New Generation Computing (2020) 38:49–64
https://doi.org/10.1007/s00354-019-00081-z

Attentional Reinforcement Learning in the Brain

Hiroshi Yamakawa[1,2]

Received: 28 October 2019 / Accepted: 29 November 2019 / Published online: 2 January 2020

Abstract

Recently, attention mechanisms have significantly boosted the performance of natural language processing using deep learning. An attention mechanism can select the information to be used, such as by conducting a dictionary lookup; this information is then used, for example, to select the next utterance word in a sentence. In neuroscience, the basis of the function of sequentially selecting words is considered to be the cortico-basal ganglia-thalamocortical loop. Here, we first show that the attention mechanism used in deep learning corresponds to the mechanism in which the cerebral basal ganglia suppress thalamic relay cells in the brain. Next, we demonstrate that, in neuroscience, the output of the basal ganglia is associated with the action output in the actor of reinforcement learning. Based on these, we show that the aforementioned loop can be generalized as reinforcement learning that controls the transmission of the prediction signal so as to maximize the prediction reward. We call this attentional reinforcement learning (ARL). In ARL, the actor selects the information transmission route according to the attention, and the prediction signal changes according to the context detected by the information source of the route. Hence, ARL enables flexible action selection that depends on the situation, unlike traditional reinforcement learning, wherein the actor must directly select an action.

Keywords Natural language processing · Deep learning · Self-attention · Situatedness · Thalamocortical loop · Basal ganglia · Actor–critic model · Predictive coding · Brain-inspired refactoring

Introduction

Natural language is data configured as a sequence of letters and words. Since the development of deep learning, natural language processing technology using recurrent neural networks, suitable for handling sequences, has become

✉ Hiroshi Yamakawa
ymkw@wba-initiative.org
https://wba-initiative.org/

[1] The Whole Brain Architecture Initiative, Tokyo, Japan

[2] The University of Tokyo, Tokyo, Japan

mainstream. However, the heightened effectiveness of using attention in such processing methods [1] was recognized several years ago; after that, a group of technologies was developed, termed transformers using attention [2, 3]. In a specific range of tasks, transformers that use attention demonstrate performance beyond that of humans. This technology has a function that calls attention to the part of the language-processing method by a dictionary-like mechanism. For example, with the task of translating the German "Ich gehe zur Schule" into the English "I go to school" as an example, the mechanism of this dictionary-like attention is simplified as follows. First, associative relations such as {"I", "go", "to", "school"} corresponding to the input German words are activated. In English, there are grammatical rules such as "the subject is placed at the beginning of the sentence" and "the verb follows the subject." For this reason, the verb, "go", is then selected immediately after the subject, "I", is spoken.

In neuroscience, the foundation for speech function is thought to be the cortico-basal ganglia-thalamo-cortical loop (CBGTC loop). The "response-release semantic feedback (RRSF) model" has been proposed as an explanation of its function [4–6]. The basal ganglia appearing in this loop are known to achieve reinforcement learning as an actor–critic model in the brain [7–10].

Recently, empirical results that support the predictive coding theory [11, 12], have increased [13]. The predictive coding theory hypothesizes that prediction signals propagate a recognition hierarchy from the top-down. It has also been suggested that basal ganglia may select the prediction signal flowing through the thalamus [14]. These are used as fundamental assumption for subsequent discussion.

In Sect. 2, we first demonstrate that the attention mechanism used in deep learning is associated with the CBGTC loop. In Sect. 3, attentional reinforcement learning (ARL), which controls the transmission of the prediction signal to maximize the prediction reward, is proposed as a computational model integrating attention mechanisms and reinforcement learning for natural language processing. Then, the actor–critic model-based ARL, which has high affinity with brain architecture, is examined. In Sect. 4, we describe how the process of promoting the merging of multiple programs because of constraints, referring to the brain architecture discussed here, can be regarded as a brain-inspired AI development method. We call this process brain-inspired refactoring and assert that there are more opportunities to develop and apply it in the future.

Dictionary-Like Attention in the Brain

In this section, we focus on the mechanism used to select next words with appropriate timing for sentence generation in natural language processing. First, we explain that in deep learning, this word selection is achieved via an attention mechanism. Next, we state that this is explained by the RRSF model based on the CBGTC loop in the brain. Additionally, we show that the attention mechanism used in deep learning is associated with this neural loop.

Attention in Deep Learning for Natural Language Processing

As a recent development in language processing using deep learning, the dictionary-like mechanism achieved via a neural network model is called attention. The dictionary mechanism is an abstract data type that is common in computer programming. Specifically, this mechanism finds a key, K, that matches the input query, Q, and outputs an arbitrary memory, V, corresponding to the key, K. This output can be regarded as an array that can use data types other than scalar numeric values as subscripts. A dictionary is also called an associative array, associative list, associative container, hash, map, etc. [15].

