Towards universal quantification in distributional semantic space

Matthew Capetola
Universtity of Oxford Faculty of Linguistics, Philology, and Phonetics
Clarendon Press Institute, Walton Road
Oxford OX1 2HG, United Kingdom
matthew.capetola@wolfson.ox.ac.uk

Abstract

This paper defines a representation of universal quantification within distributional semantic space. We propose a discourse-internal approach to the meaning of limited instances of *every*, highlighting the possibilities and limitations of doing textual logic in a purely distributional framework.

1 Introduction

Research in recent years has moved to applying distributional semantic space models to tasks that deal in more complicated meaning structures like phrases and sentences. The underlying question in those applications is how to model *compositionality*, or the idea that the meaning of a larger linguistic unit is a function of its parts. This has typically amounted to describing a correspondence between the combinatorial operations available for linear algebraic structures, like vector addition and matrix multiplication, and the (hypothetical) compositional operations of natural language (Baroni and Zamparelli, 2010; Coecke et al., 2010; Mitchell and Lapata, 2008).

While this has yielded high performance on semantic tasks like sentiment analysis (Socher et al., 2012; Socher et al., 2013) and paraphrase detection (Mitchell and Lapata, 2008; Mitchell and Lapata, 2010; Blacoe and Lapata, 2012), the issue of capturing textual logic, also inextricably linked to compositionality, remains an open problem. Part of the reason for this is that there exists structural opposition between distributional and formal semantics (Grefenstette, 2013). The benefit of distributional semantics is that its vectors represent meanings with high-dimensional subtlety. This allows us to model the way content words modify each other compositionally, better than we might using the comparably flat meaning representations of formal semantics. However, formal semantics is far better equipped to handle the meanings of function words like quantifiers on both individual and interstitial levels. The syntax of formal calculi unambiguously denotes the interrelation of functional operators in a logical expression. In essence, formal semantics treats content words as properties that obtain of entities in a referent domain, or model world. Quantification is then conceived of as a higher-order operation describing relations between the sets of entities circumscribed by such properties. In this way, a comprehensive semantic system in natural language is comprised of a universe of functions built over a universe of entities (Peters and Westerstahl, 2008).

So while the the distributional framework has been successful precisely by doing away with the set-theoretic approach to semantics (Baroni and Zamparelli, 2010), it faces several foreseeable problems, two of which we focus on here.

1. The meaning of a quantifier expression doesn't appeal to features of the domain of quantification. That is, no matter what entities engaging in whatever kind of verbal relation, the meaning of a quantifier like *every* does not change: it stands for the inclusion relation between sets. (Peters and Westerstahl, 2008).

2. Without a model world, or universe of entities over which to quantify, it is unclear what quantifiers mean.

Considering the above points, there is no *a priori* reason to expect that distributional representations make sense for function words. In light of these issues, recent research has moved towards merging distributional semantics with formal compositional calculi like Lambek calculus, leveraging the distinct strengths of both approaches selectively (Lewis and Steedman, 2013).

This paper begins by highlighting some of the persistent expressive differences between distributional and formal semantics. This will help to motivate a limited definition of the universal quantifier *every*, while remaining within a purely distributional framework. It is our belief that further inquiry into this field despite the initially perceived limitations has the potential to produce theoretically and pragmatically impactful results.

2 Mending the structural opposition?

2.1 Previous work

This paper continues in the line of inquiry which has been previously referred to in the literature as "logic for vectors" (Hermann et al., 2013). It seeks to define the meaning of a function word, and textual logic in general, within distributional semantic space. Hermann et al. (2013) is one of the first papers to concern itself explicitly with the meaning of a function word *not*, *relative to* distributional representations of content words. This contrasts with the aforementioned distributional-formal hybrid approaches, as well as the recent work of Grefenstette (2013). The latter models truth-functional logic using the operations of tensor calculi, rather than redefining what logical words mean altogether when we move to a distributional context.

