Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc

Precise Embodying Tree Structures onto Their Latent Feature Vectors

Tiansi Dong

University of Bonn, Germany

ARTICLE INFO

Article History: Submitted 1.20.2022 Revised 1.26.2022 Accepted 1.30.2022

Keywords: Deep Learning Symbolic Structure Geometric Unification

ABSTRACT

The limitivism philosophy holds that an accurate connectionist account can only approximate good symbolic descriptions within certain limit. Grounding symbolic structure onto the vector space has been researched in the literature but precise solution has not yet emerged. Here, we present a geometric method that embodies symbolic tree structures precisely onto learned vector representation. This method turns vector embedding of symbols into nested sets of *n*-spheres (spheres in a higher dimensional space), with two desirable properties: (1) each vector embedding is well preserved by the central point of an *n*-sphere; (2) symbolic tree structures are precisely encoded by inclusion relations among *n*-spheres. This unified representation bridges the gap between Deep Learning and symbolic structural knowledge. Significant experiment results are obtained by embodying a large hypernym trees word-sense tree onto GloVE word embeddings of tree nodes. Our geometric method shows a new way to completely resolve the antagonism between connectionism and symbolicism.

© 2022 KSI Research

1. Introduction

A concept can be understood from two perspectives, one from the inside – its content, in terms of a set of features, the other from outside – its connections with other concepts, in terms of a symbolic structure. In the battle between the two perspectives, both sides believe the they explain the same phenomena [26]. If we imagine the two perspectives as different eyes of the monster *Artificial Intelligence*, how can this monster construct the external world in its mind using the inputs from the two heterogeneous eyes? Precisely, How, if possible at all, can discrete symbolic structures be (precisely) unified with their own feature vectors?

The two perspectives belong to the two paradigms in Artificial Intelligence, namely the *symbolic paradigm* and the *connectionist paradigm*. The symbolic paradigm is concerned with structural knowledge and rules for inference and decision making. A typical symbolic system consists of three components [9]: (i) symbols, either primitive or constructed; (ii) the meaning of constructed symbols, interpreted via the meaning of primitive symbols and the way of construction; (iii) reasoning via symbolic manipulations. The connectionist paradigm is inspired by the physiology of the brain and

dongt@bit.uni-bonn.de (T. Dong)
https://tiansidr.github.io/ (T. Dong)
ORCID(s):

models cognitive capabilities in terms of networks of simple computational nodes. Problem solving in this paradigm is to design and to train networks using exemplary data, in which knowledge is implicitly represented by weights of connections between nodes.

Approaches in the two AI paradigms are based on complementary mechanisms and target different levels of cognitive analysis [45, 50]. Symbolic approaches excel at reasoning but can hardly learn and are vulnerable to noisy inputs. Connectionist approaches, in particular Deep Learning [32], are robust to noisy or unforeseen inputs and capable of learning from data. However, they lack explainability, and are limited to approximated reasoning [25, 10], can be deliberately fooled by adversarial inputs [44, 28], and requiring much more training data than human learnes would need [29]. Nevertheless, connectionist approaches can make sense of data via *similarity judgments* [25, 52] and thus simulate one of three judgment methods under uncertainty [53].

An open challenge remains with respect to the question of how connectionist approaches can reach symbolic levels of reasoning [3, 49] or achieve cognitive modeling [39]. Researches in hybrid neuro-symbolic approaches aim at realising robust connectionist learning and sound symbolic reasoning [1, 2, 21, 16, 5]. Most of the approaches utilise interface neural networks to approximately bridge vector embed-

DOI reference number: 10-18293/JVLC2022-N1-100

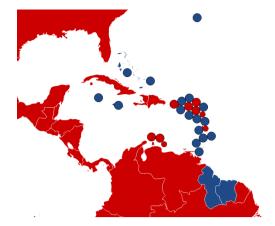


Figure 1: Blue areas represent islands with Left Hand Traffic System; red areas represent islands with Right Hand Traffic System

dings and symbols [17], in a way to roughly ground abstract symbols into a vector space [23, 24]. Geometrical structure is advocated as a potential cognitive representation apart from symbols or connectionist networks [18, 19].

Here, we geometrically construct a unified representation of a symbolic tree structure and its node embeddings: (1) vectorial node embeddings are promoted onto spheres in higher dimensional space; (2) the symbolic tree is spatialised onto these spheres such that inclusion relations among spheres precisely embody symbolic tree structures. The existence of these spheres shows the possibility to create a geometric continuum between symbolic and connectionist models, to completely resolve the antagonism between connectionism and symbolicism, and to realise the hope that to design artificial cognitive systems (the mind of the AI monster) that combine the two complementary paradigms (inputs from its two eyes) and solve problems from both perspectives [38, 9].