Even using a neural network, an operation similar to the above-described dictionary method can be implemented. In that case, an output value, O, is obtained for the input query, Q. Inside, key, K, and value, V, are implemented as numeric vectors. This mechanism is expressed as follows:

$$O(Q, K, V) = \mathrm{softmax}(QK^{\mathrm{T}})V. \tag{1}$$

Here, attention is a mechanism used to increase the weight of the value, V, corresponding to the key, K, similar to the query, Q. Therefore, the attention-calculating part, A, is expressed, as follows:

$$A(Q, K) = QK^{\mathrm{T}}. \tag{2}$$

Next, the operation of gating V using this attention, A, is expressed by the following equation.

$$O(V, A) = \mathrm{softmax}(A)V. \tag{3}$$

In a system built with deep learning, interpreting the meaning explicitly is difficult. However, it is possible to conceptually examine the process in which a natural language processing system using deep learning performs a German-to-English translation task using attention. The system will have acquired at least two pieces of knowledge from large amounts of data in advance. The first is grammatical rules as procedural knowledge, such as “subject followed by the verb”, acquired from a large volume of English corpora. In fact, it is suggested that the Bidirectional Encoder Representations from Transformers (BERT) contains knowledge related to lexical categories [16]. The second type of information acquired in advance is the various German–English associative-word relationships as declarative knowledge, learned from a large translation corpus. The translation task involves a generation of English words in an appropriate order (e.g., “I go to school”) corresponding to a given German sentence (e.g., “Ich gehe zur Schule”). The behavior in a simple case is shown in Fig. 1. First, in the system encoder, {{“Ich”, “I”}, {“gehe”, “go”}, {“zur”, “to”}, {“Schule”, “school”}} is used to interpret the German sentence, and in the expression V with the output word candidate, a context is formed in which English expressions such as {“I”, “go”, “to”, “school”} are activated. At the first step of English sentence generation, the “subject” is formed as attention A by the grammatical rule that “the subject is placed at the beginning of a sentence”. Then, a possible candidate word “I” is selected depending on attention A. In the next step, by entering

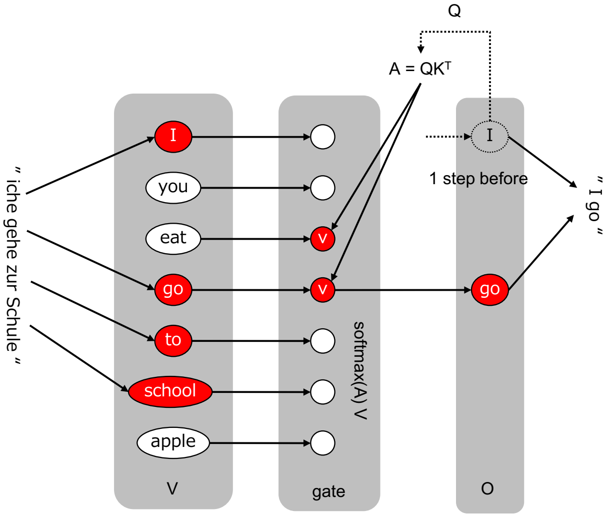

Fig. 1 Conceptual process of German–English translation: After outputting the word “I”, attention is given to the lexical category of the verb (v) according to grammar rule *K*, and “go” is selected from the activated word beyond the gate. Later in this paper, it will be shown how the generation of attention *A* from query *Q* corresponds to the actor part of the basal ganglia, and the role of the gate to the thalamic relay cells

a query that the decoded “I” is the subject in the grammatical rule “the verb follows the subject,”attention *A* to the verb is formed. Then, a possible candidate word “go” is selected depending on attention *A*. After this, the same procedure is repeated until the end of the sentence. The above process is summarized as the repetition of the following two operations alternately: first, attention *A* of the lexical category is formed according to the grammatical rules (accumulated as key–value relationships) from the previous utterance as the query; second, a word that matches the attention *A* is selected from the candidate set *V* that matches the context.

