Compositional Distributional Models of Meaning

Mehrnoosh Sadrzadeh **Dimitri Kartsaklis**

Queen Mary University of London
School of Electronic Engineering and Computer Science
Mile End Road, London E1 4NS, UK
{mehrnoosh.sadrzadeh;d.kartsaklis}@qmul.ac.uk

1 Description

Distributional models of meaning (see Turney and Pantel (2010) for an overview) are based on the pragmatic hypothesis that meanings of words are deducible from the contexts in which they are often used. This hypothesis is formalized using vector spaces, wherein a word is represented as a vector of co-occurrence statistics with a set of context dimensions. With the increasing availability of large corpora of text, these models constitute a well-established NLP technique for evaluating semantic similarities. Their methods however do not scale up to larger text constituents (i.e. phrases and sentences), since the uniqueness of multi-word expressions would inevitably lead to data sparsity problems, hence to unreliable vectorial representations. The problem is usually addressed by the provision of a compositional function, the purpose of which is to prepare a vector for a phrase or sentence by combining the vectors of the words therein. This line of research has led to the field of compositional distributional models of meaning (CDMs), where reliable semantic representations are provided for phrases, sentences, and discourse units such as dialogue utterances and even paragraphs or documents. As a result, these models have found applications in various NLP tasks, for example paraphrase detection; sentiment analysis; dialogue act tagging; machine translation; textual entailment; and so on, in many cases presenting state-of-the-art performance.

Being the natural evolution of the traditional and well-studied distributional models at the word level, CDMs are steadily evolving to a popular and active area of NLP. The topic has inspired a number of workshops and tutorials in top CL conferences such as ACL and EMNLP, special issues at high-profile journals, and it attracts a substantial amount of submissions in annual NLP conferences. The approaches employed by CDMs are as much as diverse as statistical machine leaning (Baroni and Zamparelli, 2010), linear algebra (Mitchell and Lapata, 2010), simple category theory (Coecke et al., 2010), or complex deep learning architectures based on neural networks and borrowing ideas from image processing (Socher et al., 2012; Kalchbrenner et al., 2014; Cheng and Kartsaklis, 2015). Furthermore, they create opportunities for interesting novel research, related for example to efficient methods for creating tensors for relational words such as verbs and adjectives (Grefenstette and Sadrzadeh, 2011), the treatment of logical and functional words in a distributional setting (Sadrzadeh et al., 2013; Sadrzadeh et al., 2014), or the role of polysemy and the way it affects composition (Kartsaklis and Sadrzadeh, 2013; Cheng and Kartsaklis, 2015). The purpose of this tutorial is to provide a concise introduction to this emerging field, presenting the different classes of CDMs and the various issues related to them in sufficient detail. The goal is to allow the student to understand the general philosophy of each approach, as well as its advantages and limitations with regard to the other alternatives.

2 Some background on CDMs

The purpose of a compositional distributional model is to provide a function that produces a vectorial representation of the meaning of a phrase or a sentence from the distributional vectors of the words therein. One can broadly classify such compositional distributional models to three categories:

Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Tutorial Abstracts, pages 1–4, Osaka, Japan, December 11-17 2016.

- **Vector mixture models:** These are based on simple element-wise operations between vectors, such as addition and multiplication (Mitchell and Lapata, 2008). Vector mixture models constitute the simplest compositional method in distributional semantics. Despite their simplicity, though, have been proved a very hard-to-beat baseline for many of the more sophisticated models.

- **Tensor-based models:** In these models, relational words such as verbs and adjectives are tensors and matrices contracting and multiplying with noun (and noun-phrase) vectors (Coecke et al., 2010; Baroni and Zamparelli, 2010). Tensor-based models provide a solution to the problems of vector mixtures: they are not bag-of-words approaches and they respect the type-logical identities of special words, following an approach very much aligned with the formal semantics perspective. In fact, tensor-based composition is considered as the most linguistically motivated approach in compositional distributional semantics.

- **Neural-netword based models:** Models in which the compositional operator is part of a neural network and is usually optimized against a specific objective (Socher et al., 2012; Kalchbrenner et al., 2014; Cheng and Kartsaklis, 2015). Architectures that are usually employed are that of recursive or recurrent neural networks and convolutional neural networks. The non-linearity in combination with the layered approach in which neural networks are based make these models quite powerful, allowing them to simulate the behaviour of a range of functions much wider than the linear maps of tensor-based approaches.

Figure 1 provides an overview and a taxonomy of CDMs based on their theoretical power.