2. The Structuralist Eye Cannot Be Replaced by the Connectionist Eye

The *eliminativism* claims that connectionist approaches can sweep away (*eliminate*) symbolic approaches, and was refuted by observing that neural-networks were computed by symbolic computational devices (Turing machine) [7]. In this section, we give more evidences to refute the *eliminativism* perspective, starting from a thought experiment.

2.1. You won't trust neural navigators in way-finding

On your birthday party, your wonderful birthday gift is placed in the middle of a large maze. You are given a route instruction to find it at the end of the route. You are extremely excited, take the route instruction, and dive into the maze, and forget to ask how to get out of the maze. You call your friends outside the maze for help. They tell you the gift is a neural navigator. Given the current route instruction, it produces the next route instruction. Your friends try to convince you that a sequence of route instructions can be understood as a long sentence, and the neural navigator is exactly trained by the route you have, and routes of all other mazes around the world, with the same training mechanism as that for word-embeddings, e.g., [36]. Will you trust such kinds of neural navigator? If you know symbolic approaches to get out of the maze, such as the wall-follower method, or simply reversing the current route instruction, will you choose this neural navigator?

2.2. Connectionist eyes cannot see anything that structuralist eyes see

Figure 1 illustrates places where different traffic systems are used. The blue areas use the left-hand traffic system, the red areas use the right-hand traffic system. Only being fed with sufficient traffic scenarios, can autonomous driving cars learn there are two traffic systems by themselves? If they switch among right-hand traffic and left-hand traffic places, they cannot learn whether there are two traffic systems at all. It is hard to imagine that deep-learning systems can be intelligent enough to generate concepts of left and right, if they are only fed with traffic images without labeling which traffic systems. Because the concept of being left or right is not originated from images of street scenes. The original meaning of the left hand refers to the hand that is close to the heart of the body. Even for humans, the term of being left/right may not exist. For example, Guugu Yimithierr people only have absolute orientation corrdinates, such as north, south in their spatial conceptual system. A Guugu Yimithirr speaker would say something like "I left some tobacco on the southern edge of the western table in the house" [33, 41]. Connectionist eyes cannot see anything that structuralist eyes see. To teach connectionist eyes see objects, we have to teach them by correctly imposing object names with object images [4, 36, 46, 14, 57, 22, 47, 32]. External knowledge must be imposed onto the connectionist networks. This is not new to connectionists. In image recognition, they shall first precisely label object names to each image in the training set. If each cat image is labelled as 'dog', the well-trained networks will recognize each cat image as 'dog'.

2.3. Structures can exist without data

The existence of laws lies in the fact that violation exists in the reality. That *stealing is not allowed* as a law is due to the fact that *stealing* behaviors exist in the society. Even *stealing* behaviors does not exist, it still holds that *stealing is not allowed* – A phrase may be denoting, and yet not denote anything [43, p.41]. Only fed with data of stealing behaviors, connectionist networks would be more likely to mimic stealing behavior, rather than to be enlightened that stealing is not allowed. Excluding all stealing data from the training set, connectionist networks may not learn the concept of *stealing* at all.

2.4. Mental Representation of Partial Tree Structures

Structural knowledge is often modeled in terms of relations between or among entities. Figure 2(a) illustrates a

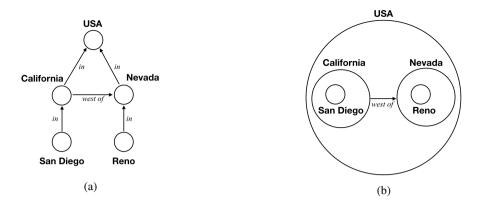


Figure 2: (a) A simple semantic network of spatial knowledge among California, Nevada, San Diego, and Reno; (b) A partially region-based hierarchical structure among California, Nevada, San Diego, and Reno

spatial structure among the cities San Diego and Reno, the stated of California and Nevada and the USA. Most people mistakenly judge that San Diego (CA) is further west than Reno (NV). The reason is that tree nodes are represented as regions in mind and that only relations between locations inside the same spatial region are explicitly stored [48], as illustrated in Figure 2(b). This region-based representation can explain systematical errors that people made in reasoning with spatial knowledge [54, 35].

The real challenge for connectionists is not to defeat symbolic theorists, but rather to come to terms with the ongoing relevance of the symbolic level of analysis [3, p.16]. Here, we examine the possibility to promote vectors of entities into regions such that inclusion relation shall precisely represent the child-parent relation in the tree, as illustrated in Figure 2(b).

3. The Statement of the Problem, and the Challenges

The problem addressed here can be stated as follows: Given a symbolic tree structure, and vector representations of its nodes, can we embody each tree node into a sphere such that (1) each child-parent relation in the tree structure (seen from the structuralist eye) is precisely encoded as inclusion relations among spheres; (2) the vector representation of a tree node (seen from the connectionist eye) is very well preserved by the sphere of the tree node.

