DEEPCOMPOSE: GROUNDING COMPOSITIONAL VECTOR SPACE SEMANTIC MODELS TO MDPS

Benjamin A. Spiegel
Department of Computer Science
Brown University
Providence, RI 02912, USA
bspiegel@cs.brown.edu

ABSTRACT

Vector space semantic models (VSMs) are the dominant form of semantic representation for natural language in modern NLP techniques. However, the prevailing VSMs used in large language models (LLMs) lack many properties commonly present in classical linguistic representations of natural language semantics, namely direct compositionality and model-theoretic grounding. For my Research Comprehensive, I will introduce Deep Semantic Composition (DeepCompose) a novel class of VSMs for semantics that are both directly compositional and grounded to Markov Decision Processes (MDPs). I will run experiments in multiple domains including the visual Taxi domain, which uses RGB images to represent state. This approach offers a promising solution for unifying symbolic and connectionist approaches to model-theoretic natural language semantics.

1 INTRODUCTION

Vector space semantic models (VSMs) – a class of models that represent the semantics of natural language as vectors and operations on vectors (Turney & Pantel, 2010) – are the dominant form of semantic representation for natural language in modern NLP techniques. However, the prevailing VSMs used in large language models (LLMs) lack many properties commonly present in classical linguistic representations of natural language semantics. For one, LLM-based VSMs are learned by "modeling" language, i.e. predicting which token will appear next in a partial sentence drawn from a large corpus. This contradicts a long-standing notion in linguistics that language has meaning beyond its surface form, i.e. is comprised of *signifiers* that refer to *signifieds* that exist in a separate *content plane* (Saussure, 1916). Since the VSMs learned by LLMs arise solely from reading text, they have restricted access to the meanings that language refers to, limiting their semantic understanding.

Secondly, the precise semantics for any given word, phrase, or linguistic operation are opaque in LLM-based VSMs. In LLMs, static high-dimensional vectors corresponding to natural language tokens are fed into the model as input, which are manipulated throughout by black-box vector operations in order to predict the next token. The murkiness of these semantic representations make it more difficult to know *a priori* if they adhere to any of the many properties of natural language that have been discovered by linguists.[1] One long-standing hypothesis about natural language introduced by Leibniz and further developed by Frege is that that the meaning of a phrase is composed of the meaning of its parts (Frege et al., 1879), though it is unclear whether LLM-based VSMs have this property. Furthermore, it's even less clear as to whether these models are in accordance with the stronger notion of model-theoretic direct compositionality introduced by Montague, that the meaning of a phrase is arrived at by directly composing the semantic meanings of its parts into a lambda calculus that is grounded in a formal world model (Montague, 1970).

Despite the criticisms surrounding classical approaches to natural language semantics, there is an undeniable beauty to the direct compositional approach set forth my Montague and continued by many linguists. Subsequent works have rendered precise machinery for describing the semantics of

[1]The field of mechanistic interpretability is concerned with deciphering these representations.

