Modelling Ambiguity with Density Matrices in Compositional Distributional Models of Meaning

Robin Piedeleu[1] Dimitri Kartsaklis[2] Bob Coecke[1] Mehrnoosh Sadrzadeh[2]

[1]University of Oxford, Department of Computer Science

{robin.piedeleu;bob.coecke}@cs.ox.ac.uk

[2]Queen Mary University of London, School of Electronic Engineering and Computer Science

{d.kartsaklis;m.sadrzadeh}@qmul.ac.uk

Abstract

This abstract summarizes the categorical compositional distributional model of Piedeleu, Kartsaklis, Coecke and Sadrzadeh (Piedeleu et al., 2015) which is capable of explicitly dealing with lexical ambiguity and its various levels. Ambiguous words are represented as mixed states, that is, as probability distributions over all their available meanings.

1 Introduction

Originally inspired by categorical quantum mechanics, the categorical compositional distributional model of natural language meaning of Coecke, Sadrzadeh and Clark (Coecke et al., 2010) provides a conceptually motivated procedure to compute the meaning of a sentence, given its grammatical structure within a Lambek pregroup and a vectorial representation of the meaning of its parts. The predictions of this first model have outperformed that of other models in mainstream empirical language processing tasks on large scale data. Moreover, just like CQM allows for varying the model in which we interpret quantum axioms, one can also vary the model in which we interpret word meaning.

Recent work by Piedeleu et al. (2015) shows that further developments in categorical quantum mechanics are relevant to natural language processing too. Firstly, Selinger's CPM-construction allows for explicitly taking into account lexical ambiguity and distinguishing between the two inherently different notions of homonymy and polysemy. In terms of the model in which we interpret word meaning, this means a passage from the vector space model to density matrices. Despite this change of model, standard empirical methods for comparing meanings can be easily adopted, which the above paper demonstrates by a small-scale experiment on real-world data. Secondly, commutative classical structures as well as their non-commutative counterparts that arise in the image of the CPM-construction allow for encoding relative pronouns, verbs and adjectives, and finally, iteration of the CPM-construction, something that has no counterpart in the quantum realm, enables one to accommodate both entailment and ambiguity. In the following sections we summarize the main ideas of this work.

2 Composition and ambiguity

In a distributional model of meaning where a homonymous word is represented by a single vector, the ambiguity in meaning has been collapsed into a convex combination of the relevant sense vectors; the result is a vector that can be seen as the average of all senses, inadequate to reflect the meaning of any of them in a reliable way (Kartsaklis and Sadrzadeh, 2013). We need a way to avoid that. In natural language, ambiguities are resolved with the introduction of context, which means that for a compositional model of meaning the resolving mechanism is the compositional process itself. We would like to retain the ambiguity of a homonymous word when needed (i.e. in the absence of appropriate context) and allow it to collapse only when the context defines the intended sense, during the compositional process.

In summary, we seek an appropriate model that will allows us: (a) to express homonymous words as probabilistic mixings of their individual meanings; (b) to retain the ambiguity until the presence of sufficient context that will eventually resolve it during composition time; (c) to achieve all the above in the multi-linear setting imposed by the vector space semantics of our original model.

3 Encoding ambiguity

We represent ambiguous words as *mixed states* expressed by *density operators*.[1]

Definition 3.1. *Let a distributional model be given in the form of a Hilbert space M, in which every word w_t is represented by a statistical ensemble $\{(p_i, |w_t^i\rangle)\}_i$—where $|w_t^i\rangle$ is a vector corresponding to a specific unambiguous meaning of the word that can occur with probability p_i. The distributional meaning of the word is defined as:*

$$\rho(w_t) = \sum_i p_i |w_t^i\rangle\langle w_t^i| \tag{1}$$

We recast the categorical model of Coecke et al. (2010) to an *open quantum system* setting by using the CPM construction (Selinger, 2007). Taking $\tilde{M}$ to be the canonical functor $\tilde{M} : \mathcal{C} \to \mathbf{CPM}(\mathcal{C})$ and $Q : \mathbf{C_F} \to \mathcal{C}$ a strongly monoidal functor from the free compact closed category generated over a pregroup grammar (Lambek, 2008) to a generic semantic category, the new model is defined by:

$$\tilde{M}Q : \mathbf{C_F} \to \mathcal{C} \to \mathbf{CPM}(\mathcal{C}) \tag{2}$$

Definition 3.2. *Let $\rho(w_i)$ be a state $I \to \tilde{M}Q(p_i)$ corresponding to word w_i with type p_i in a sentence $w_1 \dots w_n$. Given a type-reduction $\alpha : p_1 \cdot \dots \cdot p_n \to s$, the meaning of the sentence is defined as:*

$$\rho(w_1 \dots w_n) := \tilde{M}Q(\alpha)\big(\rho(w_1) \otimes_{\mathbf{CPM}} \dots \otimes_{\mathbf{CPM}} \rho(w_n)\big)$$

4 Measuring ambiguity

Von Neumann entropy can be used to measure how ambiguity evolves from individual words to text compounds; we would expect that ‘bank’ is highly ambiguous, but in ‘river bank’ the ambiguity is resolved. A small experiment with 5 nouns modified by adjectives and relative clauses was considered, in which density matrices for the compounds were formed by composition over the words and Frobenius operations; in all cases the entropy of the compound was indeed lower than that of the unmodified noun. For example, the entropy of ‘vessel’ was 0.25, but the entropy of ‘vessel that sails’ was almost zero (0.01); in other words, the density matrix of ‘vessel that sails’ expressed a *pure state*, that is, an unambiguous meaning. This demonstrates and important aspect of the model: *disambiguation = purification*. See (Piedeleu et al., 2015) for the complete experiment.

References

Blacoe, W., E. Kashefi, and M. Lapata (2013, June). A quantum-theoretic approach to distributional semantics. In *Proceedings of NACL 2013*, pp. 847–857. Association for Computational Linguistics.

Coecke, B., M. Sadrzadeh, and S. Clark (2010). Mathematical Foundations for a Compositional Distributional Model of Meaning. Lambek Festschrift. *Linguistic Analysis 36*, 345–384.

Kartsaklis, D. and M. Sadrzadeh (2013, October). Prior disambiguation of word tensors for constructing sentence vectors. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Washington, USA, pp. 1590–1601. Association for Computational Linguistics.

Lambek, J. (2008). *From Word to Sentence*. Milan: Polimetrica.

Piedeleu, R., D. Kartsaklis, B. Coecke, and M. Sadrzadeh (2015). Open System Categorical Quantum Semantics in Natural Language Processing. *arXiv preprint arXiv:1502.00831*.

Selinger, P. (2007). Dagger compact closed categories and completely positive maps. *Electronic Notes in Theoretical Computer Science 170*, 139–163.

[1] See also (Blacoe et al., 2013), which presents a distributional (but *not compositional*) model based on a different form of density matrices created from grammatical dependencies.