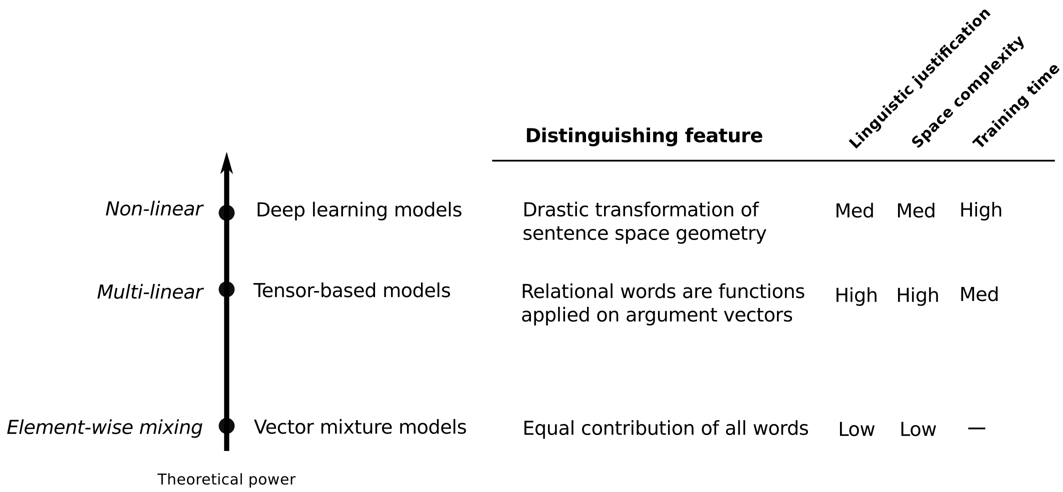

Figure 1: A hierarchy of compositional distributional models based on their theoretical power (Kartsaklis, 2015).

3 Outline

The tutorial aims at providing an introduction to these three classes of models, covering the most important aspects. Specifically, it will have the following structure (subject to time limitations):

- **Introduction.** The distributional hypothesis – Vector space models – The necessity for compositionality – Applications – An overview of CDMs

- **Vector mixture models.** Additive and multiplicative models – Interpretation – Practical applications

- **Tensor-based models.** Unifying grammar and semantics – Relational words as multi-linear maps – Extensions of the model

- **Machine learning/Deep learning models.** Introduction to NNs – Recursive and Recurrent NNs for composition – Connection to image processing – Convolutional NNs

- **Advanced issues and conclusion.** Logical and functional words – Lexical ambiguity and composition – Moving to discourse level – Concluding remarks

4 Prerequisites

The only prerequisite for attending the tutorial will be a knowledge of standard linear algebra, specifically with regard to vectors and their operations, vector spaces, matrices and linear maps. No special knowledge on advanced topics, such as category theory or neural networks, will be necessary.

5 Instructors

Mehrnoosh Sadrzadeh has taught undergraduate/graduate courses on logics of information from 2009-2012 in Oxford and in the Sino-European Winter School in Logic Language Computation in 2010 in China. She has experience in tutoring advance courses in logic for Oxford colleges. Since 2013, she teaches a course on computability in Queen Mary University London. She has long term experience in organising departmental and group research seminars in Oxford and Queen Mary and has experience in organising interdisciplinary workshops and conferences including the 11th International Conference on Computational Semantics (IWCS) and its satellite workshop Advances in Distributional Semantics in 2014. Her papers (Coecke et al., 2010; Grefenstette and Sadrzadeh, 2011) on compositional distributional semantics are among the highly cited papers of the field.

Dimitri Kartsaklis has been tutoring in various NLP and CS courses at the University of Oxford from 2012 to 2015. Recently he organized the 2016 Workshop on Semantic Spaces at the Intersection of NLP, Physics and Cognitive Science (with Martha Lewis and Laura Rimell). His work, focused on both theoretical and experimental aspects of compositional distributional models of meaning (for example, (Kartsaklis et al., 2012; Cheng and Kartsaklis, 2015)) has been published in various top-tier NLP and Computer Science conferences and journals.

6 Recommended reading

- Baroni and Zamparelli (2010). *Nouns are Vectors, Adjectives are Matrices: Representing adjective-noun constructions in semantic space*

- Coecke et al. (2010). *Mathematical Foundations for a Compositional Distributional Model of Meaning*

- Kartsaklis et al. (2012) *A Unified Sentence for Categorical Compositional Distributional Semantics: Theory and Experiments*

- Mitchell and Lapata (2010). *Composition in Distributional Models of Semantics*

- Socher et al. (2012). *Semantic Compositionality through Recursive Matrix-Vector Spaces*

- Turney and Pantel (2010). *From frequency to meaning: Vector space models of semantics*

References

M. Baroni and R. Zamparelli. 2010. Nouns are Vectors, Adjectives are Matrices. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Jianpeng Cheng and Dimitri Kartsaklis. 2015. Syntax-aware multi-sense word embeddings for deep compositional models of meaning. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1531–1542, Lisbon, Portugal, September. Association for Computational Linguistics.

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

Edward Grefenstette and Mehrnoosh Sadrzadeh. 2011. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404, Edinburgh, Scotland, UK., July. Association for Computational Linguistics.

Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 655–665, Baltimore, Maryland, June. Association for Computational Linguistics.

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

Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Stephen Pulman. 2012. A unified sentence space for categorical distributional-compositional semantics: Theory and experiments. In *Proceedings of 24th International Conference on Computational Linguistics (COLING 2012): Posters*, pages 549–558, Mumbai, India, December. The COLING 2012 Organizing Committee.

Dimitri Kartsaklis. 2015. *Compositional Distributional Semantics with Compact Closed Categories and Frobenius Algebras*. Ph.D. thesis, University of Oxford.

J. Mitchell and M. Lapata. 2008. Vector-based Models of Semantic Composition. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pages 236–244.

Jeff Mitchell and Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388–1439.

M. Sadrzadeh, S. Clark, and B. Coecke. 2013. The Frobenius anatomy of word meanings I: subject and object relative pronouns. *Journal of Logic and Computation*, Advance Access, October.

M. Sadrzadeh, S. Clark, and B. Coecke. 2014. The Frobenius anatomy of word meanings II: Possessive relative pronouns. *Journal of Logic and Computation*, June.

R. Socher, B. Huval, C. Manning, and Ng. A. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Conference on Empirical Methods in Natural Language Processing 2012*.

Peter D Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1):141–188.