Theme-Rheme structure and Nakano-Amari-Kohonen-Hopfield networks

Yoichi Uetake

Faculty of Mathematics and Computer Science
Adam Mickiewicz University
ul. Uniwersytetu Poznańskiego 4, 61-614 Poznań
Poland
E-mail: uetake@amu.edu.pl

Abstract

First we describe the theme-rheme structure of linguistics. Then we discuss what we call the Nakano-Amari-Kohonen-Hopfield (neural) network as a biologically-inspired model of (dynamic) associative memory. This model connects existing concepts, whereas its learning process connects new items to existing knowledge. Thus it may explain a cognitive mechanism underlying memorizing, learning, creativity, etc. in various brain activities including language processing.

Key words: theme-rheme structure, functional linguistics, constructive mathematics, Nakano-Amari-Kohonen-Hopfield network, Transformer, attention mechanism

1. Introduction

This paper consists of two parts, which are loosely related. In §2, I will describe my COLING paper (Uetake, 1994a) titled "A formal representation of the thematic-rhematic structure of sentences based on a typed λ-calculus" written nearly 30 years ago. Although I left this subject just after completing the paper, switching my interest to number theory, the paper's aim was to formalize the theme-rheme (shortly, T-R) structure in functional linguistics using typed λ-calculus, following an analogue of the use of λ-calculus in Montague grammar. In functional linguistics, the theme (or topic, old information) of a sentence is what is being talked about, and the rheme (or comment, focus, new information) is what is being said about the theme. The T-R structure represents an information flow in a discourse. In that paper, I also suggested that concepts similar to the typed λ-calculus used in my T-R structure analysis are used in formalization of constructive mathematics, *i.e.* **r**-realizability, constructive type theory, *etc.* of Beeson (Beeson, 1985), Martin-Löf.

In §3, I will turn to neural networks and my paper (Uetake, in preparation) "Not only CNN, but also CMM: Nakano-Amari-Kohonen-Hopfield networks". I will explain mainly only the motive of this curious title, not going into the technical results of the paper. I will show that what I call the Nakano-Amari-Kohonen-Hopfield (shortly, NAKH) network is nothing but the one called the Hopfield network today. Not a few researchers (*e.g.* D. Amit, J. Schmidhuber) argue that the Hopfield network should be called the Amari-Hopfield network. I was a student of the courses taught by Amari and Nakano. Although Amari is well-known worldwide (*e.g.* a foreign member of the Polish Academy of Sciences), Nakono's visionary work seems, sadly, to have been overlooked or even forgotten. I will argue that (along with Kohonen) Nakano's associative model of memory stretches into today's language model. The language model like Transformer should be considered and renovated as a large scale NAKH network (possibly hybridized with backprop), which is physiologically plausible. Although the dynamic equation of NAKH networks was well-studied (*e.g.* chaos, Milnor attractor *etc.*), there seems still to be more room for research in the context of pure mathematics and/or brain modeling theory. Most of the theories and algorithms behind today's hype of (energy hungry) A.I. had been developed by the late 1990s.

Yoichi Uetake, *Theme-Rheme Structure and Nakano-Amari-Kohonen-Hopfield Networks*. In: Z. Vetulani, P. Paroubek, I. Kardava (Eds.), Human Language Technologies as a Challenge for Computer Science and Linguistics – 2025, pp. 81–85. Poznań: Adam Mickiewicz University Press, 2025. © Author, 2025. https://doi.org/10.14746/amup.9788323245414.15. Open Access chapter, distributed under the terms of the CC licence (BY-NC-ND, https://creativecommons.org/licenses/by-nc-nd/4.0/).

I will suggest that the notion of NAKH networks is worth revisiting and its generalization may give a framework for the next future breakthrough.

