

LEARNING AND ANALYZING VECTOR ENCODING OF SYMBOLIC REPRESENTATIONS

Roland Fernandez, Ash Çelikyılmaz

Microsoft Research AI
Redmond, WA 98052, USA
{rfernand, aslicel}@microsoft.com

Paul Smolensky

Microsoft Research AI & Johns Hopkins University
Redmond, WA 98052, USA & Baltimore, MD 21218, USA
psmo@microsoft.com, smolensky@jhu.edu

Rishabh Singh

Google Brain
Mountain View, CA, USA
rishabh.iit@gmail.com

ABSTRACT

We present a formal language with expressions denoting general symbol structures and queries which access information in those structures. A sequence-to-sequence network processing this language learns to encode symbol structures and query them. The learned representation (approximately) shares a simple linearity property with theoretical techniques for performing this task.

1 OVERVIEW: S-LANG, S-NET AND S-REP

Embedding words and sentences in vector spaces has brought many symbolic tasks (especially in NLP) within the scope of deep neural network (DNN) models (Hinton, 1988; Palangi et al., 2016; Pollack, 1990; Socher et al., 2010; Weston et al., 2015). In general, DNNs may be expected to benefit if they can incorporate some of the power of symbolic computation without compromising the power of deep learning. The problem of embedding general symbol structures in vector spaces, and performing symbolic computation with these vectors, has been addressed theoretically, but these methods can require very large embedding spaces — e.g., Tensor Product Representation, TPR (Lee et al., 2016; Smolensky, 1990; Smolensky & Legendre, 2006) — or major error-correction/clean-up processes — e.g., Holographic Reduced Representation, HRR (Crawford et al., 2016; Plate, 1993; 2002; 2003) (Sec. 4; see also Kanerva (2009); Touretzky (1990)). We show here (Sec. 3) that deep learning can itself discover satisfactory methods of embedding general symbol structures, methods that operate in relatively small vector spaces without need for clean-up processing.

We define a general formal language scheme in which expressions denote symbol structures. Such a language will be called an S-LANG (Sec. 2). Information within these structures is accessed by evaluating query expressions within the language. The model that learns to encode structure-denoting expressions and to evaluate queries over these structures (Zaremba & Sutskever, 2014) is a simple bidirectional encoder-decoder model that operates on symbols in the formal language one at a time (Cho et al., 2014). We call such a model an S-NET (Sec. 3), and call the vector embedding of an S-LANG learned by an S-NET an S-REP.

Our S-NET learns to evaluate S-LANG expressions with a high degree of accuracy (Sec. 3). Furthermore, upon analysis, the S-REP that S-NET learns turns out to exhibit a simple linearity property — the Superposition Principle — that is crucial to both theoretical models, TPR and HRR (Sec. 4).

2 THE TASK: EMBEDDING GENERAL SYMBOLIC STRUCTURES IN VECTOR SPACES AND ACCESSING THEIR CONTENTS

Symbol-structure-denoting expressions: Structural-role binding. In general, a symbol structure can be characterized as a set of symbols each bound to a role that it plays in the structure (Newell,

Table 1: Example input/output pairs of the S-NET model

Expression	Value	Type
as:U	as:U	binding
qf:N ? N	qf	unbind
qf:N ? X	\$	unbind (not found)
(ao:N & ax:F & wh:A ? F):K	ax:K	3-bind, unbind, rebind
((((sf:W & fr:V):N):R):R):Y ? Y ? R ? R ? N	sf:W & fr:V	4-nested, unbind

Table 2: Performance and hyperparameters of the trained S-NET model

Accu- racy	Test per- plexity	Train loss	Mini- batches	Batch size	Hidden dimension	Drop- out	Learn- ing rate	Optim- izer	Attention; Beam
96.16	1.02	0.187	54K	128	128	0.2	0.001	Adam	None

1980, 141). The method is applicable to any type of symbol structure, but we focus on binary trees here. The simple binary tree $\mathcal{T} = [a [b c]]$ consists of the symbols a, b, c respectively bound to the roles R_0, R_{01}, R_{11} , where R_{01} is the role of left-child ('0') of right-child ('1') of root, etc. Symbolic structural roles are typically recursive. The recursive character of binary tree roles can be seen by viewing R_{01} as the symbol R_0 bound to the role R_1 . In the simple formal language we develop here, S-LANG, the tree \mathcal{T} will be denoted by the expression $a:L & b:L:R & c:R:R$, where L, R respectively abbreviate the roles R_0, R_1 . The grammar of S-LANG is shown in Fig. 1.

```

<exp> ::= <nil> | <filler-id> | <binding> | <unbinding> | <superimpose> | ( <exp> )
<binding> ::= <exp> : <role-id>
<unbinding> ::= <exp> ? <role-id> [ ? <role-id> ... ]
<superimpose> ::= <exp> & <exp>
<filler-id> ::= <id that starts with lowercase letter>
<role-id> ::= <id that starts with uppercase letter>
<nil> ::= $

```

Figure 1: Grammar of S-LANG: expressions denoting symbol structures and access queries.

