tensor representation $Q^{\exists M}$ is introduced to existential quantifiers themselves for the first time as far as we know. Since our approach does not rely on propositionalization of first-order formulas, it can derive tensor equations from logical equivalences. We exploited this property to derive recursive matrix equations to evaluated Datalog programs. We empirically demonstrated the effectiveness of our linear algebraic approach by showing that it runs orders of magnitude faster than existing symbolic approaches when matrices are not too sparse.

References


