Graded Entailment for Compositional Distributional Semantics

Dea Bankova Bob Coecke Martha Lewis Dan Marsden
University of Oxford
{coecke, marlew, daniel.marsden}@cs.ox.ac.uk

The categorical compositional distributional model of natural language provides a conceptually motivated procedure to compute the meaning of sentences, given grammatical structure and the meanings of words. However, until recently it has lacked the crucial feature of lexical entailment. We propose a graded measure of entailment, exploiting ideas from partial knowledge in quantum computation.

Our main theorem shows that entailment strength lifts compositionally to the sentence level, giving a lower bound on sentence entailment. We describe the essential properties of graded entailment such as continuity, and provide a procedure for calculating entailment strength.

This is an abstract of the paper *Graded Entailment for Compositional Distributional Semantics*, available at http://arxiv.org/abs/1601.04908

1 Introduction

Categorical compositional distributional models of meaning [4] provide techniques for constructing the meaning of a sentence from its grammatical structure and the meanings of its parts. Until recently, these models lacked the crucial feature of lexical entailment. Like [1] we capture a notion of entailment by using density matrices rather than vectors. Previous work used a discrete Boolean form of entailment. We provide a graded measure quantifying the strength of the entailment that can model realistic linguistic phenomena. We show that our approach applies to a wider range of sentence types than previous models, and that it is robust to small perturbations in its inputs.

Density matrices have been used in other areas of distributional semantics such as by [5, 10] to encode ambiguity. This work is partially based on the first author's MSc thesis [2].

2 Categorical Compositional Distributional Meaning

The grammatical structure of language can be modelled by Lambek's pregroup grammars [6]. In distributional semantics, word meanings are described by finite real vectors, derived from text corpora using word co-occurrence statistics [8, 9, 3]. Both pregroups and the category **FHilb**, of finite dimensional real Hilbert spaces, are monoidal categories with duals. This is the key insight exploited by categorical composition distributional models of natural language [4]. By functorially mapping grammatical reductions into **FHilb**, language meaning can be derived from grammatical structure.

Our aim is to provide a satisfactory account of entailment in categorical composition distributional semantics. There is no meaningful ordering on real vectors, so we must adjust our model. We exploit Selinger's CPM construction [11] to construct a new model **CPM**(**FHilb**) in which meanings are now represented by density matrices rather than simple vectors. We use this extra flexibility to capture the concept of hyponymy, where one word may be seen as an instance of another. For example, *red* is a hyponym of *colour*. The hyponymy relation can be associated with a notion of logical entailment. Some

Submitted to:
QPL 2016

entailment is crisp, for example: *dog* entails *animal*. However, we may also wish to permit entailments of differing strengths. For example, the concept *dog* gives high support to the the concept *pet*, but does not completely entail it: some dogs are working dogs. The hyponymy relation we describe here can account for these phenomena. We should also be able to measure entailment strengths at the sentence level. For example, we require that *Cujo is a dog* crisply entails *Cujo is an animal*, but that the statement *Cujo is a dog* does not completely entail *Cujo is a pet*. Again, the relation we describe here will successfully account for this behaviour at the sentence level.

3 Predicates and Entailment

We view positive operators as predicates and use a generalization of the Löwner order [7] to give a notion of entailment.

Definition 1 (k-hyponym)**.** *We say that A is a k-*hyponym *of B for a given value of k in the range $(0,1]$ and write $A \preceq_k B$ if:*

$$0 \sqsubseteq B - kA$$

In general, we are interested in the maximal value k_{max} for which k-hyponymy holds This is given by:

Theorem 1. *For positive self-adjoint matrices A, B such that $supp(A) \subseteq supp(B)$ the maximum k such that $B - kA \geq 0$ is given by $1/\lambda$ where λ is the maximum eigenvalue of B^+A, A^+ is the Moore-Penrose pseudo-inverse and $supp(A)$ is the the support of A.*

The notion of k-hyponymy is continuous. For $A \preceq_k B$, a small perturbation to A gives a small change in k.

Theorem 2. *Given $A \preceq_k B$ and density operator ρ such that $supp(\rho) \subseteq supp(B)$, then for any $\varepsilon > 0$ we can choose a $\delta > 0$ such that: $A' = A + \delta\rho \implies A' \preceq_{k'} B$ and $|k - k'| < \varepsilon$.*

4 Compositionality

Crucially, k-hyponymy 'lifts' to the sentence level in a very intuitive way, and we can infer entailment at the sentence level from entailment at the word level, in a graded manner.