Natural Language Processing in the Brain

The mechanism of human natural language processing has been researched since the nineteenth century via the study of aphasia with various symptoms; from this, it seems as though natural language processing functions are achieved in networks localized in multiple neocortical areas of the left hemisphere. The Lichtheim brain anatomy model, which was particularly influential, linked the posterior superior temporal gyrus (Wernicke area) to auditory language understanding, and the inferior frontal gyrus (Broca area) to speech function [17]. Geschwind then incorporated a bundle of nerve fibers from the Wernicke field to the Broca field (an arcuate bundle) into the model, positioning the language’s repetitive function there, and induced visual input. It is known that the ventral and inferolateral aspects of the anterior temporal lobe play an essential role in language understanding; this is supported by neuroscientific research [18, 19] and by case studies such as those investigating semantic dementia in patients with brain injuries [20]. Following these developments in proposed models of natural language processing, a model called Lichtheim 2 [21], which consists of a double circuit of the dorsal and ventral pathways of the left hemisphere, has been proposed.

However, since about 1950, it has been considered that the mechanism of human natural language processing should include the thalamus and basal ganglia as opposed to the neocortical view, with the following remarks by Lieberman [22]:

> The traditional theory equating the brain bases of language with Broca's and Wernicke's neocortical areas is wrong. Neural circuits linking activity in anatomically segregated populations of neurons in subcortical structures and the neocortex throughout the human brain regulate complex behaviors such as walking, talking, and comprehending the meaning of sentences. (Lieberman, 2002)

The background associated with these remarks by Lieberman involved clinical studies of many aphasias caused by the thalamus [23–26]. However, in the case of thalamus aphasia, the symptoms are more complex than neocortical aphasia, and it is difficult to identify the correspondence with the damaged site in the brain (due to bleeding, etc.). The RRSF model [4–6] is known as a computational model related to utterances considering the subcortical function. In the RRSF model, the basal ganglia controls the signal that flows through the thalamocortical loop, such that a word is selected according to the context of the text, and a phoneme or character is selected according to the word selected.

The following points (A)–(C) explain the following, respectively: "the selective engagement model" that models the function of the thalamocortical loop, the function of the basal ganglia, and "the RRSF model" based on the CBGTC loop.

(A) Selective engagement model of the thalamocortical loop:

The thalamocortical loop circuit has a hub-like connection structure that receives excitatory neural projection from a wide range, namely from the cerebral neocortex to the thalamus, and simultaneously projects from the thalamus to various areas of the neocortex. In other words, it has an anatomical structure suitable for relaying information between areas of the neocortex. The thalamus contains relay cells for this function. The function of the thalamocortical loop is considered in light of the "selective engagement model" hypothesis, which posits that thalamic cells monitor the activity state of widely dispersed neocortical areas and control their functional connections through connections with the hypothalamus [4, 27]. This hypothesis is supported by the fact that the thalamic relay nucleus plays an important role in changing the dynamics of cortical processing by setting different frequency-synchronous vibration patterns [28–36]. Predictive coding theory [11, 12] states that the prediction signal propagates from the higher-level related region to the lower-level sensory cortex, in a top-down manner. Since brain organs that can flexibly exchange prediction signals between various neocortical regions can be assumed only in the thalamus, the prediction signals are thought to pass through the thalamus.

(B) Function of basal ganglia:

Basal ganglia receive information by projection from the fifth layer in a broad neocortical area. Basal ganglia have been modeled as reinforcement learning that selects actions from a wide range of external prediction signals [7–10]. Specifically, basal ganglia have long been considered in actor–critic-type reinforcement learning. A circuit in which the striatum (patch) receives a projection from the neocortex and projects it to the substantia nigra pars compacta (SNc), and a circuit in which dopamine projection from the SNc feeds back to the projection from the neocortex to the striatum plays the role of a critic. Here, dopamine projection is considered a predictive reward error and controls both actor and critic learning. In the circuit corresponding to the actor, the striatum (matrix) receives an excitatory projection from the neocortex; this part of learning is controlled by dopamine. The projection from the striatum (matrix) to the internal globus pallidus (GPi)/substantia nigra pars reticulata (SNr) is inhibitory. Furthermore, projection to a thalamic relay cell that transmits a prediction signal from the GPi/SNr is also inhibitory. In this manner, information transmitted between neocortical areas through the thalamic relay cells is gated by basal ganglia control. On the actor side of the basal ganglia, the projection path from the striatum (matrix) to the GPi/SNr has a direct path as well as an indirect path. Additionally, there is a hyper-direct path projected from the neocortex. That mechanism coordinates the timing of deactivating the GPi/SNr and releasing thalamic relay-cell transmission. This is consistent with the lexical selection model [37] related to language generation, wherein the basal ganglia are regarded as the machine that aligns word-related input with ongoing language plans.