The first criteria can be re-formulated within the connectionists' community as the criteria of reaching global loss zero. In the literature of connectionism, the termination condition of training processes only needs to be a local minimum. As we need to precisely encode all symbolic relations into inclusion relations among regions, we require global loss zero. A small scaled experiment in [11] shows that it is not possible to reach global minimum zero only by utilising the back-propagation method.

The second criteria may suggest us to create a sphere for a tree node by taking its vector as the central point of the sphere. This turns out to be not realistic. Take a hypernym tree as the example. In many cases, words, such as

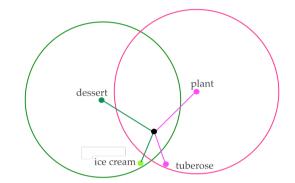


Figure 3: dessert sphere partially overlaps with plant sphere, although they should be disconnected from each other

ice_cream, tuberose, and their superordinate words, such as dessert, plant, seldom occur in the same context, their vector embeddings differ to such a degree that the cosine value is less than zero. For example, using GloVE embedding [40], the cos value of ice_cream and dessert is -0.1998, the cos value of tuberose and plant is -0.2191. This follows that the dessert sphere by taking the vector embedding as the central point, will contain the origin point of the embedding space, if it contains ice_cream sphere. The plant sphere by taking the vector embedding space, if it contains the origin point of the embedding space, sphere. Then, dessert sphere overlaps with plant sphere, as illustrated in Figure 3. Such overlapping is not allowed, as there is no entity which are both dessert and plant.

4. Constructing Spheres in Higher Dimensional Spaces

Promoting vectors into spheres appears deceptively simple, as it seems that we only need to add two new elements for each vector: one representing the length of the central point vector, the other representing the radius. Could the back propagation method be successful for this task? Experiments show that it cannot guarantee to achieve the target configuration precisely [12]. We abandon back propagation method, and use geometric construction and illustrate the

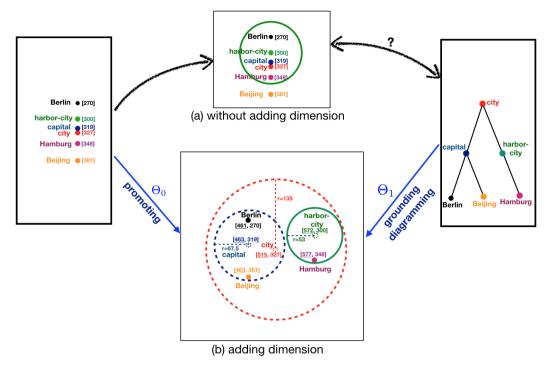


Figure 4: Spatializing symbolic structures onto vector space. Connectionist network represents a word as a one-element vector. Biased by training sentence, Hamburg vector [348] is closer to city vector [327] and captial vector [319] than to harbor-city vector [300]. To precisely encode symbolic tree structures, we promote them into circles, for example, harbor-city vector to a circle with central point [572, 300] with radius=53, with the target that inclusion relations among circles encode child-parent relations in the tree structure

method with the example as follows: Let the sentence "it was proposed to construct a maglev train between Berlin, capital of Germany, and harbor-city Hamburg" be in a training set, connectionist networks can capture co-occurrence relations among words in terms of vectors (word-embeddings). Figure 4 illustrates one dimensional word embeddings, e.g., the word embedding of Berlin is [270]. Suppose that Hamburg is a harbor-city, as a piece of truth-knowledge, be not clearly stated in the training text. As a result, Hamburg vector is closer to city vector and capital vector than to harbor-city vector, which violates the symbolic structure that Hamburg is-a harbor-city and Hamburg is-not-a capital. Such bias may lead connectionists and statisticians to stay at the level of approximation in reasoning and to believe the limitivism philosophy that an accurate connectionist account can only approximate good symbolic descriptions within certain limit. Our novel method is to promote a one dimensional wordembedding [y] into two dimensional circles with central point [x,y] and radius r. Then, we gear all xs and rs, so that inclusion relations among circles precisely encode the symbolic tree structure, as illustrated in Figure 4. Adding dimension is due to the fact that the space to embed semantic relations of words may not be the same as the space to embed their co-occurrence relations. Here, a symbol is embodied as a sphere in high dimensional space with the restriction that a part of the center vector is the vector embedding from connectionist networks. That is, symbols are only partially landed onto the vector embedding space. Through

such a symbol spatialising process, we embody a symbolic tree structure into a continuous space, so that the brittleness problem of symbolic approaches is removed. We describe the symbol spatializing process for a tree structure as follows.

