International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:2, No:1, 2008

Self-Assembling Hypernetworks for Cognitive Learning of Linguistic Memory

Byoung-Tak Zhang, and Chan-Hoon Park

***Abstract*—**Hypernetworks are a generalized graph structure representing higher-order interactions between variables. We present a method for self-organizing hypernetworks to learn an associative memory of sentences and to recall the sentences from this memory. This learning method is inspired by the "mental chemistry" model of cognition and the "molecular self-assembly" technology in biochemistry. Simulation experiments are performed on a corpus of natural-language dialogues of approximately 300K sentences collected from TV drama captions. We report on the sentence completion performance as a function of the order of word-interaction and the size of the learning corpus, and discuss the plausibility of this architecture as a cognitive model of language learning and memory.

***Keywords*—**Linguistic recall memory, sentence completion task, self-organizing hypernetworks, cognitive learning and memory.

I. INTRODUCTION

WE investigate the use of the hypernetwork structure as a model of linguistic memory at the sentence level. Hypernetworks are originally proposed as an associative memory model inspired by and realized in molecular self-assembly [10, 5]. A hypernetwork consists of a huge number of hyperedges, each of which links vertices of arbitrary size and thus is able to encode higher-order interactions among the variables. In language modeling, this combinatorial property can be used for learning the higher-order associations of the words from massive cognitive data. The hypernetworks are constructed by a self-organizing process based on three primitive operations of matching, selection, and amplification of hyperedges. Given a learning-data item, the hypernetwork memory is matched against the hyperedges and the matching elements are selected and amplified. Each of these primitive operations occurs in a massively parallel way using the molecular self-assembly, i.e. the hyperedges as memory elements recognize each other by molecular recognition. This chemical recognition property of the hypernetworks provides an interesting analogy to the "mental chemistry" model of the mind stated by John Stuart Mill more than a hundred years ago [1].

Here we report the experimental results of using hypernetworks to learn language models from a corpus of 290K sentences collected from TV dramas. This task would be trivial if we use hyperedges of order *n*, where *n* is the length of the longest sentence in the corpus. This would be equivalent to indexing and storing the whole sentences into a text database. However, this will not lead to any generalization and cannot generate unseen sentences. We want the hypernetwork to learn in a cognitively plausible way and demonstrate linguistic recall capability, such as making a complete sentence or generating a new sentence given a partial list of words.

Conceptually, the hyperedges can represent memory chunks or micromodules or cognitive schema [1, 2, 8]. These mental codes can be mapped to neural microcircuits [3, 4, 7]. Thus, applied to cognitive tasks such as language and vision, the self-organizing hypernetwork model can simulate the cognitive learning and memory processes with potential mapping to their neural substrates. Much data have been collected recently to bridge the gap between cognitive computation and neural computation [2,3,7]. This cognitive learning approach will also make a step forward to the human-level intelligence [6, 9].

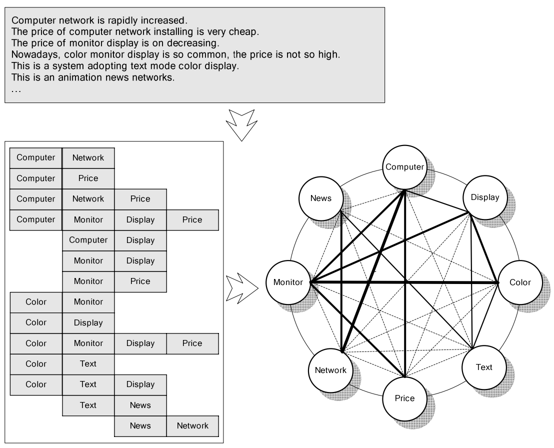

Fig. 1 Constructing a hypernetwork from a corpus of sentences. The hyperedges of the hypernetwork are made by randomly sampling the words in a sentence. For a given sentence, a large number of hyperedges are made with duplication allowed. The number of copies of the hyperedge represents the strength of the connections among the words. Thus, the hypernetwork represents an associative linguistic memory of the language corpus, where the vertices denote the words and the weighted edges the association strengths among the words.

This work was supported by the Ministry of Science and Technology (NRL 2002-2007), the Ministry of Commerce, Industry, and Energy (MEC 2006-2009), the Ministry of Education and Human Resources (BK21-IT 2006-2009).

