An idealistic treatment of quantifiers in categorical compositional distributional semantics
Extended Abstract

Ondřej Rypáček
University of Oxford
Dept. of Computer Science
ondrej@cs.ox.ac.uk

Mehrnoosh Sadrzadeh
Queen Mary University of London
School of Electronic Eng. and Computer Science
mehrs@eecs.qmul.ac.uk

Abstract. We show how one can formalise quantifiers in the categorical compositional distributional model of meaning. Our model is based on the generalised quantifier theory of Barwise and Cooper. We develop an abstract compact closed semantics and instantiate it in the category of relations.

1 Quantification in Sets

The category Rel of sets and relations is a basic example of a symmetric monoidal compact closed category. The tensor is Cartesian product, $\times$, and monoidal the unit is the one element set, 1, $A^* = A$. Monoidal closure is therefore just the product. As an example consider the correspondence between relations and subsets of the Cartesian product expressed formally as

$$\mathrm{Rel}(1, X \times Y) \;\cong\; \mathrm{Rel}(X, Y) \;\cong\; \mathrm{Rel}(X \times Y, 1)$$

where we are repeatedly using compact closure and the fact that relation beween 1 and Z are demonstrably subsets of Z.

O.R. This is probably best put where compact closed cats are introduced.

The frobenius maps Δ is just the diagonal relation $x \sim (x, x)$ and ∇ its transpose; $\perp : 1 \longrightarrow X$ is the so-called *fan* relation: $* \sim x,\ \forall x$, $\top$ its converse. The required axioms are easily checked.

O.R. This is probably best put where Frobenius structure is introduced.

In order to introduce quantification in Rel, we first discuss the well known case in Set where the existential and universal quantifiers are left and right adjoints, respectively, to the inverse image functor. More formally this can be described as follows. For a set X, the set of functions from X to the two-element set, 2^X, is the powerset of X. For a function $f : X \longrightarrow Y$, precomposition with f is internally the map $2^f : 2^Y \longrightarrow 2^X$ that takes a h to hf. It has both left and right adjoints with respect to the pointwise subset ordering. In more detail, the powerset of X ordered by subset inclusion is isomorphic to the set of characteristic functions on X, 2^X, ordered pointwise by the ordering $0 < 1$ of 2. Any pre-ordered set can be seen as a category[1]. The function 2^f becomes a functor[2] from the category 2^Y to 2^X. This functor has both adjoints. It is well known that when f is the first projection $\pi_1 : X \times Y \longrightarrow X$, 2^{π_1} corresponds to *weakening* in the sense that it takes predicates over context X to predicates over the larger context, $X \times Y$, where the second component in Y doesn't occur. The left adjoint to 2^{π_1} interprets

[1] internal category?
[2] internal functor?

existential quantification, and the right adjoint interprets universal quantification. In more detail, the left adjoint, $\exists_f : 2^X \longrightarrow 2^Y$, is defined by

$$\exists_f(X' \subseteq X) = \{y \in Y \mid \exists x \in X. fx = y \wedge x \in X'\} \tag{1}$$

i.e. the image of X' under f. It is also known as the *direct image* of f. The right adjoint $\forall_f : 2^X \longrightarrow 2^Y$ is defined by

$$\forall_f(X' \subseteq X) = \{y \in Y \mid \forall x \in X. fx = y \implies x \in X'\}$$

There are other functions from $2^{X \times Y}$ to 2^X, such as the one sending an f to the g for which $g(x) = 1$ if and only if $f(x, y) = 1$ for exactly two distinct values of y. This quantifier could be called "two". Another example is "none" for which $g(x) = 1$ iff $f(x, y) = 0$ always. This exhibits the existential and universal quantifiers as two extreme cases of a spectrum of all possible quantifiers[3] .

In the following we assume for simplicity that the ambient context, X, is empty, i.e. $X = 1$, and then quantifiers become certain functions $2^Y \longrightarrow 2$. Everything we say generalises straightforwardly for arbitrary X. In this case, the existential quantifier becomes the characteristic function of nonempty subsets of Y; the universal quantifier is the characteristic function of the singleton $\{Y\}$, and "two" is the characteristic function of two-element sets.

In single sorted first order predicate logic, quantification ranges over the whole domain Y, which is simply assumed to be fixed, and often implicit, throught the formula. For example, in $\exists x. x > 0$, the bound variable, x, ranges over some ordered set. In linguistics, however, the range of quantification is explicitly stated, as in "all men sleep", where the word "men" restricts the range of the quantifier, "all", to just men. In other words, the quantifier is "all men", and "all" is a particle which expects a linguistic category to become a quantifier. Formally, when we fix a domain of all possible subjects of quantification, Y, such as all nouns, a quantifier Q must have type

$$Q \ : \ 2^Y \longrightarrow 2^{2^Y} \tag{2}$$

Here, the first argument to Q is the range of the quantification. The result is a quantifier which possibly ignores everything that is outside the range.