2. Theme-rheme structure of sentences and a typed λ-calculus

In functional linguistics of Prague School, the *theme* (or *topic*, *old information*) of a sentence is what is being talked about, and the *rheme* (or *comment*, *focus*, *new information*) is what is being said about the theme. The theme-rheme structure represents an information flow (or dynamics) in a discourse.

$$\begin{array}{lccc} \text{1st sentence:} & T & - & R \\ & & & \vdots \\ \text{2nd sentence:} & & T & - & R \\ \vdots & & & & \vdots \end{array}$$

In the English language this structure is realized with the particles "the" and "a". Here is an example.

"There is $\underbrace{\textbf{a}\ \text{book}}_{R}$ $\underbrace{\text{on}\ \textbf{the}\ \text{table}}_{T}$.
$\underbrace{\textbf{The}\ \text{book}}_{T}$ $\underbrace{\text{is interesting}}_{R}$."

The Japanese for this is realized with postpositions "wa" and "ga". See (Uetake, 1991–1992) and (Uetake, 1994b).

"$\underbrace{\text{Teburu-no ueni-}\textbf{wa}}_{T}$ $\underbrace{\text{hon-}\textbf{ga}}_{R}$ arimasu.
Sono $\underbrace{\text{hon-}\textbf{wa}}_{T}$ $\underbrace{\text{omoshiroi-desu}}_{R}$."

The aim of my COLING paper (Uetake, 1994a) was two fold. See also (Uetake, 1993).

(1) Formalize the theme-rheme structure in discourse using typed λ-calculus.
(2) Seek connection of (1) with formalisation of constructive mathematics.

In (Uetake, 1993) and (Uetake, 1994a), types $T, R, T \to R$ and $R \to R$ are introduced. For example, $\lambda f.f(\text{book})$ and $\lambda x.\text{interesting}(x)$ are assigned types T and $T \to R$, respectively. Then by β-reduction we get interesting(book) of type R, which vindicates the "correctness" of the discourse.

To give a flavor of what constructive mathematics is, here is an example of a famous non-constructive proof.

Theorem 2.1. There are irrational numbers a and b (possibly $a = b$) such that a^b is rational.

Proof. If $\sqrt{2}^{\sqrt{2}}$ is not rational, then put $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$. Invoke the law of excluded middle. (For this proof, "$\sqrt{2}^{\sqrt{2}}$ is either rational or irrational".) ■

Remark 2.2. Actually $\sqrt{2}^{\sqrt{2}}$ is transcendental thus irrational as a consequence of theorems proven by Gel'fond (1934) and Schneider (1934). Hence $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$.

3. Nakano-Amari-Kohonen-Hopfield networks and Transformer

3.1. The NAKH network

We now turn to neural networks and my paper in preparation "Not only CNN, but also CMM: Nakano-Amari-Kohonen-Hopfield networks" (Uetake, in preparation). I will explain mainly only the motive of this curious title, not going into the technical results of the paper. In 1989 LeCun developed CNN (Convolutional Neural Network) by plugging error back propagation (backprop) of Rumelhart-Hinton-Williams (1986) in Fukushima's Neocognitron (1980). Fukushima's aim was to model visual primary cortex activity discovered by Hubel and Wiesel in 1959.

$$\text{CNN} = \text{Neocognitron} + \text{Backprop}$$

This CNN was used in AlexNet by Krizhevsky-Sutskever-Hinton (2012) and AlphaGo by Google DeepMind (2016), which were the start of today's A.I. revolution.

Although backprop is a central tool of learning in today's A.I., F. Crick cast dout on its physiological plausibility as a brain activity already in 1980s (Crick, 1989). Given a recent huge success of backprop, this plausibility issue seems to be back in the neuroscience literiture. For example there is a paper (Lillicrap et al., 2020), but it hardly seems to be convincing for backprop's plausibility.