Byoung-Tak Zhang is with the School of Computer Science and Engineering, Graduate Programs in Cognitive Science and Brain Science, Seoul National University, Seoul 151-744, Korea (phone: +82-2-880-1833; fax: +82-2-875-2240; e-mail: btzhang@bi.snu.ac.kr).

Chan-Hoon Park is with the School of Computer Science and Engineering, Seoul National University, Seoul 151-744, Korea (e-mail: chpark@bi.snu.ac.kr).

International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:2, No:1, 2008

Technically, there is a fundamental tradeoff between using large and small hyperedges (chunks) for building the hypernetwork memory. Large-chunk hypernetworks will be good at correctly memorizing the training sentences but too-large memory chunks do not generalize well to unseen sentences. In contrast, small-chunk hypernetworks might show low performance in memorizing training sentences but will tend to generalize better to unobserved sentences. Figuring out the best balance would be interesting from the cognitive science point of view, but challenging from the computational point of view, especially from machine learning. To see the effect of hyperedge size, we also investigate the effect of the order of the hyperedges on the memory recall performance.

The paper is organized as follows. In Section II we describe how the sentences are stored in the hypernetwork model and how new sentences are generated from it. In Section III we report on the experimental results on the language corpus in the TV drama domain. Section IV discusses the implications of this work.

II. Hypernetwork Models of Language

A. Hypernetwork Models

A corpus *D* of sentences $x^{(n)}$ is given as a learning material. The goal is to encode the sentences onto the hypernetwork structure to build a language model. After learning, the hypernetwork memory will be used to generate new sentences from some cues, such as a corrupt sentence or a partial list of words. Fig. 1 shows an example corpus consisting of sentences related to computers.

B. Encoding Sentences onto the Hypernetwork

The hypernet *H* is initialized by getting sentences from the corpus and making hyperedges from them. For the *n*-th sentence, $x^{(n)} = (x_1, x_2, \ldots, x_n)$, a large number of *k*-th order hyperedges $E_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$ are made by randomly sampling the words. The network is then adapted or "self-organized" by repeatedly observing the sentences. In each observation of a sentence, a large collection of query hyperedges $E_j^{(q)}$ are randomly sampled from $x^{(n)}$ and then they are matched against or "self-assembled with" the hyperedges E_i in the learned hypernetwork *H*. The matching E_i are then copied. When the weight of the hyperedges normalized, the distribution of the weights represents the probability distribution of the word association patterns. The method has been implemented as a language game platform shown in Fig. 2.

C. Generating Sentences from the Hypernetwork

The decoding process is similar to the query process described above. The difference is that the query in decoding can be missing. A missing word is generated by collecting all the hyperedges having the same context. For example, given the corrupt sentence query $x^{(q)}$ = ("who", ?, "you"), the missing word ? can be generated by getting all the hypereges E_i = ("who", *, "you") and choosing the word in the position * that appear most frequently.

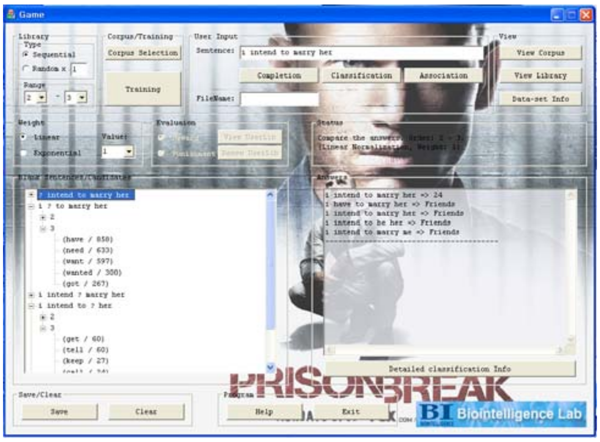

Fig. 2 The language game platform. This software allows for learning hypernetworks from a text corpus and generating sentences from corrupt query sentences using the learned hypernetwork recall memory of language. The display in this particular screen shot shows that given a series of five corrupt versions (a missing word for each position) of the learned sentence "I intend to marry her" the hypernetwork can generate variations of the original sentence learned, such as "I have to marry her" or "I intend to be her". The left panel of the window shows the relative strengths of the candidate words in recall.