For instance, "some" ($\exists$) takes a subset $\mathrm{men} \subseteq X$ to all nonempty subsets of men; all ($\forall$) maps $\mathrm{men} \subseteq X$ to $\{\mathrm{men}\}$; "two" takes $\mathrm{men} \subseteq X$ to two-element subsets of men.

2 Relations

We show two alternative categorical views on relations. The first one understands a relation R between sets X and Y as a function into a powerset of Y, namely the function r that assigns to each x the set of all $y \in Y$ related to x. Formally, $r(x) = \{ \ y \in Y \mid x \sim_R y \ \}$. On the other hand the same relation R is a subset of the Cartesian product $X \times Y$ of those pairs (x, y) for which $x \sim_R y$. We now give the standard Categorical presentation of these two points of view on relations.

Relations as Kleisli arrows

The assignment $X \mapsto 2^X$ extends to a functor $2^{(-)} : \mathrm{Set} \longrightarrow \mathrm{Set}$ where the action on arrows is direct image (1). The functor is a monad where the unit, η, maps x to $\{x\}$ and multiplication, μ, is set union. Recall [2] that for any monad $(\mathbf{T}, \mu, \eta)$ on $\mathbb{C}$, the Kleisli category of $\mathbf{T}$, denoted $\mathbb{C}_{\mathbf{T}}$, has the same objects as $\mathbb{C}$. Arrows $X \longrightarrow Y$ in $\mathbb{C}_{\mathbf{T}}$ are arrows $X \longrightarrow \mathbf{T}Y$ in $\mathbb{C}$. Moreover, there is an adjunction

[3] Uhm, yes. That is precisely what it means to be a left and right adjoint.

$\mathbf{F} \dashv \mathbf{U} : \mathbb{C} \longrightarrow \mathbb{C}_{\mathbf{T}}$ such that $\mathbf{F}$ is identity on objects, and $f \mapsto \eta \cdot f$ on arrows and $\mathbf{U}$ is $\mathbf{T}$ on objects and $\mu \cdot \mathbf{T}f$ on arrows. Moreover, $\mathbf{UF} = \mathbf{T}$.

Thus by Rel one can understand the Kleisli category of $2^{(-)}$, which is understood as the category of sets and relations where a function $r : X \longrightarrow 2^Y$ defines a relation $\sim_R$ by

$$x \sim_R y \quad \equiv \quad y \in fx$$

Relations as spans in Set

On the other hand the same relation R defines a subset of $X \times Y$ by $R = \{ (x, y) \mid x \sim_R y \}$. Such a set is called the *tabulation of* R. In Set, R with the two projections is a *span* over X and Y:

$$\begin{tikzcd}
 & R \arrow[dl, "p"'] \arrow[d, dashed, "\langle p,q \rangle"] \arrow[dr, "q"] & \\
X & X \times Y \arrow[l, "\pi_1"'] \arrow[r, "\pi_2"] & Y
\end{tikzcd}$$

Spans form a (bi)category where composition is defined using pullback as in the following diagram:

$$\begin{tikzcd}
 & & R \times_Y S \arrow[dl] \arrow[dr] & & \\
 & R \arrow[dl, "p"'] \arrow[dr, "q"] & & S \arrow[dl, "r"'] \arrow[dr, "s"] & \\
X & & Y & & Z
\end{tikzcd}$$

where $R \times_Y S$ is the pullback of q and r. The unit is then just the span of identities:

$$\begin{tikzcd}
 & X \arrow[dl, "\mathsf{id}"'] \arrow[dr, "\mathsf{id}"] & \\
X & & X
\end{tikzcd}$$

For an arbitrary category $\mathcal{C}$, the category of spans in $\mathcal{C}$, formally $\mathrm{Span}(\mathcal{C})$, is a bicategory where a 2-cell $R \Rightarrow S : X \longleftrightarrow Y$ is the vertical arrow in the following commuting diagram:

$$\begin{tikzcd}
 & R \arrow[dl, "p"'] \arrow[dr, "q"] \arrow[dd, dashed] & \\
X & & Y \\
 & S \arrow[ul, "r"] \arrow[ur, "s"'] &
\end{tikzcd}$$

We leave out the details that this indeed forms a bicategory. We leave the compact closed monoidal structure of $\mathrm{Span}(\mathrm{Set})$ as an easy exercise.