Integral to the discussion here, as well as the "tripartite representation" of meaning in Hermann et al. (2013), is the concept of a dual-space representation similar to those of Turney (2012). A dual-space model posits that the mathematical structure of a word is comprised of two block components: a domain and a value. A domain is extracted via similarity metrics, and serves to group a term with others according to overarching semantic similarity in the space (Turney, 2012). Hermann et al. give the example that terms *red* and *blue* have very different values, but share the common conceptual domain *colors*. Important for the ideas here is that semantic domains are defined by appeal to other terms within the same semantic space. Taking this further, we will treat semantic domains as higher-order structures: divisions of the semantic space into subspaces, or sets of concepts.

2.2 Domains vs. ontologies

Previous work has shown that imposition of domain structure on a semantic space model affords some additional expressiveness for defining the meanings of function words. We ought to ask to what extent this is the case. Of particular interest in this section is the relationship between *domain* structure of distributional semantics and *model world* structure of formal semantics.

Consider the sentence *All boys are good* $\equiv \forall x : boy(x) \Rightarrow good(x)$. The quantifier is integral to the logical meaning of this sentence. If we eliminate it, we can express the general notion that the concept *boy* is good, by composing distributional representations of the two lexemes. This however, is not as ontologically rich as the formal interpretation. In a model-theoretic semantics, *boy* serves as an ontological domain of entities which are boys. A distributional model, on the other hand, does not postulate the existence of hypothetical wold that is populated by entities. It intentionally does away with this set-theoretic representation. Keeping this structural assumption, what can the universal quantifier mean?

Now consider the sentences *Every country attended* and *Every color is good*. Unlike *boys*, *colors* and *countries* can serve as hypernyms denoting sets of concepts that are learned in discourse. So while while *red* is indeed a color, it is also lexically and conceptually distinct, and a distributional model would learn a representation for *red* which maintains this duality. In contrast, boys in a model world are distinct by virtue of being separate entities, as opposed to distinct concepts. Similarly, for the sentence *Everything red is good*, the donation of *red* in our model are those things in the world which bear the property of being red. When *red* serves as a conceptual domain however, as in the sentence *All reds are good*, it is referring not to entities, but to the set of concepts denoting shades of red.

Another distinction to be drawn is that that our definition of quantification with *every* must be further-confined to cases in which semantic domains are denoted by count nouns. Count nouns are common nouns that can be enumerated and can appear in both single and plural form. In contrast, for other kinds of semantic domains like *politics*, which is a viable concept under which one could group terms in a semantic space, the meaning of quantification changes in subtle ways. *All politics is interesting* ought to have a very different semantic content than a sentence like *Every country attended*.

These distinctions allow us to define, within disributional space, a notion of quantification over countable concepts, but not quantification over mass nouns, entities, or topical concepts. As a general result, we see that dual-space approaches eschew some of the need for an entity-coded ontological structure. It can be thought that the imposition of domains on a semantic space is a way of reclaiming part of the higher-order structure of an ontology, just not all of it. In general, it would seem that the significance of quantifiers in a semantic model is directly proportional to the descriptive capacity and structural advancement of the ontology of that model.

3 Discursive *every*

Consider the sentence *Red is good*. Ignoring the copula *is*, the meaning of the sentence is, formally, a function application of the meaning of *good* to the meaning of *red*, producing *good(red)*. Now consider the sentence *Every color is good*. The formal semantic interpretation of this sentence is as follows:

1. (a) Every color is good. $\equiv$
   (b) $\{x|color(x)\} \subseteq \{x|good(x)\} \equiv$
   (c) $\forall x : color(x) \Rightarrow good(x)$

What is being said is structurally distinct from the meaning of the first sentence considered, and this is because of the intervention of the function word *every*. As in the formal semantics presented above, the sentence means that for any term bearing the domain *color*, that color is good. The quantifier is said to range over entities for which the property *color* obtains. So, this returns not a single sentential representation, but a set of sentential representations such that the property *good* is applied to the elements contained in the semantic domain *color*:

{*Red is good. Blue is good. ...*}

Provided with our dual-space representation, and assuming $\bullet$ represents our compositional strategy and $*$ represents the Kleene Star, a compact representation of this in vectorial calculi is as follows:

$$\begin{bmatrix} f_{good} \end{bmatrix} \bullet \begin{bmatrix} d_{color} \\ * \end{bmatrix}$$

This example puts forth two claims.