TABLE I
ILLUSTRATIVE RESULTS FOR SENTENCE COMPLETION AND CLASSIFICATION TASKS

Query	Completion	Classification
who are you	**Corpus: Friends, 24, Prison Break**	
? are you	what are you	Friends
who ? you	who are you	Friends
who are ?	who are you	Friends
you need to wear it	**Corpus: 24, Prison Break, House**	
? need to wear it	i need to wear it	24
you ? to wear it	you want to wear it	24
you need ? wear it	you need to wear it	24
you need to ? it	you need to do it	House
you need to wear ?	you need to wear a	24

Table I shows the example results for two different decoding procedures for the same hypernetwork memory, i.e. recall and recognition tasks. In this particular experiment, the sentences were labeled with the title of the movie, i.e. the source of the sentence. For example, a training sentence consists of ("who", "are", "you", Friends), where Friends denotes the sentence "Who are you" came from the drama Friends. In the recall task, the hypernetwork is given, say, (?, "are", "you") and to complete the missing word to produce "What are you". In the recognition or classification task, the hypernetwork is to output "Friends" as the source of the sentence.

For the sample sentence "You need to wear it" which appeared in movies 24 as well as House, the learned hypernetwork could generate, for example, the sentences like "I need to wear it", "You want to wear it", and "You need to do it" with the right recognition of the sources.

III. Experimental Results

The corpus of sentences consists of approximately 300,000 sentences from DVD dramas, including Friends, House, 24, Grey Anatomy, Sex and the City, and Gilmore Girls. The hypernetwork was trained by the sentences in the entire corpus. The recall capability of the hypernetwork memory was tested on a collection of 700 sentences randomly chosen out of the training corpus. Note that we use the training sentences for the test purposes, but the recall is not trivial because the original sentences are encoded and learned in small fragments (chunks) rather than as a complete sentence. This makes the evaluation of recall capability simple and easy.

Fig. 3 shows some example sentences generated by the recall process of the hypernetwork memory. Only two of the six sentences are perfect (“**I appreciate it if you call her by the way”**), but the rest also makes sense (“**Would you *nice* to meet you in Tuesday *and*”**) or syntactically correct (“**I think but I am *met* him somewhere before”**).

The whole space of the sentences to be generated is huge while we usually have a small size of corpus available. Thus, it is interesting to see how the performance or robustness of the memory behaves as the size of training corpus increases. Can the completion task be (rather) robustly learned by a collection of a single drama or it is too small to show any interesting recall performance?

Table II summarizes the experimental results. Here the rows show the number of sentences correctly recalled out of 700 test sentences (which was used during training). We experimented with three different values of orders, i.e. k = 2, 3, 4. For each fixed *k*, we varied the number of missing words (the number of ?’s in the query sentence). Each column is for a different number of training sentences (corpus size), increasing from left to right.

The results show that as the corpus size increases the number of recalls also increases (i.e. becomes more robust). For the same corpus size, as the order *k* increases the number of recalls tends to decrease. This makes sense because for large k the probability of finding a matching hyperedge decreases. For a given k, the increase of corpus size from 40K to 290K increases makes the number of recalls increasing, but the effect is seems linear and very slow. This can be interpreted in two ways. One is that the size of 40K is already relatively big enough making the memory already robust. The other interpretation is that 40K and 290K are in the range of a too small corpus size, making no significant difference in improving the recall performance. In the second case, we would need a much bigger corpus to get a small amount of improvement in recall.

To see this, we examined the real sentences reconstructed by the hypernetworks. The following examples show the sentences made by the recall memory for various combinations of order *k* and the number of missing words.

i) Generating a single word from context:

It was my mother’s ? (using order k=2)
⇨ It was my mother’s been (~ 80,000)
⇨ It was my mother’s life (~ 160,000)
⇨ It was my mother’s not (~ 200,000)
⇨ It was my mother’s life
It was my mother’s ? (using order k=3)
⇨ It was my mother’s the (~ 40,000)
⇨ It was my mother's living (~ 80,000)
⇨ It was my mother’s life (~ 160,000)
⇨ It was my mother’s dead