(C) Response-release semantic feedback model:

In the above thalamocortical loop, there is a CBGTC loop as a pathway for gating the output of the basal ganglia to relay the thalamocortical loop. In basal ganglia, there are parallel loops related to the control of movement and thought; these are mainly divided into a skeletomotor loop, oculomotor loop, prefrontal-cortex loop, and limbic loop [38]. The RRSF model is known as a computational model based on the hypothesis that the CBGTC loop is responsible for language processing [4–6]. For language processing, the left hemisphere language-related areas (Broca area, Wernicke area, etc.) are mainly used in the above loop. Klostermann [27] stated:

> In the specific context of language processing, the “Response-Release Semantic Feedback model” claims thalamic and [basal ganglia (BG)] functions in language production [4–6, 39]. As in the Selective Engagement model, thalamic nuclei are posited to control the interaction between fronto-opercular and temporo-cortical cortices for the integration of lexico-syntactic with seman-

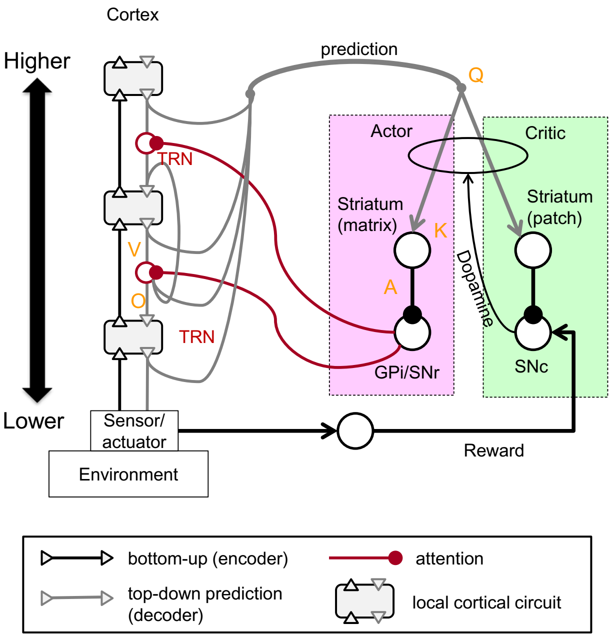

Fig. 2 Schematic relationship between the cortex, thalamus, and basal ganglia: the neocortex is composed of multiple areas and has a hierarchical structure that extends from the lower-order areas close to sensors/actuators in contact with the environment to the abstract higher-order areas. Encoding is performed in the flow up the hierarchy, in a bottom-up manner, and decoding is performed in the flow in which the prediction signal descends the hierarchy, in a top-down manner. The striatum of basal ganglia receives the prediction signals from the neocortex. In the basal ganglia, the actor part in the actor–critic model is a path from the striatum (matrix) to the GPi/SNr, and the critic part corresponds to a loop of the striatum (patch) and the SNc. The SNc receives a reward calculated based on inputs from sensors. The dopamine output by the SNc is a prediction reward error and controls the learning of actors and critics. The output of the GPi/SNr selects the signal by disinhibiting the thalamic relay cell (TRN) that mediates the prediction signal

tic information. The resulting signal is further passed on to the BG which are thought to coordinate the release of the provided language plan into speech.

The above model is consistent with the “declaration/procedure model” [40] that executes the combination of declarative knowledge accumulated in the neocortex through the procedural function of applying grammatical rules.

Dictionary-Like Attention Mechanism in the Brain

In this section, we build a hypothesis regarding the neural foundation of the dictionary-like attention used in deep learning for language processing. In the proposed hypothesis (Fig. 2), the output of the striatum (matrix), which corresponds to the actor part of the basal ganglia, is assigned to the output of attention A. Next, by controlling thalamic relay cells through disinhibition based on attention A, the outputs of value V from the cortical areas are gated and action O is produced.

“The mapping hypothesis, between the dictionary-like attention and brain circuit, involves the following:”

- The cortex sends the previous word as a query (potentially including lexical category) to the striatum (matrix).
- The striatum (matrix) calculates attention A according to grammatical rules acquired from long-term experience.
- Learning, such as the above-mentioned grammar rules, is based on the fact that the plasticity of projections from the cortex to the striatum is modulated by dopamine projection from the SNc.
- The activation of attention A disinhibits thalamic relay cells through the GPi/SNr, and part of value V is transmitted as the output value O.

The following explains the rationale that supports the validity of this hypothesis. First, disinhibition mechanisms, which release the suppression of the relay, achieve a pure gating function needed for the dictionary-like mechanism. These mechanisms are positioned at thalamic relay cells to control transmissions between areas of the neocortex.

Second, timing is considered. The timing at which attention A acts, which determines the next word in deep learning, corresponds to the timing at which the basal ganglia release speech output in the RRSF model. Therefore, it is reasonable to associate the GPi/SNr output with attention A.