Part of speech	Syntactic category	Example usage	Semantic type	Example log. form
Noun	N	**person** : N	$\mathbb{R}^d$	v_{person}
Adjective	N/N_x	**good** person : N	$\langle \mathbb{R}^d, \mathbb{R}^d \rangle$	$\lambda x.A_{good}x$
Determiner	NP/N_x	**the** person : NP	$\langle \mathbb{R}^d, \mathbb{R}^d \rangle$	$\lambda x.x$
Intrans. Verb	$S\backslash NP_x$	the person **ran** : S	$\langle \mathbb{R}^d, \mathbb{R}^d \rangle$	$\lambda x.A_{ran}x + b_{ran}$
Trans. Verb	$S\backslash NP_y/NP_x$	the person **ran** home : S	$\langle \mathbb{R}^d, \langle \mathbb{R}^d, \mathbb{R}^d \rangle \rangle$	$\lambda x.\lambda y.(\mathcal{T}_{ran}x)y$
Adverb	$(S\backslash NP)\backslash(S\backslash NP)$	ran **lazily** : $S\backslash NP$	$\langle \langle \mathbb{R}^d, \mathbb{R}^d \rangle, \langle \mathbb{R}^d, \mathbb{R}^d \rangle \rangle$	$[\lambda y.Ay \rightarrow \lambda y.(\mathcal{T}_{lazy}A)y]$
	$(S\backslash NP)/(S\backslash NP)$	**lazily** ran : $S\backslash NP$	$\langle \langle \mathbb{R}^d, \mathbb{R}^d \rangle, \langle \mathbb{R}^d, \mathbb{R}^d \rangle \rangle$	$[\lambda y.Ay \rightarrow \lambda y.(\mathcal{T}_{lazy}A)y]$
	$(N/N)/(N/N)$	**very** good person : N	$\langle \langle \mathbb{R}^d, \mathbb{R}^d \rangle, \langle \mathbb{R}^d, \mathbb{R}^d \rangle \rangle$	$[\lambda y.Ay \rightarrow \lambda y.(\mathcal{T}_{very}A)y]$
Preposition	$(N\backslash N_y)/N_x$	person **in** France : N	$\langle \mathbb{R}^d, \langle \mathbb{R}^d, \mathbb{R}^d \rangle \rangle$	$\lambda x.\lambda y.(\mathcal{T}_{in}x)y$
	$(S\backslash NP_y)\backslash(S\backslash NP)_f/NP_x$	ran **in** France : $S\backslash NP$	$\langle \mathbb{R}^d, \langle \langle \mathbb{R}^d, \mathbb{R}^d \rangle, \langle \mathbb{R}^d, \mathbb{R}^d \rangle \rangle \rangle$	$\lambda x.\lambda f.\lambda y.(\mathcal{T}_{in}x)(f(y))$

Table 1: Common syntactic categories in CCG, paired with their semantic types and example logical forms according to VSSP. DeepCompose uses a nearly identical representation for the extensional semantics of a lexicon, but differs from VSSP in how their specific weights are learned and generated.

large fragments of natural language, most commonly of English. Such approaches have accounted for complex linguistic phenomenon, such as generalized quantifiers (Barwise & Cooper, 1981), negative polarity items (Cresswell, 1976; Klein, 1980), and the licensing of anaphora (Reinhart, 1976). I argue that the successes of these approaches merit further inquiry into building VSMs for natural language semantics that are both directly compositional and model-theoretic.

2 Background

One line of research that predates the LLM explosion is concerned precisely with directly compositional VSMs (Mitchell & Lapata, 2010; Socher et al., 2012; 2013). Vector space semantic parsing (VSSP) (Krishnamurthy & Mitchell, 2013), parses a sentence into lambda calculus and assigns each token a vector or function over vectors based on its combinatory categorial grammar (CCG) type (see 1) (Steedman, 1996). Evaluating the resulting expression yields a vector, which acts as a point of supervision for training the weights of the semantics for the constituents. Formally, VSSP semantics are trained using a dataset of tuples, $\{(l^i, y^i)\}_{i=1}^n$, where l is a logical form and y is a label vector. Each logical form l is a function from lexicon parameters θ to vectors in $\mathbb{R}^d$, and the parameters are learned by maximizing the objective:

$$O(\theta) = \sum_{i=1}^{n} \mathcal{L}(y^i, l^i(\theta)) + \frac{\lambda}{2}||\theta||^2$$

Where $\mathcal{L}$ is some loss function defined over pairs of label vectors and λ is a regularization parameter.

In experiments, VSSP vastly outperformed prior non-compositional VSMs on a propositional logic task for predicting the truth-values of simple `and/or/xor/true/false` expressions as well as an adjective-adverb-noun composition task. In contrast to LLM-based VSMs, VSSP semantics are grounded in meanings outside of the language domain (e.g. truth values in the case of propositional logic). However, its semantics are *extensional*, they are fixed and apply only to a single timeless domain. Language is a general and useful tool because humans can interpret it across worlds and times; its semantics are *intensional*, i.e. are functions from world-time pairs to semantic extensions. There has been much successful work formalizing and integrating this notion of *possible worlds* into classical semantics (Von Fintel & Heim, 2011), but there does not seem to be work in NLP inspired by this approach.