Theorem 3 (Generalised Sentence k-Hyponymy)**.** *Let Φ and Ψ be two positive phrases of the same length and grammatical structure, expressed in the same noun spaces N and sentence spaces S. Denote the nouns and verbs of Φ, by $A_1, \ldots, A_n$ and in Ψ by $B_1 \ldots B_n$, with density matrices $[\![A_1]\!], \ldots, [\![A_n]\!]$ and $[\![B_1]\!], \ldots, [\![B_n]\!]$ respectively. Suppose that $[\![A_i]\!] \preceq_{k_i} [\![B_i]\!]$ for $i \in \{1, \ldots, n\}$ and some $k_i \in (0,1]$, and that φ is the linear map induced by the grammatical structure of Φ and Ψ. Then:*

$$\varphi(\Phi) \preceq_{k_1 \cdots k_n} \varphi(\Psi)$$

so $k_1 \cdots k_n$ provides a lower bound on the extent to which $\varphi(\Phi)$ entails $\varphi(\Psi)$

We consider a concrete example. Suppose we have a noun space N with basis $\{|e_i\rangle\}_i$, and sentence space S with basis $\{|x_j\rangle\}_j$ We consider the verbs *nibble*, *scoff* and the nouns *cake*, *chocolate* given by pure states. The more general *eat* and *sweets* are given by:

$$[\![eat]\!] = \frac{1}{2}([\![nibble]\!] + [\![scoff]\!]) \quad \text{and} \quad [\![sweets]\!] = \frac{1}{2}([\![cake]\!] + [\![chocolate]\!])$$

Then $[\![scoff]\!] \preccurlyeq_{1/2} [\![eat]\!]$ and $[\![cake]\!] \preccurlyeq_{1/2} [\![sweets]\!]$. We consider the sentences $s_1 = $ *Mary scoffs cake* and $s_2 = $ *Mary eats sweets*. The semantics of these sentences are:

$$[\![s_1]\!] = \varphi([\![Mary]\!] \otimes [\![scoffs]\!] \otimes [\![cake]\!]) \quad \text{and} \quad [\![s_2]\!] = \varphi([\![Mary]\!] \otimes [\![eats]\!] \otimes [\![sweets]\!])$$

We will show that $[\![s_1]\!] \preccurlyeq_{kl} [\![s_2]\!]$ where $kl = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Expanding $[\![s_2]\!]$ and subtracting $\frac{1}{4}[\![s_1]\!]$ we obtain:

$$\begin{aligned}[\![s_2]\!] - \frac{1}{4}[\![s_1]\!] = \varphi([\![Mary]\!] \otimes [\![choc]\!] \otimes [\![choc]\!]) + \varphi([\![Mary]\!] \otimes [\![nibbles]\!] \otimes [\![cake]\!]) \\ + \varphi([\![Mary]\!] \otimes [\![nibbles]\!] \otimes [\![choc]\!]))\end{aligned}$$

We can see that $[\![s_2]\!] - \frac{1}{4}[\![s_1]\!]$ is positive by positivity of the individual elements and the fact that positivity is preserved under addition and tensor product. Therefore, $[\![s_1]\!] \preccurlyeq_{kl} [\![s_2]\!]$, as required.

5 Conclusion

Integrating a logical framework with compositional distributional semantics is an important step in improving this model of language. By moving to the setting of density matrices, we have described a graded measure of entailment that may be used to describe the extent of entailment between two words represented within this enriched framework. This approach extends uniformly to provide entailment strengths between phrases of any type. We have also shown how a lower bound on entailment strength of phrases of the same structure can be calculated from their components.

References

[1] Esma Balkır, Dimitri Kartsaklis & Mehrnoosh Sadrzadeh (2015): *Sentence Entailment in Compositional Distributional Semantics*. To appear ISAIM 2016.

[2] Dea Bankova (2015): *Comparing Meaning in Language and Cognition - P-Hypononymy, Concept Combination, Asymmetric Similarity*. Master's thesis, University of Oxford.

[3] John A Bullinaria & Joseph P Levy (2007): *Extracting semantic representations from word co-occurrence statistics: A computational study*. Behavior research methods 39(3), pp. 510–526.

[4] Bob Coecke, Mehrnoosh Sadrzadeh & Stephen Clark (2010): *Mathematical foundations for a compositional distributional model of meaning*. arXiv:1003.4394.

[5] Dimitri Kartsaklis (2015): *Compositional Distributional Semantics with Compact Closed Categories and Frobenius Algebras*. Ph.D. thesis, University of Oxford.

[6] Joachim Lambek (1999): *Type grammar revisited*. In: Logical aspects of computational linguistics, Springer, pp. 1–27.

[7] Karl Löwner (1934): *Über monotone matrixfunktionen*. Mathematische Zeitschrift 38(1), pp. 177–216.

[8] Kevin Lund & Curt Burgess (1996): *Producing high-dimensional semantic spaces from lexical co-occurrence*. Behavior Research Methods, Instruments, & Computers 28(2), pp. 203–208.

[9] Jeff Mitchell & Mirella Lapata (2010): *Composition in distributional models of semantics*. Cognitive science 34(8), pp. 1388–1429.

[10] Robin Piedeleu, Dimitri Kartsaklis, Bob Coecke & Mehrnoosh Sadrzadeh (2015): *Open System Categorical Quantum Semantics in Natural Language Processing*. arXiv:1502.00831.

[11] Peter Selinger (2007): *Dagger compact closed categories and completely positive maps*. Electronic Notes in Theoretical Computer Science 170, pp. 139–163.