4.1. Method

A tree \mathfrak{T} is a relational structure that can be described as a relational structure (T, S) [6] in which

- 1. T is the set of nodes $\{t_1, t_2, \dots, t_n\}$
- 2. S is the set of node pairs $\{(t_i, t_j) | t_i, t_j \in T\}$
- 3. \mathfrak{T} has a unique root node *r* that for any other node *t* there is a unique chain $[r = t'_1, \ldots, t'_i, \ldots, t = t'_w]$ under the condition that neighborhood nodes, t'_j and t'_{i+1} , are node pairs in S.
- 4. For every non-root node *u*, there is a unique node pair (v, u) in S.
- 5. For any node x, (x, x) does not exist in S

We geometrically represent each tree node *t* as a sphere \odot_t with central point O_t and radius r_t , and node pair as the *cover* relation COV as follows: (t_i, t_j) is in *S*, if and only if \odot_{t_i} covers \odot_{t_j} , written as COV $(\odot_{t_i}, \odot_{t_j})$. Geometrically, we define that \odot_{t_i} covers \odot_{t_j} , if and only if radius r_{t_i} is greater than the sum of r_{t_j} and the distance between their central points $||O_{t_i} - O_{t_j}||$, that is, $r_{t_i} > r_{t_j} + ||O_{t_i} - O_{t_j}||$. This greater than relation excludes the case that a sphere covers

sphere	contained by spheres	
\odot (beijing.n.01)	$\bigcirc(city.n.01) \subset \bigcirc(municipality.n.01) \subset \bigcirc(region.n.03)$	
$\odot(berlin.n.01)$	$\subset \odot(location.n.01)$	
$\odot(berlin.n.02)$	\bigcirc (songwriter.n.01) $\subset \bigcirc$ (composer.n.01) $\subset \bigcirc$ (musician.n.02)	
	$\subset \odot(artist.n.01)$	
⊙(hamburg.n.01)	$\odot(port.n.01) \subset \odot(point.n.02) \subset \odot(location.n.01) \subset \odot(object.n.01)$	

Table 1

Child-parent relations are encoded by continuous inclusion relations among spheres. The direct hypernym of a word-sense w is the word-sense x whose sphere is the smallest sphere that covers w's sphere. $\odot(w)$ represents the sphere of $w, A \subset B$ represents COV(B, A)

word-sense 1	word-sense 2	word
beijing.n.01	london.n.01, atlanta.n.01,	china, taiwan, seoul, taipei, chinese, shanghai,
	washington.n.01, paris.n.0,	korea, mainland, hong, wen, kong, japan,
	potomac.n.02, boston.n.01	hu, guangzhou, chen, visit, here, tokyo, vietnam
berlin.n.01	madrid.n.01, toronto.n.01,	vienna, warsaw, munich, prague, germany,
	rome.n.01, columbia.n.03,	moscow, hamburg, bonn, copenhagen, cologne,
	sydney.n.01, dallas.n.01	dresden, leipzig, budapest, stockholm, paris,
berlin.n.02	simon.n.02, williams.n.01,	frankfurt, amsterdam,german,stuttgart,brussels
	foster.n.01, dylan.n.01,	petersburg, rome, austria, bucharest, düsseldorf,
	mccartney.n.01, lennon.n.01	zurich, kiev, austrian, heidelberg, london
hamburg.n.01	glasgow.n.01, bristol.n.01,	munich, stuttgart, bundesliga, frankfurt,
	oslo.n.01, santos.n.01,	freiburg, bayern, borussia, vfb, fc,,
	colon.n.04, hull.n.05	germany, werder, bremen, eintracht, berlin

Table 2

Top-6 sphere nearest neighbors compared with top-N GloVe nearest neighbors. Neighbours of a sphere are strictly constrained by hypernym structures. In our tree structure, 'hamburg.n.01' is the neighbor of other ports. In contrast, GloVe neighbours are biased training corpus, mixing with or neglecting other word-senses. Glove neighbors of 'hamburg' are severely biased to football-related training sentences, neighbors 'beijing' mixes with names of countries and persons, the word-sense of family names of 'berlin' is totally neglected in its neighborhood

itself (condition 5). The *cover* relation is transitive, that is, if \bigcirc_{t_i} covers \bigcirc_{t_j} , and \bigcirc_{t_j} covers \bigcirc_{t_k} , then \bigcirc_{t_i} covers \bigcirc_{t_k} . This follows that the root sphere covers all other spheres.

We adopt depth-first recursive process to traverse the nodes in a tree structure. A parent sphere will be constructed after all its child spheres are constructed, as illustrated in Figure 5(a). Geometric construction is carried out as a sequence of operations selected from three geometric transformations as follows.