There are, however, notions of formal models of the world elsewhere in the field of Artificial Intelligence: in Reinforcement Learning and Planning. In these fields it is common to model the world as a Markov Decision Process (MDP) Puterman (1994), a class of stochastic sequential decision processes whose dynamics are totally determined by the current state and action. Formally, an MDP is a tuple (S, A, T, R, γ), where S is a set of states, A is a set of actions the agent can take, $T : S \times A \rightarrow S$ is a transition function that determines the next state for a given state-action

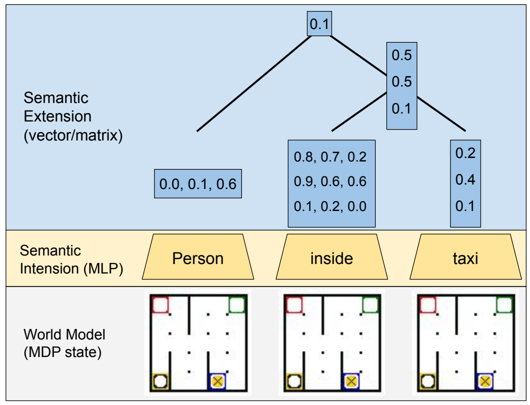

Figure 1: A simplified example of DeepCompose that uses matrix multiplication to compose semantic extensions. In experiments, vectors and matrices are significantly larger than the number of elements illustrated, and ReLU activation functions are applied after each matrix operation.

pair, $R : S \times A \rightarrow \mathbb{N}$ is a reward function, and γ is a discount factor. This simple MDP formalism has been augmented with additional machinery to represent environments in which the state is only partially observed (Kaelbling et al., 1998), in which some actions are abstract and temporally extended (Sutton et al., 1999), and in which there is additional structure in the state or action space (Diuk et al., 2008; Koller & Parr, 2013; Masson et al., 2016). Furthermore, recent work has highlighted remarkable parallels between the structure of natural language and various structured MDP formalisms (Patel et al., 2020), which motivates their use as a target for grounding natural language.

3 Proposed Research: Deep Semantic Composition

For my Research Comprehensive, I will introduce Deep Semantic Composition (DeepCompose) a novel class of VSMs for natural language that are both directly compositional and model-theoretic. Similarly to VSSP, DeepCompose represents the extensional semantics for natural language tokens as vectors and functions over vectors according to their CCG type, and directly composes them together to arrive at the meanings of whole declarative sentences, which are represented as scalars denoting truth value between 0 and 1. To generate the semantic extensions for a given token, each token is assigned a unique *semantic intension module* $\phi : S \rightarrow \mathcal{E}$, a function from MDP states $s \in S$ to semantic extensions $e \in \mathcal{E}$. As we would hope, the semantic extensions for the lexicon are dependent on the state of the MDP, as the truth value for a given proposition is a function of the current state of the world.

While in VSSP, the semantics for a lexicon are learned from a dataset of sentences and label vectors, DeepCompose learns from a dataset $\{(l^i, s^i, y^i)\}_{i=1}^n$ of propositions, MDP states, and truth values. Propositions l^i, which are in logical form, select out from a lexicon of semantic intensions Φ a subset of modules $\{\phi_1, \phi_2, ...\phi_t\}$ corresponding to the tokens in the proposition and knits them together[2] into a function from MDP states to truth values. The parameters for the intensions in Φ are learned by maximizing the objective:

$$O(\Phi) = \sum_{i=1}^{n} \mathcal{L}(y^i, l^i(\Phi)(s^i)) + \frac{\lambda}{2}||\Phi||^2$$

[2]My representation of choice for extensions are vectors and matrices, though more sophisticated function classes could be used in theory. Extension vectors/matrices are composed together in the order given by the logical form, and after each step a sigmoid function is applied. See 1.