Finally, note that the correspondence between spans and relations is an equivalence rather than an isomorphism, as there may be many spans representing (tabulating) the same relation. For instance, whenever $X \xleftarrow{p} R \xrightarrow{q} Y$ tabulates a relation, $X \xleftarrow{[p,p]} R+R \xrightarrow{[q,q]} Y$ tabulates the same relation. However, all such spans are equivalent via suitable 2-cells[4]. Thus, strictly speaking Rel is the quotient posetal[5] bicategory $\mathrm{Span}(\mathrm{Set})/_{\simeq}$ where $\simeq$ is the above-described equivalence.

2.1 Relations in arbitrary categories

The construction of the category of relations $\mathrm{Span}(\mathcal{C})$ can be generalised to an arbitrary category with pullbacks. The resulting category is compact closed if $\mathcal{C}$ is.

Remark 1. It seems that associativity of composition in $\mathrm{Rel}\mathcal{C}$ has to do with the existence of a factorisation system (both our key examples have one).

3 Powerset Relations

In this section we study relations over powersets of sets, i.e. relations between sets 2^X and 2^Y for some sets X, Y.

3.1 Definition

The following is the key observation of this section:

Proposition 1. *A relation between powersets is a relation in* Rel.

Proof. Given a relation $R : 2^X \longleftrightarrow 2^Y$ let T with $\pi_X : T \longrightarrow 2^X$ and $\pi_y : T \longrightarrow 2^Y$ be its tabulation. Each of the projections is a Kleisli arrow, i.e. a relation in the first sense:

$$\begin{tikzcd} & T \arrow[dl, "\pi_X"'] \arrow[dr, "\pi_2"] & \\ 2^X & & 2^Y \end{tikzcd}$$

This is a span

$$\begin{tikzcd} & T \arrow[dl, "\pi_X"'] \arrow[dr, "\pi_2"] & \\ X & & Y \end{tikzcd}$$

in Rel. On the other hand given a relation of relations R:

$$\begin{tikzcd} & R \arrow[dl, "\pi_X"'] \arrow[dr, "\pi_2"] & \\ X & & Y \end{tikzcd}$$

[4] Not true! $R + R$ is not $\cong R$

[5] where between each pair $f, g : X \longrightarrow Y$ of 1-cells is at most one 2-cell.

by passing from the Kleisli to the underyling category one obtains a span in Set whose domain and codomain are powersets. Note that we are not applying the underlying set functor, just forgetting the Kleisli category structure. Explicitly, the ordinary relation obtained from R is:

$$x \subseteq X \;\sim_R\; y \subseteq Y \quad \equiv \quad \exists r \in R.\, \pi_X(r) = x \wedge \pi_Y(r) = y$$

Less formally: a relation is a subset[6], thus a relation on relations is a relation on subsets.

3.2 Limits in Rel

Remark 2. The key question in all this is how does taking Rel of something preserve limits. Are there limits in spans? In or at least in the concrete case of $\mathrm{Rel}(\mathrm{Set})$? Does FdVect have limits?

4 Quantifiers as Relations

It follows from what was said thus far that equation (2) exhibits a quantifier as a relation $2^Y \longrightarrow 2^Y$.

So in order to interpret generailed quantifiers in Rel, we look into relations over powersets.

A table of some well known and some less known quantifiers is below. Here, X, Y are all subsets of a set U. I.e. a quantifier is a relation $2^U \longleftrightarrow 2^U$. We define a relation in the Kleisli style.

$$\begin{aligned} \forall: &\quad X \sim \{X\} \\ \exists: &\quad X \sim \{\, Y \mid Y \neq \emptyset \,\} \\ 2: &\quad X \sim \{\, Y \mid |Y| = 2 \,\} \\ \text{none}: &\quad X \sim \{Y \mid Y \cap X = \emptyset\} \end{aligned}$$

Where $|Y|$ is the cardinality of Y.

5 Relations in FdVect

The category FdVect is compact closed and has pullbacks, so we can form $\mathrm{RelFdVect}$, which is compact closed.

Definition 1. *Relations in* $\mathrm{Rel}(\mathrm{FdVect})$ *are called* linear relations

Proposition 2. *For a linear relation $\sim$ holds that for all i, $v_i \sim u_i$ iff $\sum_i \alpha_i v_i \sim \sum_i \alpha_i u_i$. In words: linear relations are closed under linear combinations.*

Remark 3. The above proposition does not imply that if $u+v$ is related to some $x+y$ than u is related to x and v is related to y or something similar. That would not even make sense in general. When defining linear relations, we explicitly name the vectors (linear combinations of basis vectors) that are related and then close under linear combinations. For example if $u + v \sim x$ and $v \sim y$ then $u + 2v \sim x + y$ follows but $u + v$ is not related to y.