? starts looking at me (using order k=2)
⇨ That starts looking at me (~ 40,000)
⇨ He starts looking at me (~ 200,000)
⇨ It starts looking at me
? starts looking at me (using order k=3)
⇨ Anyone starts looking at me

Why ? you ? come ? down ? ⇨ **Why are you *go come* on down here**
? think ? I ? met ? somewhere before ⇨ **I think but I am *met* him somewhere before**
? appreciate it if ? call her by ? ? ⇨ **I appreciate it if you call her by the way**
I'm standing ? the ? ? ? cafeteria ⇨ **I'm standing in *the one* of the cafeteria**
Would you ? to meet ? ? Tuesday ? ⇨ **Would you *nice* to meet you in Tuesday *and***
? gonna ? upstairs ? ? a shower ⇨ **I'm gonna go upstairs and take a shower**
? have ? visit the ? room ⇨ **I have to visit the ladies' room**
We ? ? a lot ? gifts ⇨ **We don't have a lot of gifts**
? ? don't need your ? ⇨ **if I don't need your help**
? ? ? decision ⇨ **to make a decision**
? still ? believe ? did this ⇨ **I still can't believe you did this**
What ? ? ? here ⇨ **What are you doing here**
? you ? first ? of medical school ⇨ **Are you go first day of medical school**
? ? a dream about ? in Copenhagen ⇨ **I had a dream about you in Copenhagen**

Fig. 3 Results of the sentence completion task with a relatively large number, i.e. half of the sentence length, of missing words

TABLE II
The Number of Correct Recalls as a Function of the Size of the Training Corpus

Order	?	The size of the training corpus							
		40K	80K	120K	160K	200K	240K	280K	290K
2	1	691	697	698	699	699	699	699	699
	2	515	522	527	528	530	531	532	532
	3	295	302	309	309	309	309	309	309
	4	174	175	176	178	178	178	178	178
3	1	565	608	617	628	632	634	636	637
	2	619	642	654	658	661	663	666	667
	3	528	553	565	567	569	571	577	577
	4	460	481	488	489	489	489	490	490
4	1	405	462	500	516	536	546	550	550
	2	545	580	594	601	605	608	612	612
	3	570	595	608	614	621	626	633	633
	4	550	577	589	592	596	600	602	602

the roman ? (using order k = 2)
- ⇨ the roman numeral (~ 160,000)
- ⇨ the roman empire

the roman ? (using order k = 3)
- ⇨ the roman numeral (~ 160,000)
- ⇨ the roman empire

They seemed to take it pretty ? (using order k = 2)
- ⇨ They seemed to take it pretty much (~ 40,000)
- ⇨ they seemed to take it pretty good

They seemed to take it pretty ? (using order k = 3)
- ⇨ They seemed to take it pretty much (~ 40,000)
- ⇨ they seemed to take it pretty well

He's coming into the ? (using order k = 2)
- ⇨ He's coming into the other (~ 40,000)
- ⇨ he's coming into the way

He's coming into the ? (using order k = 3)
- ⇨ he's coming into the house (~ 160,000)
- ⇨ he's coming into the white (~ 200,000)
- ⇨ he's coming into the room (~ 280,000)
- ⇨ he's coming into the house

ii) Generating two words from a context:

? was my ? house (using order k=2)
- ⇨ I was my life house (~ 40,000)
- ⇨ I was my god house

? was my ? house (using order k=3)
- ⇨ It was my friend house (~ 40,000)
- ⇨ It was my own house (~ 120,000)
- ⇨ It was my best house (~ 160,000)
- ⇨ It was my fault house

? coming into the ? (using order k = 2)
- ⇨ not coming into the other (~ 40,000)
- ⇨ not coming into the way

? coming into the ? (using order k = 3)
- ⇨ constantly coming into the house (~ 80,000)
- ⇨ is coming into the house (~ 160,000)
- ⇨ is coming into the white (~ 200,000)
- ⇨ is coming into the room (~ 280,000)
- ⇨ is coming into the house

But I think you let ? manipulate ? (using order k = 2)
- ⇨ But I think you let me manipulate and (~ 120,000)
- ⇨ But I think you let me manipulate you (~ 240,000)
- ⇨ But I think you let me manipulate me