Third, basal ganglia are considered to comprise a circuit that exhibits procedural functions, wherein they apply empirically acquired grammatical rules in the declaration/procedure model [40]. It is, thus, reasonable to assign the function of generating attention A to basal ganglia, as they contain the grammatical rules.

There are three possibilities in regard to which part of the neural circuit the softmax function (Eq. 3) is associated. The first possibility, which directly corresponds to Eq. 3, is that softmax function stay within the striatum (matrix), and these activities are controlled sparsely by means of interneurons. The second possibility is that the disinhibition mechanism from the striatum (matrix) through the GPi/SNr to the thalamus has an unknown effect. Third, it may be brought about by the regulation of thalamic relay-cell output [39] as a function of the thalamic reticular nucleus (RTN) effectively, which comprises inhibitory neurons that surround the thalamus, as shown in Fig. 3.

The feature of this hypothesis is that the basal ganglia output lexical category, as attention comprised of a set of words, is gated. This is different from selecting a specific word. With this feature (shown in Fig. 1), it is possible to achieve the function of further narrowing down a candidate set of next words weighted separately according to context.

The prediction signal in the brain is often transmitted top-down (see Fig. 2). However, in natural language processing, as shown in Fig. 1, attention is generated and used in the same representation hierarchy, such as in the selection of words. In deep learning research, this attention generation and utilization corresponds to a mechanism called self-attention.

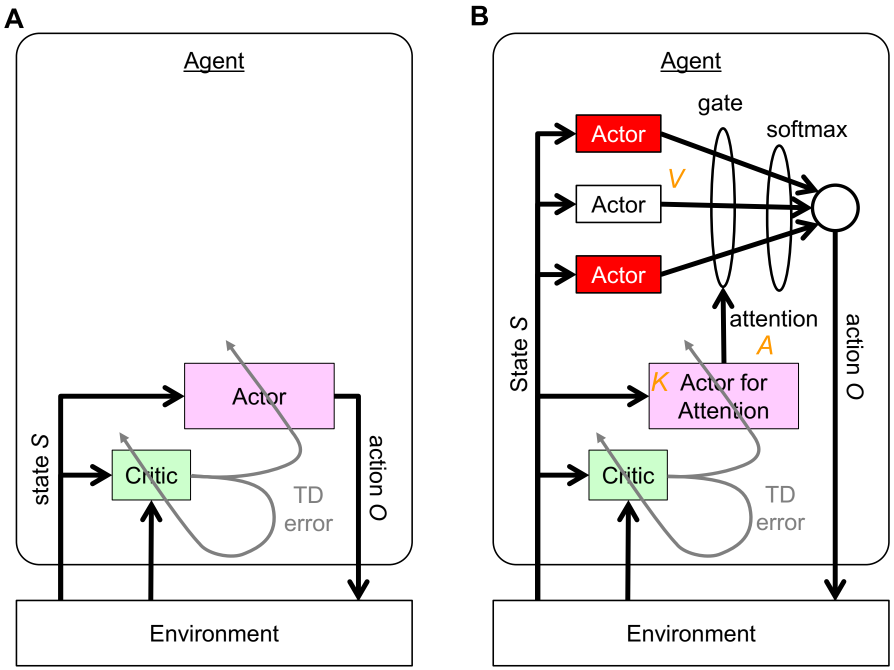

Fig. 3 Overview of attentional reinforcement learning: **a** Conventional actor–critic model. **b** ARL model based on actor–critic. The TD error output by the critic is used for learning the actor that outputs attention. Value *V* is the whole output of candidate values by multiple actors; *V* is gated by attention *A* and output to the environment as action *O*

Attentional Reinforcement Learning (ARL)

Defining Attentional Reinforcement Learning (ARL)

A mechanism for giving attention to a lexical category, i.e., a set of words as actions, can be generalized as a mechanism for gating a specific information path. Contrastingly, as previously noted in Sect. 2.2, the mechanism of reinforcement learning repeatedly appears as a loop in which various neocortical areas include the basal ganglia and the thalamus. Therefore, the mechanism for gating the flow of information between areas of the neocortex with an attention signal is not restricted to language processing and will be widely diverted. With regard to the aspect of this attention-giving mechanism, attentional reinforcement learning (ARL) is defined as follows:

> ARL is a kind of reinforcement learning. ARL maximizes the expected reward by modulating the flow of prediction signals.

In this case, the "action output" part in reinforcement learning is limited to the form of the "predictive signal gating." We can say that gating is one of the types of action outputs of reinforcement learning. Agents with ARL will be able to generate context-dependent output through gating. This is different from typical reinforcement learning that outputs a specific action (or action selection probability).