On the contrary, Nakano's Associatron and Kohonen's CMM (Correlated Matrix Memory)

naturally realize the Hebbian rule of learning and plausible. Thus "Not only CNN, but also CMM"! The so called Hopfield network was already proposed by Nakano and Amari. Therefore "Nakano-Amari-Kohonen-Hopfield networks"! In the reference of the Nobel Committee's Scientifc Report (The Nobel Committee, 2024), Nakano and Amari are cited as reference numbers 7 and 8. However, curiously their papers are not referenced (even numbers) in the main text. Kohonen's important paper is not referenced there. See (Schmidhuber, 2024) for controversy surrounding the 2024 Nobel Physics.

• Associatron (Nakano), Correlated Memory Model or **CMM (Kohonen)**

Nakano's Associatron (Nakano, 1972) and Kohonen's CMM (Kohonen, 1972) are a brain model of association, which assigns output to input. Let's consider the following simple example. For simplicity, each animal has only one sample picture.

$$\begin{array}{c}\text{dog picture} \mapsto \text{“dog”}\\ \text{cat picture} \mapsto \text{“cat”}\\ \text{hamster picture} \mapsto \text{“hamster”}\end{array}$$

We assign input vectors as follows.
q_1 =column vector of dog picture
q_2 =column vector of cat picture
q_3 =column vector of hamster picture
(Each element of $q_i = \pm 1$.)

Output vectors are assigned as follows.
v_1 =column vector of the answer "dog"
v_2 =column vector of the answer "cat"
v_3 =column vector of the answer "hamster"
(Each element of $v_i = \pm 1$.)

Let $m = \dim q_i$. Let W be a $\dim v_i \times \dim q_i$ matrix defined by

$$\begin{aligned} W &= \frac{1}{m}\sum_{j=1}^{3} v_j q_j^T \\ &= \frac{1}{m}(v_1 q_1^T + v_2 q_2^T + v_3 q_3^T). \end{aligned}$$

• Hebbian learning

If an active neuron "i" is cyclically stimulated by another neuron "j", it becomes more sensitive to that neuron's stimulation. (D. O. Hebb *Organization of Behavior* 1949)

In other words: Neurons that fire together wire together.

For example, if $\dim v_i = 2$ and $\dim q_i = 3$ then we have the following sistuation ("i" = "j" = "2").

$$v_1 q_1^T = \begin{array}{|c|}\hline \\ \hline \text{“i”} \\ \hline \end{array} \begin{array}{|c|c|c|}\hline & & \\ \hline & \text{“j”} & \\ \hline \end{array}$$

$$\text{“}i\text{”} \cdot \text{“}j\text{”} = \begin{cases} (+1)\cdot(+1) = +1 \\ (+1)\cdot(-1) = -1 \\ (-1)\cdot(+1) = -1 \\ (-1)\cdot(-1) = +1 \end{cases}$$

Remark 3.1. Actually one considers $\frac{\text{“}i\text{”}+1}{2}\cdot\frac{\text{“}j\text{”}+1}{2}$. If $\frac{\text{“}i\text{”}+1}{2}\cdot\frac{\text{“}j\text{”}+1}{2} = 1\cdot 1$, then the synapse between neurons "i" and "j" will be strengthened.

• Threshold function

For the column vector $x = [x_i]$ define the threshold function

$$x \mapsto y = [y_i] = \phi(x)$$

by

$$y_i = \begin{cases} -1 & \text{if } x_i < 0 \\ +1 & \text{if } x_i \geq 0. \end{cases}$$

Definition 3.2. *Associatron* (Nakano, 1972), *CMM* (Kohonen, 1972) are a mapping

$$x \mapsto y = f(x) = \phi(Wx).$$

Definition 3.3. If vectors x and y are of the same dimension, one can define a dynamical system (Nakano, 1972), (Amari, 1972), (Hopfield, 1982)

$$x_{t+1} = f(x_t).$$

We call it the *Nakano-Amari-Kohonen-Hopfield network* (shortly, the *NAKH network*).

Remark 3.4. $f(x)$ is nothing but the McCulloch-Pitts neuron model!