1. A sentence which contains the quantifier *every* is by some measure semantically richer than an unquantified sentence.
2. So presented, universal quantification over conceptual domains does not require postulation of a hypothetical model world. Instead, we can treat it as a function from a sentence in discourse to a set of sentences of lower-order meaning comprised of terms from the same discourse.

Formally, the function mentioned in 2. is as follows:

$$f_{every} : S^2 \rightarrow 2^S$$

Where S^2 represents the set of higher-order sentences as described, and 2^S the power set of the set of lower-order sentences.

The obvious appeal of such a representation is that given a more comprehensive treatment and assuming an appropriate compositional model, the values manipulated in this variety of textual logic are of the same mathematical form as the sentences upon which we wish to do inference. They are themselves sentences.

With this definition, we can more formally express the difference between quantification and this proposed idea of quantification over concepts, revisiting a comparison of the domains *boys* and *colors*. If we have learned M subelements of the domain of *colors*, of the N possible colors in a universe of concepts, then $M \leq N$ and:

- *Every color is good.* $\mapsto \bigcup_{i=1}^{M \leq N} color_i$ *is good.*

In contrast, this does not work for quantification over *boy*, because a distributional representation does not learn separate, indexable representations boy_i. These indexed "boys" would denote separate entities, not separate concepts.

As of now, even for conceptual domains of countable concepts, the definition of f_{every} we've provided has a marked shortcoming. It is limited to the subelements of domains for which our model has learned distributional representations. Leaving the functional value of *every* as defined, we would be treating the semantic space as a static proxy for a more complicated ontological structure. So, for example, if we haven't learned a representation for *cerulean*, the projection from *every color* will not include it. In order to do so, this will likely require a dynamic representation of quantification, perhaps one which is capable of modeling inference on subconcepts predictively, or stochastically.

4 Concluding remarks

Confined to the discussion here, progress needs to be made to extend the applicability of f_{every} towards the goal of dynamic inference. This should be rooted in specific computational semantic tasks. The implications of this approach to quantification should then be brought to bear on more complex issues. Can we conceive of constructing a consistent system of "logic for vectors" such that we can consider more syntactically and semantically complex sentences?

References

M. Baroni and R. Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1183–1193. Association for Computational Linguistics.

W. Blacoe and M. Lapata. 2012. A comparison of vector-based representations for semantic composition. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing*.

Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *CoRR*, abs/1003.4394.

E. Grefenstette. 2013. Towards a formal distributional semantics: Simulating logical calculi with tensors. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics*.

Karl Moritz Hermann, Edward Grefenstette, and Phil Blunsom. 2013. "not not bad" is not "bad": A distributional account of negation". *Proceedings of the 2013 Workshop on Continuous Vector Space Models and their Compositionality*, August.

Mike Lewis and Mark Steedman. 2013. Combined logical and distributional semantics. *Transactions of the Association for Computational Lingusitics*, 1:179–192.

J. Mitchell and M. Lapata. 2008. Vector-based models of semantic composition. In *Proceedings of ACL*, volume 8.

J. Mitchell and M. Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science*.

Stanley Peters and Dag Westerstahl. 2008. *Quantifiers in Language and Logic*. OUP, Oxford.

Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. 2012. Semantic Compositionality Through Recursive Matrix-Vector Spaces. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Richard Socher, Alex Perelygin, Jean Wu, Christopher Manning, Andrew Ng, and Jason Chuang. 2013. Recursive models for semantic compositionality over a sentiment treebank. In *Conference on Empericial Methods in Natural Language Processing*.

Peter D. Turney. 2012. Domain and function: A dual-space model of semantic relations and compositions. *J. Artif. Intell. Res. (JAIR)*, 44:533–585.