- 1. A Homothetic operation on sphere \odot with the ratio k(k > 0), written as $H(\odot, k)$, zooms out the length of the vector of its central point and the radius with the same ratio k, $H(\odot(O, r), k) = \odot(kO, kr)$, as illustrated in Figure 5(g).
- 2. A Shifting operation on sphere \odot with vector \vec{v} , written as $S(\odot, \vec{v})$, shifts this sphere with vector \vec{v} , $S(\odot, \vec{v}) = \odot(O + \vec{v}, r)$, as illustrated in Figure 5(h).
- 3. A Rotation operation rotates sphere $\bigcirc(O, r)$ with unit vector $\vec{\beta}$ in the subspace spanned by the *i*-th and the *j*-th basis, written as $R(\bigcirc(O, r), \vec{\beta}, i, j) = \bigcirc(O', r)$, in which $O_k = O'_k$ for all $k \neq i, j, O'_i = O_i \cos \beta + O_j \sin \beta, O'_j = -O_i \sin \beta + O_j \cos \beta$, as illustrated in Figure 5(i).

4.2. Experiments

GloVe word-embeddings [40] are used as the pre-trained word-embedding, hypernym trees among word-senses are extracted from WordNet 3.0 [37], totaling 291 hypernym trees and 54,310 spheres [12], each representing a wordsense in a hypernym tree. These sphere embeddings have 32,503 word-stems. Precise spatialisation has been achieved, pre-trained GloVe word-embeddings have been very wellpreserved. Only a tiny portion (1.3%) of pre-trained wordembeddings indicates a small variation (std $\in (0.1, 0.7666]$). Symbolic tree structure is precisely embedded onto the continuous space, which leads to precise encoding of category information, as illustrated in Table 1, and precise separation of word-senses in nearest neighborhood experiments, as illustrated in Table 2. In order to prevent the deterioration of already constructed relations, we will apply the same transformation for all its child spheres, if we apply a geometric transformation for a sphere. The recursive geometric construction process will generate a sequence of transformations for the construction of the sphere of a tree node that transform a sphere from its initial status to the final status. This dynamic information can be likened as a route instruction that tells a baby's home address starting from the hospital address where it is born. If we already have the route instruction of its siblings, we can use it to send the new born baby home. Using this idea, we have conducted experiments

T.Dong / Journal of Visual Language and Computing (2022) 16-24

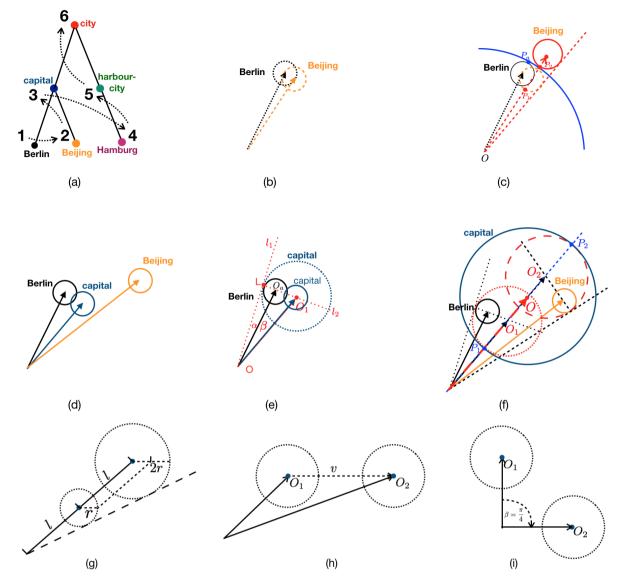
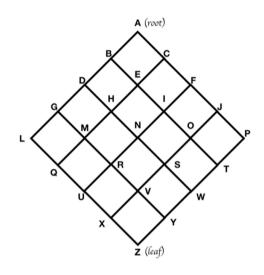


Figure 5: (a) Depth-first sequence to update spheres; (b-c) homothetic transformation will be applied for overlapped sibling spheres; (d-e) construct one *capital* sphere that covers the Berlin sphere; (f) construct the final *capital* sphere that covers all existing *capital* spheres; (g) homothetic transformation with k = 2, (h) shifting transformation with \vec{v} , (i) rotation transformation with $\beta = \frac{\pi}{4}$

to validate the category of an unknown word that appears in corpus. For example, when we read a text *Solingen has long been renowned for the manufacturing of fine swords, knives, scissors and razors*..., we wonder whether *Solingen* is a *city* or a *person*? Supposing it is a city, we initialize its sphere by using the type information of *city*. One of its sibling is *Berlin* that we have its route instruction that guides it into the sphere of *city*. So, we use this route to guide *Solingen*. If it is finally located inside sphere of *city*, we will predict that *Solingen* is a *city*, otherwise not. We have experimented this method to predict the type of unknown entities in knowledge graph [13]. Our experiments show that this geometric approach greatly outperforms traditional embedding approaches, especially when the route is long.