3.1 Proposed Experiments

Propositional logic on low-dimensional states I will run a variation of the propositional logic experiment run in VSSP that includes a simple low-dimensional state. Instead of evaluating `and/or/xor/true/false` expressions, I will evaluate `and/or/xor/a/b` expressions, where `a` and `b` are variables bound to values in a primitive state vector `[a, b]`.

Simplified grammar on high-dimensional states To assess the viability of DeepCompose representations, I propose to run experiments in the visual Taxi domain (pictured in 1), which represents the state of the world as an RGB image. Using a DeepCompose representation, I will attempt to learn the object-oriented abstraction provided by Diuk et al. (2008), i.e. for every state learn the set of true and false object-oriented propositions.

Re-composability Perhaps the most essential property of compositional semantics is the out-of-the-box re-composability of lexical items to form new meaningful sentences. To assess the compositionality of DeepCompose representations I will generate a more sophisticated grammar for an English fragment that applies to the Taxi domain, and include a subset of all possible sentences in the training data. I will then assess the accuracy of DeepCompose on sentences outside of the training data.

First order logic Given a more complex domain, can DeepCompose learn representations for sets and quantifiers?

4 Conclusion

In this proposal I presented a preliminary sketch of Deep Semantic Composition (DeepCompose), a novel VSM class that represents tokens as *semantic intension modules*, which are functions from states in a Markov Decision Process to semantic extensions. Similar to previous works, these extensions are represented by vectors and matrices that compose together to form truth values for the underlying proposition. The resulting representation is a natural extension of VSMs to the domain of intensional semantics, as it is both directly compositional and grounded to a modern world formalism.

References

Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2):159–219, 1981. ISSN 01650157, 15730549. URL `http://www.jstor.org/stable/25001052`.

Max J Cresswell. The semantics of degree. In *Montague grammar*, pp. 261–292. Elsevier, 1976.

Carlos Diuk, Andre Cohen, and Michael L Littman. An object-oriented representation for efficient reinforcement learning. In *Proceedings of the 25th international conference on Machine learning*, pp. 240–247, 2008.

Gottlob Frege et al. Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. *From Frege to Gödel: A source book in mathematical logic*, 1931:1–82, 1879.

Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.

Ewan Klein. A semantics for positive and comparative adjectives. *Linguistics and philosophy*, 4:1–45, 1980.

Daphne Koller and Ron Parr. Policy iteration for factored mdps. *arXiv preprint arXiv:1301.3869*, 2013.

Jayant Krishnamurthy and Tom Mitchell. Vector space semantic parsing: A framework for compositional vector space models. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, pp. 1–10, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL `https://aclanthology.org/W13-3201`.

Warwick Masson, Pravesh Ranchod, and George Konidaris. Reinforcement learning with parameterized actions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

Jeff Mitchell and Mirella Lapata. Composition in distributional models of semantics. *Cognitive science*, 34 8:1388–429, 2010.

Richard Montague. Universal grammar. *Theoria*, 36(3), 1970.

Roma Patel, Rafael Rodriguez-Sanchez, and George Konidaris. On the relationship between structure in natural language and models of sequential decision processes. In *Language in Reinforcement Learning Workshop at ICML 2020*, 2020. URL `https://openreview.net/forum?id=-KDnP4X1-0_`.

Martin L Puterman. Markov decision processes: Discrete stochastic dynamic programming, 1994.

Tanya Miriam Reinhart. *The syntactic domain of anaphora*. PhD thesis, Massachusetts Institute of Technology, 1976.

Ferdinand de Saussure. *Cours de linguistique générale*. Payot, 1916.

Richard Socher, Brody Huval, Christopher D. Manning, and A. Ng. Semantic compositionality through recursive matrix-vector spaces. In *Conference on Empirical Methods in Natural Language Processing*, 2012.

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

Mark Steedman. *Surface structure and interpretation*, volume 30. MIT press Cambridge, MA, 1996.

Richard S Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211, 1999.

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

Kai Von Fintel and Irene Heim. Intensional semantics. *Unpublished lecture notes*, 2011.