In the act of uttering, the information used by the actor to determine the next lexical category has been simplified as the previous word (i.e., the query Q). However, even in this case, the words before the most recent should be taken into consideration, and generally, the action is determined by factors from various external environments. It is not rare that the influence on the next action from the external environment becomes larger than that of its own actions. Metaphorically, when a person looks at a door with a knob, they intuitively want to turn it. In other words, candidate actions appear in the relationship between a person and an object. Gibson called this “affordance,” according to which action is embedded in the environment [41]. However, even in this example, it is often assumed that the act of moving in front of the door is performed before the door is recognized. In many cases, the previous action influences the next action.

Attentional Actor–Critic Model (AAC Model)

In this section, we examine ARL based on the actor–critic model, which is compatible with brain architecture, heretofore referred to as the attentional actor–critic (AAC) model. In the following, we explain the AAC model in terms of its implementation in the brain. As shown in Fig. 3, the actor–critic model is a function that calculates rewards to be acquired in the future from the state S and learning from data based on the temporal difference error (TD error) calculated by the model itself. The actor learns a function that performs a mapping from state S to action O using TD error. However, from the discussion in the latter half of Sect. 3.1, the information that the actor should use is generally regarded as the state S. For state S, sensor inputs may be used directly, or information after preprocessing (for example, in deep learning) may be used. The AAC model assumes that there are multiple actors that generate specific actions for a state. These actors generate concrete action candidates V from state S and may be constructed by human design or by learning from data. The actor for attention in the actor–critic model outputs attention signal A, as shown in the following equation.

$$A = SK^{\mathrm{T}} = \pi(S). \tag{4}$$

Here, Q (in Eq. 2) was replaced with S. Furthermore, since the key K is an internal parameter determined through learning, it is considered to be incorporated in the policy function π. This function corresponds to the part in which basal ganglia output the attention A according to state S in the brain.

Next, using attention A, gating the candidate value V (Eq. 3) is used as it is. This function corresponds to thalamic relay cells in the brain. In Fig. 3, an actor that outputs V is depicted as a plurality of modules for intuitive understanding. However, it should be noted that when the system is constructed with a neural network, each actor outputs from multiple neurons. Below is a list of descriptions for operating the AAC model. Although the definitive descriptions are given here, they may be extended to a probabilistic description; i.e., the AAC model described below is one example of the ARL model.

- Critic: same as standard actor–critic model
- $A = \pi(S)$
- $O = \mathrm{softmax}(A)V$

Effective Application Scenarios for ARL

As previously described, there are two operations for selecting the next word in machine translation. First, grammar rules are applied to the word that was spoken immediately before, and attention for the lexical category of the next word is generated. Next, from the possible word candidates in the context obtained from the input source text, a word that matches the generated attention is selected and output.

A generalization of these operation is the ARL operation. In other words, the actor function is applied to the immediately preceding state to generate attention as an output category. The advantage of this model is that the output of reinforcement learning can be made dependent on the situation.

Considering the characteristics of ARL, ARL is effective if applied to scenarios that require both rapid and timely responses as well as flexibility, depending on the context. If the previous information does not affect the decision or if the response does not require timeliness, the immediacy provided by ARL is less effective because agents can make time-consuming inferences. Even if the response to the immediately preceding information needs to be rapid, regular reinforcement learning can be used if the response is not flexible, whereby the characteristics of ARL are not sufficiently utilized.

In terms of daily life, many tasks require both real-time operations and flexibility, and are, thus, suitable for ARL.

As a first example, the process for a singer to achieve a praiseworthy performance can be considered to consist of two operations. First, deliberate attention to volume, voice type, tempo, pitch, etc. is generated based on singing skills. Next, utterances that match the attention are selected and output in the context of lyrics.

As a second example, the process for hitting an effective shot in tennis can also be considered to consist of two operations. First, intentional attention about the possible hitting method, speed, spin, and course is selected by considering hitting an opponent's ball according to his or her technique. Next, the action is selected that matches the attention from the possible racket swing movements within the context, such as the position and posture of the player and the opponent, and is output.

As a third example, the process of safely operating a car is also composed of two operations. First, based on the driver's skill, intentional attention is selected for the direction of travel, speed, etc. from the current surrounding conditions. Next, an operation that matches the attention is selected, and an output is generated from the possible steering, acceleration, and deceleration operations in view of the driving context, such as road surface conditions and vehicle performance.

Discussion

Here, prediction signals were assumed to pass through thalamic relay cells, and gated by attention signals outputted by the basal ganglia. Indeed, there are many reviews regarding the cognitive importance of predictions in the brain [42–49]. It is also assumed that the top-down information gated by ARL is the prediction signals. The concept of reinforcement learning, which selects the desired behavior based on rewards, is reasonable. In contrast, selecting a prediction based on rewards in the ARL looks to do something different. However, this gap is solved by the emulation theory described below.