5. Conclusion

The relation between neural networks and symbolic structures remains an open debate. This debate is nothing new and dates back to the antagonism between Connectionism and Symbolicism, e.g., [45]. The difficulty is that precise encoding of symbolic relations cannot be achieved by the back-propagation algorithm – the fundamental algorithm of Connectionism. Here, sphere embeddings are created using geometric construction, and by abandoning back propagation method. Our methodology does not belong to the connectionism paradigm. The created sphere embeddings only exist in a space whose dimension is higher the vector space produced by connectionist networks. This refutes



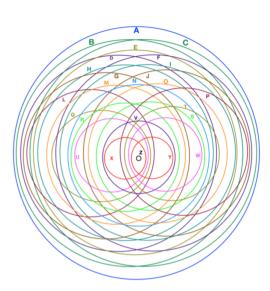


Figure 6: Spatializing a complex lattice

eliminativism, implementationalism, and *revisionism*¹, and also refutes the analogy between the symbolic-subsymbolic approaches and the relation between macro–micro physical theories [45]. The sphere embeddings should appeal psychologists, for discrete symbolic tree structures have been precisely transformed into sphere configurations in a continuous space [45, 8]. They also appeal cognitive linguists [30], for now bounded regions and paths can be implemented by geometric transformations.

The nested structure of sphere configurations would solve the *elaboration tolerance* problem² in connectionist networks [34]. The sphere configuration favors both *limitivism* and hybridism³, and serves as a big geometric *patchwork* that bridges symbolic and network components, and promises to merge the two complementary methods and theories in cognitive science [39]. For example, the grounding model of metaphor [31, 20], the embodiment of language and thoughts [41, 15], and spatial thinking [55]. It also favors the way of killing two birds with one stone [56], in the sense that it is able to precisely reproduce the two birds: subsymbolic vectors, and symbolic structure - It kills the two birds without loosing any information about them. The success of the geometric approach largely depends on being able to precisely spatialise more complex symbolic structures onto vector space. Figure 6 shows an on-going work to spatialize a

complex grid onto vector space.

Connectionism is not a complete theory for learning [27]. Learning through huge amount of data using back-propagation is bottom-up [51] and inefficient, and only establishes a harmony between input and output. An important learning style in schools and universities is learning under instruction (topdown style of learning [51]). It is easy to translate "white as snow" into German ("weiß wie Schnee"), French ("blanc comme neige"), and many other languages. How shall we translate it into the Natemba language? People who speak Natemba live in Benin, a country near to the equator, where the temperature is around 20°C in the winter, so no snowing in the winter, as a consequence, no word for snow in the Natemba language. To describe something very white there, people would say "white as pelican" (pelican is a kind of white bird. This is an example of elaboration tolerance in translation, see footnote 2). We would feel this translation is interesting and reasonable, after having been informed about the right background knowledge. Methods should be developed to precisely inform (or impose) external knowledge to pure data-driven machine learning systems [42].

Acknowledgement

I am indebted to S.K. Chang for the invitation and his long-term interest in my work, and to Ron Sun for the comments on an early version of this paper. This research is funded by the Federal Ministry of Education and Resarch of Germany as part of the competence center for machine learning ML2R under grant number 01/S18038C.

References

- Bader, S., Hitzler, P., 2004. Logic programs, iterated function systems, and recurrent radial basis function networks. Journal of Applied Logic, Special Issue on Neural-Symbolic Systems 2, 273–300.
- [2] Bader, S., Hitzler, P., 2005. Dimensions of neural-symbolic integration — a structured survey, in: Artemov, S., Barringer, H., Garcez,

¹*eliminativism* philosophy holds that connectionist approach can achieve all symbolic approach can do; *implementationalism* philosophy holds that neural-network is the "hardware" of symbolic system; *revisionism* philosophy holds that a symbolic account can be generated by connectionist networks

²*elaboration tolerance* challenges whether connectionist network can be elaborated with additional phenomenon.

³*limitivism* philosophy holds that accurate connectionist account can approximate good symbolic descriptions within certain limit; *hybridism* philosophy holds that a *patchwork* can be created for the gap between symbolic and neural-network components

⁴*killing two birds with one stone* refers to a two-system hypothesis of the mind that the same neural event that is capable of simultaneously manipulate conceptual-level symbols and perform subsymbolic operations

A.S.d., Lamb, L.C., Woods, J. (Eds.), We Will Show Them: Essays in Honour of Dov Gabbay. King's College Publications. volume 1, pp. 167–194.