Query-denoting expressions: Structural-role unbinding. A minimal requirement for a vector embedding a symbol structure is that it be possible to extract, with vector computation, the (embedding of) the symbol that is bound to any specified role. In S-LANG (see Fig. 1), the query denoted $? S$ asks for the structure (possibly a single symbol), bound to the role denoted S; this is role-unbinding. Thus the expression $a:L & b:L:R & c:R:R ? L$ asks for the structure filling the role L in \mathcal{T} , i.e., the structure forming the left child of the root, which in \mathcal{T} happens to be a single symbol, a. Similarly, the expression $a:L & b:L:R & c:R:R ? L:R$ asks for the left child of the right child of the root, and so has the value b. An S-LANG query can return a structure rather than a single symbol. The expression $a:L & b:L:R & c:R:R ? R$ asks for the right child of the root of \mathcal{T} , which is the structure $b:L & c:R$, denoting \mathcal{T} 's right sub-tree, $[b c]$.

Expressions combining querying and structure-building. The general expression in S-LANG allows structure that is returned by queries to be used to build new structures. Table 1 provides examples of expressions correctly evaluated by S-NET.

3 THE S-NET MODEL AND EXPERIMENTAL RESULTS

S-NET is a standard bidirectional encoder-decoder network where the output of the bidirectional LSTM encoder is the S-REP embedding of the input S-LANG expression. The S-REP vector is then fed as input to an LSTM decoder. Some implementation details are given in Table 2, which also gives the results of training S-NET on randomly-selected input/output pairs.

4 ANALYSIS OF S-REP: THE SUPERPOSITION PRINCIPLE

The Superposition Principle in theoretical structure-embedding schemes. Theoretical solutions to performing the task defined in Sec. 2 were proposed in the previous generation of neural network

modeling. Two general solutions, TPR and HRR, were introduced in Sec. 1. The TPR embedding of a symbol structure \mathcal{S} with symbols $\{s_k\}$ respectively bound to roles $\{r_k\}$ is $\text{TPR}(\mathcal{S}) \equiv \sum_k s_k \otimes r_k$, where \otimes denotes the tensor (generalized outer) product and $\{s_k\}$ and $\{r_k\}$ are embeddings of the symbols and roles, with respective dimensions σ and ρ ; the dimension of the TPR itself is then $\sigma\rho$.

If the role-embedding vectors $\{r_k\}$ are linearly independent, when collected together they form an invertible matrix \mathbf{R} ; the rows of \mathbf{R}^{-1} are the “unbinding” vectors $\{u_k\}$: $r_k \cdot u_j = \delta_{kj}$ so these vectors can be used to unbind the roles in a TPR. The symbol that fills role r_k in structure \mathcal{S} is exactly the symbol s_k with embedding $s_k = \text{TPR}(\mathcal{S}) \cdot u_k$.

A crucial property of TPR is that the embedding of a structure is the *sum* over embeddings of its symbols. This is TPR’s Superposition Principle. This is what enables extraction of symbols from any binding: $(\sum_k s_k \otimes r_k) \cdot u_j = \sum_k s_k \otimes (r_k \cdot u_j) = s_j$ since $r_k \cdot u_j = \delta_{kj}$.

HRRs are essentially contracted TPRs (Smolensky & Legendre, 2006, 260). The equation defining $\text{TPR}(\mathcal{S})$ also defines $\text{HRR}(\mathcal{S})$, provided \otimes is reinterpreted to denote circular convolution: $[a \otimes b]_\mu = \sum_\nu [a]_\nu [b]_{\mu-\nu}$. Assuming the elements of the $\{r_k\}$ are randomly (typically, normally) distributed, each role-embedding vector r_k can be used as its own unbinding vector. However the HRR unbinding equation holds only approximately: $s_k = \text{HRR}(\mathcal{S}) \cdot u_k + \text{noise}$. This noise must be eliminated by ‘clean-up’ processes. Note that, like TPR, HRR obeys the Superposition Principle.