Emulation theory has important implications for predictive coding. The theory assumes the cognitive mechanisms in the brain with multiple internal representations (emulators) that predict specific actions and the resulting expected sensor information (including the effect copy) [47]. By acting based on the emulation theory, the agent can control the sensor information that is expected to be obtained in the future within the range of freedom of action selection. This idea has been studied in some fields, such as in motion control [50, 51].

Similarly, the generative character of perception using neural architecture can also resolve the gap between cognition and action generation [52]. In the similar idea of a "generalized state," in which prediction representations are always placed in parallel with state representations, the merit, that preparing a representation only for action output is unnecessary, is emphasized [53, 54]. The reason behind this is that it is difficult to acquire representations for action from reward signals and/or teacher signals with little information.

In general, an intelligent agent generates an action to achieve desired situations by following the orientation, such as a value function, reward function, goals, or purposes. For agents based on emulation theory, this desired orientation is achieved by controlling the predictions so as to be useful to the agent. In ARL, the desired prediction is selected by a mechanism that directs attention to a prediction signal with high future value. In the brain, it is achieved by a mechanism that selects information transmission in the thalamus by the output of basal ganglia.

Conclusion

Here, the following hypothesis was outlined and explained: the dictionary-like attention mechanism used for language processing using deep learning is an attention mechanism that controls information transmission between one cortex area and another cortex area. Specifically, it was shown that the basal ganglia output an attention signal and control the thalamic relay cells as a gate. In general, the basal ganglia operate as reinforcement learning. However, the output act is attention, such that a flexible action corresponding to the context becomes the output. Here, reinforcement learning that outputs attention in this way is called attentional reinforcement learning (ARL). This mechanism is particularly effective when it is necessary to have a response that is both timely, according to the immediately preceding information, and flexible, according to the context. In addition to natural language utterances, this ability is required for various action skills. In other words, ARL is a model that provides some degree of situatedness to the nature of reinforcement learning and can react quickly.

Since ARL has a high affinity with brain mechanisms, it is promising for use in designing brain-inspired AI (including natural language processing models), together with the reference model of the neocortical local circuit [55]. Practically speaking, a mechanism that adjusts the timing at which the basal ganglia GPi/SNr release the transmission of thalamic relay cells in the brain may help to achieve more human-like responsiveness, especially if such a mechanism can be incorporated into a real-time dialog system. Top-down action generation is performed hierarchically in the brain, and the prediction of desirable sensor signals emerges as the actuator in contact with the environment. Therefore, ARL-based hierarchical reinforcement learning, that functions as brain architecture, can be built. However, it can be said that language processing is characterized by a self-attention mechanism, which gives attention to information transmission in the same hierarchy. Feedback to the same hierarchical level within the animal cognitive architecture may have been enabled in humans only through evolutionary mutations.

Lastly, the idea of ARL was built on the nonscientific fact that both the deep learning models of attention mechanisms and reinforcement learning are functionally realized in the basal ganglia. In other words, model merging was facilitated by the constraint that different computational models were implemented on the same brain organ. Thus, in the development of brain-inspired AI, a method that encourages the integration of multiple programs using the restriction of referring to the brain as an existing unique model is called brain-inspired refactoring. This idea was conceived through AI development activities of the Whole Brain Architecture Initiative.[1] In the development of an integrated AI system that references the brain, opportunities to use such a brain-inspired refactoring method are likely to increase, while keeping pace with the development of machine learning and neuroscience.

Acknowledgements We would like to thank Naoya Arakawa, Shinichi Asakawa, Satoshi Negishi, and Masahiro Suzuki for their discussion and advice in writing this paper.