- [3] Bechtel, W., Abrahamsen, A., 2002. Connectionism and the mind: Parallel processing, dynamics, and evolution in networks. Graphicraft Ltd, Hong Kong.
- [4] Bengio, Y., Ducharme, R., Vincent, P., Janvin, C., 2003. A neural probabilistic language model. J. Mach. Learn. Res. 3, 1137–1155. URL: http://dl.acm.org/citation.cfm?id=944919.944966.
- [5] Besold, T.R., d'Avila Garcez, A.S., Bader, S., Bowman, H., Domingos, P.M., Hitzler, P., Kühnberger, K., Lamb, L.C., Lowd, D., Lima, P.M.V., de Penning, L., Pinkas, G., Poon, H., Zaverucha, G., 2017. Neural-symbolic learning and reasoning: A survey and interpretation. CoRR abs/1711.03902. URL: http://arxiv.org/abs/1711.03902, arXiv:1711.03902.
- [6] Blackburn, P., 2000. Representation, Reasoning, and Relational Structures: a Hybrid Logic Manifesto. Logic Journal of the IGPL 8, 339–625.
- [7] Chalmers, D.J., 1992. Subsymbolic Computation and the Chinese Room, in: The Symbolic and Connectionist Paradigms: Closing the Gap, Erlbaum. pp. 25–48.
- [8] Dellarosa, D., 1988. The psychological appeal of connectionism. Behavioral and Brain Sciences 1, 28–29.
- [9] Dinsmore, J., 1992. Thunder in the Gap, in: The Symbolic and Connectionist Paradigms: Closing the Gap, Erlbaum. pp. 1–23.
- [10] Dong, H., Mao, J., Lin, T., Wang, C., Li, L., Zhou, D., 2019a. Neural Logic Machines, in: ICLR-19, New Orleans, USA.
- [11] Dong, T., 2021. A Geometric Approach to the Unification of Symbolic Structures and Neural Networks. volume 910 of *Studies in Computational Intelligence*. Springer-Nature.
- [12] Dong, T., Bauckhage, C., Jin, H., Li, J., Cremers, O.H., Speicher, D., Cremers, A.B., Zimmermann, J., 2019b. Imposing Category Trees Onto Word-Embeddings Using A Geometric Construction, in: ICLR-19, New Orleans, USA. May 6-9.
- [13] Dong, T., Wang, Z., Li, J., Bauckhage, C., Cremers, A.B., 2019c. Triple Classification Using Regions and Fine-Grained Entity Typing, in: AAAI.
- [14] Faruqui, M., Dodge, J., Jauhar, S.K., Dyer, C., Hovy, E., Smith, N.A., 2015. Retrofitting word vectors to semantic lexicons, in: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, ACL. pp. 1606–1615. URL: http://www.aclweb.org/anthology/ N15-1184, doi:10.3115/v1/N15-1184.
- [15] Feldman, J., 2006. From Molecule to Metaphor: A Neural Theory of Language. The MIT Press, Cambridge, Masachusetts.
- [16] Feldman, J., 2013. The neural binding problem(s). Cogn Neurodyn 7, 1–11.
- [17] d'Avila Garcez, A.S., Lamb, L.C., Gabbay, D.M., 2007. Connectionist modal logic: Representing modalities in neural networks. Theor. Comput. Sci. 371, 34–53.
- [18] G\u00e4rdenfors, P., 2000. Conceptual Spaces The Geometry of Thought. MIT Press, Cambridge, Massachusetts, USA.
- [19] G\u00e4rdenfors, P., 2017. The Geometry of Meaning. MIT Press, Cambridge, Massachusetts, USA.
- [20] Grady, J., 1997. Foundations of Meaning: Primary Metaphors and Primary Scenes. University Microfilms.
- [21] Hammer, B., Hitzler, P. (Eds.), 2007. Perspectives of Neural-Symbolic Integration. Springer.
- [22] Han, X., Liu, Z., Sun, M., 2016. Joint representation learning of text and knowledge for knowledge graph completion. CoRR abs/1611.04125. arXiv:1611.04125.
- [23] Harnad, S., 1990. The symbol grounding problem. Phys. D 42, 335– 346.
- [24] Harnad, S., 2003. The symbol grounding problem, in: Encyclopedia of Cognitive Science, Macmillan.
- [25] Hinton, G.E., 1981. Implementing semantic networks in parallel hardware, in: Hinton, G.E., Anderson, J.A. (Eds.), Parallel Models of Associative Memory. Erlbaum, Hillsdale, NJ, pp. 161–187.