Theorem 3.5. *If the input vectors q_i's are orthogonal to each other, i.e. scalar product $q_i \bullet q_j (= q_i^T q_j) = 0$ for $i \neq j$, then*

$$f(q_i) = v_i.$$

Remark 3.6. If orthogonality holds, the threshold function ϕ is not necessary. When data are

not orthogonal to some extent and/or noisy, ϕ has a nice role for the Associatron and CMM to still work properly.

Proof. Let's check $f(q_2) = v_2$. The proof for $i = 1, 3$ is the same.

$$\begin{aligned} Wq_2 &= \frac{1}{m}\left(\sum_{j=1}^{3} \underbrace{v_j q_j^T}_{\text{matrix}}\right) q_2 \\ &= \frac{1}{m}\sum_{j=1}^{3} (\underbrace{v_j q_j^T}_{\text{matrix}}) q_2 \\ &= \frac{1}{m}\sum_{j=1}^{3} v_j (\underbrace{q_j^T q_2}_{\text{scalar}}) \\ &= \frac{1}{m}\sum_{j=1}^{3} v_j (q_j \bullet q_2) \\ &= v_2 \quad \blacksquare \end{aligned}$$

For the NAKH network with a square symmetric matrix $W = [w_{ij}]$, Hopfield defined the energy form $E(x)$ for $x = [x_i]$ by

$$E(x) = x^T W x = \sum_{i=1}^{\dim x} \sum_{j=1}^{\dim x} x_i w_{ij} x_j.$$

Theorem 3.7. (Hopfield, 1982) *Under some assumptions (e.g. asynchronicity), the energy function $E(x_t)$ in the NAKH network dynamics $x_{t+1} = f(x_t)$ is non-increasing, i.e. $E(x_{t+1}) \leq E(x_t)$.*

Remark 3.8. In the Boltzmann machine (shortly, BM) of Hinton-Sejnowski (Ackley et al., 1985), $e^{-\frac{E(x)}{k_B T}}$ is a Boltzmann (or Gibbs) distribution in the partition function. The NAKH network is a limit of the BM with an analogue of temperature $0 < T \to 0$ (crystallization).

$$\lim_{T \to 0} \text{BM}(T) \to \text{NAKH network}$$

On the contrary to his criticism for backprop (Crick, 1989), Crick praised the learning mechanism of the Boltzmann machine as a physiologically plausible model of brain activity (Crick, 1988).

3.2. Transformer (Vaswani et al., 2017)

We now explain that attention mechanism in Transformer is a natural generalization of the NAKH network. In Transformer we have three types of vectors q_i, k_i, v_i. We recover the NAKH network just by setting $k_i = q_i$. Our toy model Transformer described below translates the English sentence "i love you" to the French "je t' aime".

• **Transformer**

$$\begin{aligned} (q_1, q_2, q_3) &= \text{“}\underbrace{\text{i love you}}_{\text{tokens}}\text{”} \\ \to (y_1, y_2, y_3) &= \text{“}\underbrace{\text{je t' aime}}_{\text{tokens}}\text{”} \end{aligned}$$

• **Attention head** (trained by backprop, Hebbian learning, ...)
q_1 ="i", q_2 ="love", q_3 ="you" (query vectors)
k_1, k_2, k_3 (key vectors)
v_1 ="je", v_2 ="aime", v_3 ="t' " (value vectors)

• **Attention mechanism**
We expect q_2 ="love" $\to y_2$ ="t' ".

$$\begin{aligned} y_2 &= \text{Attention}(q_2, K, V) \\ &= \sum_{j=1}^{3} (\underbrace{v_j k_j^T}_{\text{matrix}}) q_2 \\ &= \sum_{j=1}^{3} v_j (\underbrace{k_j^T q_2}_{\text{scalar}}) \\ &= \sum_{j=1}^{3} v_j (k_j \bullet q_2) \\ &= v_1(k_1 \bullet q_2) + v_2(k_2 \bullet q_2) + v_3(k_3 \bullet q_2) \end{aligned}$$

Therefore if the key vectors are trained so that $k_3 \bullet q_2$ is large, then y_2 ="t' " with high probability. Compare the above calculation with the proof of Theorem 3.5.