[6] of the cartesian product of the relation's domain and codomain

5.1 Examples

Let the basis be $\mathsf{J}, \mathsf{B}, \mathsf{M}, \mathsf{S}$ (for John, Ben, Mary and Sarah). Define $\sim$ for "love" as $\mathsf{J} \sim \mathsf{M}$ and $\mathsf{B} \sim \mathsf{S}$. Then necessarily, $\mathsf{J}+\mathsf{B} \sim \mathsf{M}+\mathsf{S}$. In words: if John loves Mary and Ben loves Sarah, then John + Ben love Mary + Sarah.

If John loves Mary, and Ben also loves Mary, then John + Ben love $2\times$ Mary. I.e. Mary has twice as much love.

Let $I \cong \mathbb{K}$, be the monoidal unit and let $\mathsf{men} : I \longrightarrow N$ relate the unit vector $|0\rangle$ to all vectors denoting men. Then necessarily each linear combination of men is related to $\alpha|0\rangle$, for some scalar α.

We said that quantifiers are relations over subsets. For instance, $\exists$ (or "some") relates each set X to each of its nonempty subsets. When moving to $\mathrm{Rel}(\mathrm{FdVect})$ subsets become linear combinations. Thus quantifiers become relations over linear combinations. For instance "some" relates each linear combination, eg. $u+v$ to each of its factors, ie. $u+v \sim u$, $u+v \sim v$, $u+v \sim u+v$. The last case also follows by linearity from $u \sim u$ and $v \sim v$ but not the other ones unless either u or v are related to 0. All these cases are necessary in the correct interpretation of quantification. The quantifier $\forall$ ("all") relates $u+v$ to $u+v$ only.

It's a question how to deal with scalars. Ideally we want to act as if all scalars were equal , thus $\forall$ should relate $u+v$ to $u+v$ but also $u+2v$ to $u+2v$ (or $u+v$?). This should be easily done, as the cardinality of the set of all finite sequences of all reals is just the reals (via some Church encoding using exponents of primes, right? Real exponents?! We can say an infinite countable sequence of reals, that should be a real, and the finite ones are a subset of those which taper off after some finite number).

6 Concluding Remark

Doing relations over FdVect is like doing set theory over vectors. Only in this case closed under linear combinations. We don't even have to restrict sets to the basis. By that I mean, pick some vectors $m_1, m_2, \ldots, m_i$ in N as the set of men. Then define "some" as a relation (ie. there's no requirement for linearity) on linear combinations of m_i so that each linear combination $\sum_i \alpha_i m_i$ is related to $\sum \overline{\alpha_i} m_i$ where $\overline{\alpha}$ sets α *nondeterministically* to 0, but in such a way that at least one alpha is nonzero[7]. And possibly, following the remark above, for all possible other scalar multiples[8]. Then "some men sleep" will give a nonzero if some nonempty linear combination of m_i sleeps.

In relation to the Rel case, in an ambient vector space N, let M be a subset of its vectors denoting men. Then some men are all its nonempty subsets, and we can make sense of all first order logic, and generalised quantifiers, in terms of sets. The fact that we are doing $\mathrm{Rel}(\mathrm{FdVect})$ just means we moreover close our relations under linear combinations. And by that we seem to arrive at a compact-closed category again. I have yet to understand how precisely.

The question is, what are we getting from this second step: linearity? Why don't we do just relations over vectors when we want a logic. The linearity adds something: a degree of truth perhaps. For instance, is "some men sleep" is true for more men, the truth is bigger. If for two men it's 3 (1 for each singleton, 1 for the couple), etc. Is it worth anything? It's highly dependent on the presenation, decomposition of linear compositions. Some linear compositions are more decomposable than others. If we are after a boolean meaning we can do just relations over (sets of) vectors.

[7] You know what I mean

[8] This is a bit iffy, because we would technically like vector space over the two-element field, but the only two-element field there is is the one where + acts like xor and we need proper or.

References

1. Beck, J.: Distributive laws. In: Seminar on Triples and Categorical Homology Theory, Lecture Notes in Mathematics, vol. 80, pp. 119–140. Springer Berlin / Heidelberg (1969), http://dx.doi.org/10.1007/BFb0083084, 10.1007/BFb0083084
2. MacLane, S.: Categories for the Working Mathematician. Springer
3. Mulry, P.: Lifting theorems for kleisli categories. In: Brookes, S., Main, M., Melton, A., Mislove, M., Schmidt, D. (eds.) Mathematical Foundations of Programming Semantics, Lecture Notes in Computer Science, vol. 802, pp. 304–319. Springer Berlin Heidelberg (1994), http://dx.doi.org/10.1007/3-540-58027-1_15