- [26] Hinton, G.E., 1986. Learning distributed representations of concepts, in: Proceedings of the eighth annual conference of the cognitive science society, Amherst, MA. p. 12.
- [27] Hunter, L.E., 1988. Some memory, but no mind. Behavioral and Brain Sciences 1, 37–38.
- [28] Jia, R., Liang, P., 2017. Adversarial examples for evaluating reading comprehension systems, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017, pp. 2021–2031.
- [29] Lake, B., Salakhutdinov, R., Tenenbaum, J., 2015. Human-level concept learning through probabilistic program induction. Science 350, 1332–1338.
- [30] Lakoff, G., 1988. Smolensky, semantics, and the sensorimotor system. Behavioral and Brain Sciences 1, 39–40.
- [31] Lakoff, G., Johnson, M., 1980. Metaphors We Live By. The University of Chicago Press, Chicago. Citation is based on the reprinted in 2003.
- [32] LeCun, Y., Bengio, Y., Hinton, G.E., 2015. Deep learning. Nature 521.
- [33] Levinson, S.C., 1997. Language and Cognition: The Cognitive Consequences of Spatial Description in Guugu Yimithirr. Journal of Linguistic Anthropology 7, 98–131.
- [34] McCarthy, J., 1988. Epistemological challenges for connectionism. Behavioral and Brain Sciences 1, 44.
- [35] McNamara, T.P., 1991. Memory's View of Space. The Psychology of Learning and Motivation 27, 147–186.
- [36] Mikolov, T., 2012. Statistical Language Models Based on Neural Networks. Ph.D. thesis. Brno University of Technology. Brno, CZ.
- [37] Miller, G.A., 1995. Wordnet: A lexical database for english. Commun. ACM 38, 39–41.
- [38] Newell, A., 1990. Unified theories of cognition. Harvard University Press, Cambridge, MA.
- [39] Núñez, R., Allen, M., Gao, R., Rigoli, C.M., Relaford-Doyle, J., 2019. What happened to cognitive science? Nature Human Behaviour 3, 782–791.
- [40] Pennington, J., Socher, R., Manning, C.D., 2014. GloVe: Global Vectors for Word Representation, in: EMNLP'14, pp. 1532–1543.
- [41] Regier, T., 1997. The Human Semantic Potential: Spatial Language and Constrained Connectionism. The MIT Press, Cambridge, Massachusetts.
- [42] von Rüden, L., Mayer, S., Garcke, J., Bauckhage, C., Schücker, J., 2019. Informed Machine Learning - Towards a Taxonomy of Explicit Integration of Knowledge into Machine Learning. CoRR abs/1903.12394. URL: http://arxiv.org/abs/1903.12394.
- [43] Russell, B., 1956. Logic and Knowledge. Routledge, an imprint of Taylor & Francis Books Ltd.
- [44] Seo, M.J., Kembhavi, A., Farhadi, A., Hajishirzi, H., 2016. Bidirectional attention flow for machine comprehension. CoRR abs/1611.01603. URL: http://arxiv.org/abs/1611.01603, arXiv:1611.01603.
- [45] Smolensky, P., 1988. On the proper treatment of connectionism. Behavioral and Brain Sciences.
- [46] Socher, R., Chen, D., Manning, C.D., Ng, A., 2013. Reasoning with neural tensor networks for knowledge base completion, in: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 26. Curran Associates, Inc., pp. 926–934.
- [47] Speer, R., Chin, J., Havasi, C., 2017. Conceptnet 5.5: An open multilingual graph of general knowledge, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA., pp. 4444–4451.
- [48] Stevens, A., Coupe, P., 1978. Distance estimation from cognitive maps. Cognitive Psychology 13, 526–550.
- [49] Sun, R., 2002. Hybrid connectionist symbolic systems, in: Arbib, M. (Ed.), Handbook of Brain Theories and Neural Networks (2nd Edition). MIT Press, Cambridge, MA., pp. 543–547.
- [50] Sun, R., 2015. Artificial Intelligence: Connectionist and Symbolic Approaches, in: James, D.W. (Ed.), International Encyclopedia of

the Social and Behavioral Sciences (2nd edition). Pergamon/Elsevier, Oxford, pp. 35 - 40.

- [51] Sun, R., 2016. Implicit and explicit processes: Their relation, interaction, and competition, in: Macchi, L., Bagassi, M., Viale, R. (Eds.), Cognitive Unconscious and Human Rationality. MIT Press, Cambridge, MA., pp. 257–27.
- [52] Tversky, A., 1977. Features of similarity. Psychological Review 84, 327–353.
- [53] Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. Science 185, 1124–1131.
- [54] Tversky, B., 1981. Distortions in Memory for Maps. Cognitive Psychology 13, 407–433.
- [55] Tversky, B., 2019. Mind in Motion. Basic Books, New York, USA.
- [56] Woodfield, A., Morton, A., 1988. The reality of the symbolic and subsymbolic systems. Behavioral and Brain Sciences 1, 58.
- [57] Xiao, H., Huang, M., Zhu, X., 2016. From one point to a manifold: Knowledge graph embedding for precise link prediction, in: IJCAI.