But I think you let ? manipulate ? (using order k = 3)
- ⇨ But I think you let me manipulate and (~ 120,000)
- ⇨ But I think you let me manipulate you (~ 240,000)
- ⇨ But I think you let me manipulate me

Went ? romantic ? (using order k = 2)
- ⇨ Went to romantic atmosphere (~ 40,000)
- ⇨ Went to romantic dinner (~ 240,000)
- ⇨ Went to romantic night

Went ? romantic ? (using order k = 3)
- ⇨ Went a romantic atmosphere (~ 40,000)
- ⇨ Went a romantic dinner (~ 240,000)
- ⇨ Went a romantic night

I think she blames me ? ? moving out (using order k = 2)
- ⇨ I think she blames me to you moving out (~ 40,000)
- ⇨ I think she blames me to not moving out (~ 160,000)
- ⇨ I think she blames me to we're moving out

I think she blames me ? ? moving out (using order k = 3)
- ⇨ I think she blames me to you moving out (~ 40,000)
- ⇨ I think she blames me to the moving out

Generally, the recall memory of order $k = 2$, 3 was the most robust in completing one and two missing words. As expected, the syntax was not correct in many cases, but the sentences usually make sense ("It was my friend house") or the syntax is correct but the meaning is somewhat awkward ("It was my mistake house").

IV. Concluding Remarks

We demonstrate that recall memory of a language corpus can be made using the self-assemblying hypernetwork model. For a collection of drama dialogues we show that the hypernetwork can learn a linguistic memory and generate interesting new sentences by cued recall. Since the current experiments did not take into account any syntax or semantics in the model, the learned language model produced some incorrect sentences, but these were in many cases rather plausible and even give an impression of "creative" language generation.

From the cognitive learning and memory point of view, we note that the hypernetwork model has properties of both the localized and distributed representations. The representation is localized or modular since the whole network is a collection of discrete chunks of hyperedges. The representation is distributed since the ensemble of the chunks or hyperedges represent the whole memory. We also note that the hyperedges in the memory network consists of partially overlapping elements and appear in multiple copies. Thus, the hypernetwork memory model consists of "multiple representations of partially overlapping micromodules which are partially active simultaneously". In the brain, this kind of memory encoding scheme is found at several levels, notably in population codes at the cellular and circuit levels [3, 4, 8].

In the experiments, we have used about 300K sentences. By the machine learning standard, this size of training samples is large, but, considering the whole space of sentences, this is a rather small fraction. So, it would be interesting to see how the self-assemblying hypernetwork memory behaves if the corpus size increases, for example, to the whole dialogue collection of DVD films in a video shop.

Acknowledgment

We thank the members of the Multimodal Memory Game project in the Biointelligence Laboratory for their contributions to preparing the corpus and performing simulations.

REFERENCES

[1] Crowder, R. G., *Principles of Learning and Memory*, Lawrence Erlbaum, 1976.

[2] Feldman, J., *From Molecules to Metaphor: A Neural Theory of Language*, MIT Press, 2006.

[3] Fuster, J., *Cortex and Mind: Unifying Cognition*, Oxford University Press, 2003.

[4] Grillner, S. and Graybiel, A. M. (Eds.) *Microcircuits: The Interface between Neurons and Global Brain Function*, MIT Press, 2006.

[5] Kauffman, S., *Origins of Order: Self Organization and Selection in Evolution*, Oxford University Press, 1993.

[6] Langley, P., "Cognitive architectures and general intelligent systems," *AI Magazine*, pp. 33-44, Summer 2006.

[7] O'Reilly, R. C., "Biologically based computational models of high-level cognition," *Science*, 314: 91-94, October 2006.

[8] Tse, D. et al., "Schemas and memory consolidation," *Science* 316: 76-82, April 2007.

[9] Zhang, B.-T., "Cognitive learning and the multimodal memory game: toward human-level machine learning," *IEEE World Congress on Computational Intelligence (WCCI-2008)*, 2007 (submitted).

[10] Zhang, B.-T. and Kim, J.-K., "DNA hypernetworks for information storage and retrieval," *Proc. 2006 Int. Annual Meeting on DNA Computing (DNA12), LNCS* 4287: 298-307, 2006.