Testing the Superposition Principle in the learned representation. As a test of whether the Superposition Principle holds of S-REP, let $v(\text{expr})$ denote the S-REP vector embedding of S-LANG expression expr , and consider expressions containing two symbol/role bindings, such as $aa:A \ \& \ bb:B$. Then if the Superposition Principle holds, we have¹:

1. $[v(aa:A \ \& \ bb:B) - v(aa:A \ \& \ cc:C)] - [v(dd:D \ \& \ bb:B) - v(dd:D \ \& \ cc:C)] = \mathbf{0}$.
2. $[v(aa:A \ \& \ bb:B) - v(aa:A \ \& \ cc:C)] - [v(xx:X \ \& \ uu:U) - v(xx:X \ \& \ vv:V)] \neq \mathbf{0}$.

We examine the Euclidean length of vectors of the form given on the LHS of Eq. 1, which we expect to not be exactly 0, but small — small, for example, relative to the LHS of Eq. 2. Fig. 2 shows that this is true: the AUC = 1.0 to within less than 10^{-16} .

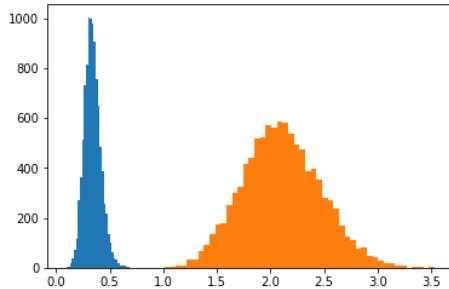


Figure 2: Distribution of lengths of vectors of the form of the LHS of Eqs. 1 and 2.

5 CONCLUSION

A standard bidirectional encoder-decoder model can generate vector embeddings of expressions denoting complex symbol structures and can successfully query the content of such representations. Like theoretical techniques for accomplishing this, the learned representation obeys the Superposition Principle (approximately; at least within the manifold of embeddings of two-binding expressions).

¹The simpler equation $v(aa:A \ \& \ bb:B) - [v(aa:A) + v(bb:B)] = \mathbf{0}$ does not hold in S-REP; it appears that the manifolds of one- and two-binding embeddings are distinct. Eq. 1 stays within the latter. Eq. 1 is analogous to the famous equation $[v(\text{king}) - v(\text{man})] - [v(\text{queen}) - v(\text{woman})] \approx 0$ (Mikolov et al., 2013).

REFERENCES

- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *Empirical Methods in Natural Language Processing-2014*, abs/1406.1078, 2014.
- Eric Crawford, Matthew Gingerich, and Chris Eliasmith. Biologically plausible, human-scale knowledge representation. *Cognitive Science*, 40(4):782–821, 2016.
- Geoffrey E. Hinton. Representing part-whole hierarchies in connectionist networks. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, pp. 48–54. 1988.
- Pentti Kanerva. Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computing*, 1:139–159, 2009.
- Moontae Lee, Xiaodong He, Wen-tau Yih, Jianfeng Gao, Li Deng, and Paul Smolensky. Reasoning in vector space: An exploratory study of question answering. In *Proceedings of the International Conference on Learning Representations-2016*, 2016.
- Tomas Mikolov, Scott Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT-2013)*. May 2013.
- Allen Newell. Physical symbol systems. *Cognitive Science*, 4(1):135–183, 1980.
- Hamid Palangi, Li Deng, Yelong Shen, Jianfeng Gao, Xiaodong He, Jianshu Chen, Xinying Song, and Rabab Ward. Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(4):694–707, 2016.
- Tony Plate. Holographic Recurrent Networks. In Stephen José Hanson and C Lee Giles (eds.), *Advances in Neural Information Processing Systems 5*. Morgan Kaufmann, San Mateo, CA, 1993.
- Tony Plate. Distributed representations. *Encyclopedia of Cognitive Science*, 2002.
- Tony Plate. *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*. CSLI Publications, Stanford, CA, 2003.
- Jordan B. Pollack. Recursive distributed representations. *Artificial Intelligence*, 46(1):77–105, 1990.
- Paul Smolensky. Tensor product variable binding and the representation of symbolic structures in connectionist networks. *Artificial Intelligence*, 46:159–216, 1990.
- Paul Smolensky and Géraldine Legendre. *The harmonic mind: From neural computation to Optimality-Theoretic grammar. 2 vols.* MIT Press, Cambridge, MA, 2006.
- Richard Socher, Christopher D Manning, and Andrew Y Ng. Learning continuous phrase representations and syntactic parsing with recursive neural networks. In *Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop*, pp. 1–9, 2010.
- David S. Touretzky. Boltzcons: Dynamic symbol structures in a connectionist network. *Artificial Intelligence*, 46:5–46, 1990.
- Jason Weston, Antoine Bordes, Sumit Chopra, Alexander M. Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. Towards AI-complete question answering: A set of prerequisite toy tasks. *arXiv:1502.05698*, 2015.
- Wojciech Zaremba and Ilya Sutskever. Learning to execute. *arXiv:1410.4615*, 2014.