Remark 3.9. One can use the threshold function instead of the softmax function as used in Transformer.

4. Conclusions

(1) The language model like Transformer should be considered and renovated as a large scale NAKH network, which is physiologically plausible. Possibly hybridized with backprop (see *e.g.*

Gupta et al., 2021).

(2) Although the dynamical system $x_{t+1} = f(x_t)$ was well-studied (*e.g.* chaos, Milnor attractor, *etc.*), there seems still to be more room for research in the context of pure mathematics and/or brain modeling theory.

(3) The NAKH network itself can be generalized (*e.g.* nonlinear scattering and S-matrix).

References

Ackley, D. H., Hinton, G. E. and Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. In: *Cognitive Science*, **9**, pp. 147 –169.

Amari, S.-I. (1972). Learning patterns and pattern sequences by self-organizing nets of threshold elements. In: *IEEE Trans. on Computers*, **C-21**, pp. 1197–1206.

Beeson, M. (1985). *Foundations of Constructive Mathematics*. Berlin: Springer.

Crick, F. (1988). Neural networks and REM sleep. In: *Bioscience Reports*, **8**, pp. 531–535.

Crick, F. (1989). The recent excitement about neural networks. In: *Nature*, **337**, pp. 129–132.

Gupta M., Ambikapathi, A. and Ramasamy S. (2021). HebbNet: A simplified Hebbian learning framework to do biologically plausible learning. In: *ICASSP 2020–2021, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 06–11 June 2021.*

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. In: *Proc. Nat. Acad. Sci. USA, Biophysics*, **79**, pp. 2554–2558.

Kohonen, T. (1972). Correlation matrix memories. In: *IEEE Trans. on Computers*, **C-21**, pp. 353–359.

Lillicrap T. P., Santoro, A., Marris, L., Akerman, C. J. and Hinton G. (2020). Backpropagation and the brain. In: *Nature Reviews Neuroscience*, **21**, pp. 335–346.

Nakano, K. (1972). Associatron — a model of associative memory. In: *IEEE Trans. on Systems, Man, and Cybernetics*, **SMC-2**, pp. 380–388.

The Nobel Committee for Physics (2024). *Scientifc Background to the Nobel Prize in Physics 2024*, October 8, 2024, pp. 12.

Schmidhuber, J. (2024). A Nobel Prize for Plagiarism. In: *Blog*, 12/7/2024 (v1), 7/31/25 (v2), 10/7/25 (v3).

Uetake, Y. (1991–1992). Analysis of the theme and rheme structure of a Japanese sentence. In: *Lingua Posnaniensis*, **XXXIV**, pp. 125–134.

Uetake, Y. (1993). Two formal representations of the thematic-rhematic structure of sentences. In: *Proc. of Pacific and Asian Conference on Formal and Computational Linguistics*, pp. 256–264.

Uetake, Y. (1994a). A formal representation of the thematic-rhematic structure of sentences based on a typed λ-calculus. In: *COLING '94: Proc. of the 15th International Conference on Computational Linguistics*, **2**, pp. 1105–1108.

Uetake, Y. (1994b). A thematic-rhematic structure of a Japanese sentence. In: *Proc. of the Warsaw Symposium on Japanese Studies 1994*, University of Warsaw (2000), pp. 432–445 (in Japanese).

Uetake, Y. (in preparation). Not only CNN, but also CMM: Nakano-Amari-Kohonen-Hopfield networks.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł. and Polosukhin, I. Attention is all you need. In: *31st Conference on Neural Information Processing Systems (NIPS 2017); Advances in Neural Information Processing Systems* **30**.