[1] https://wba-initiative.org/en/

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł.U., Polosukhin, I.: Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008. Curran Associates Inc, New York (2017)
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding (2018)
3. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8) (2019)
4. Crosson, B.: Subcortical functions in language: a working model. Brain Lang. **25**(2), 257–292 (1985)
5. Crosson, B.A.: Subcortical Functions in Language and Memory. Guilford Press, New York (1992)
6. Murdoch, B.E.: Subcortical brain mechanisms in speech and language. Folia Phoniatr. Logop. **53**(5), 233–251 (2001)
7. Bogacz, R., Larsen, T.: Integration of reinforcement learning and optimal decision-making theories of the basal ganglia. Neural. Comput. **23**(4), 817–851 (2011)
8. Gillies, A., Arbuthnott, G.: Computational models of the basal ganglia. Mov. Disord. **15**(5), 762–770 (2000)
9. Khamassi, M., Lachèze, L., Girard, B., Berthoz, A., Guillot, A.: Actor–Critic models of reinforcement learning in the basal ganglia: from natural to artificial rats. Adapt. Behav. **13**, 131–148 (2005)
10. Ren, H., Liu, C., Shi, T.: A computational model of cognitive development for the motor skill learning from curiosity. Biol. Inspir. Cogn. Archit. **25**, 101–106 (2018)
11. Huang, Y., Rao, R.P.N.: Predictive coding. Wiley Interdiscip. Rev. Cogn. Sci. **2**(5), 580–593 (2011)
12. Rao, R.P., Ballard, D.H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. **2**(1), 79–87 (1999)
13. Egner, T., Summerfield, C.: Grounding predictive coding models in empirical neuroscience research. Behav. Brain Sci. **36**(3), 210–211 (2013)
14. Colder, B.: The basal ganglia select the expected sensory input used for predictive coding. Front. Comput. Neurosci. **9**, 119 (2015)
15. Hash tables and associative arrays. In: Mehlhorn, K., Sanders, P. (eds.) Algorithms and Data Structures: The Basic Toolbox, pp. 81–98. Springer, Berlin (2008)
16. Clark, K., Khandelwal, U., Levy, O., Manning, C.D.: What does BERT look at? An analysis of BERT's attention (2019)
17. Lichtheim, L.: On aphasia. Brain **7**, 433–484 (1885)
18. Binney, R.J., Embleton, K.V., Jefferies, E., Parker, G.J.M., Lambon Ralph, M.A.: The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of Distortion-Corrected fMRI, rTMS, and semantic dementia. Cereb. Cortex **20**, 2728–2739 (2010)
19. Pobric, G., Jefferies, E., Ralph, M.A.L.: Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. Proc. Natl. Acad. Sci. USA **104**(50), 20137–20141 (2007)
20. Patterson, K., Nestor, P.J., Rogers, T.T.: Where do you know what you know? The representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. **8**(12), 976–987 (2007)
21. Ueno, T., Saito, S., Rogers, T.T., Lambon Ralph, M.A.: Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. Neuron **72**(2), 385–396 (2011)
22. Lieberman, P.: On the nature and evolution of the neural bases of human language. Am. J. Phys. Anthropol. Suppl **35**, 36–62 (2002)
23. Assaf, M., Calhoun, V.D., Kuzu, C.H., Kraut, M.A., Rivkin, P.R., Hart Jr., J., Pearlson, G.D.: Neural correlates of the object-recall process in semantic memory. Psychiatry Res. **147**(2–3), 115–126 (2006)
24. Lam, Y.W., Sherman, S.M.: Functional organization of the somatosensory cortical layer 6 feedback to the thalamus. Cereb. Cortex **20**(1), 13–24 (2010)
25. Nadeau, S.E., Crosson, B.: Subcortical aphasia. Brain Lang. **58**(3), 355–402 (1997). (discussion 418–23)

26. Sherman, S.M., Guillery, R.W.: Distinct functions for direct and transthalamic corticocortical connections. J. Neurophysiol. **106**(3), 1068–1077 (2011)
27. Klostermann, F., Krugel, L.K., Ehlen, F.: Functional roles of the thalamus for language capacities. Front. Syst. Neurosci. **7**, 32 (2013)
28. Bal, T., Debay, D., Destexhe, A.: Cortical feedback controls the frequency and synchrony of oscillations in the visual thalamus. J. Neurosci. **20**, 7478–7488 (2000)
29. Destexhe, A., Contreras, D., Steriade, M.: Mechanisms underlying the synchronizing action of corticothalamic feedback through inhibition of thalamic relay cells. J. Neurophysiol. **79**, 999–1016 (1998)
30. Destexhe, A., Contreras, D., Steriade, M.: Cortically-induced coherence of a thalamic-generated oscillation. Neuroscience **92**(2), 427–443 (1999)
31. Haber, S.N.: Integrating cognition and motivation into the basal ganglia pathways of action. In: Bédard, M.A., Agid, Y., Chouinard, S., Fahn, S., Korczyn, A.D., Lespérance, P. (eds.) Mental and Behavioral Dysfunction in Movement Disorders, pp. 35–50. Humana Press, Totowa (2003)
32. Haber, S.N.: The primate basal ganglia: parallel and integrative networks. J. Chem. Neuroanat. **26**(4), 317–330 (2003)
33. Haber, S.N., Groenewegen, H.J., Grove, E.A., Nauta, W.J.: Efferent connections of the ventral pallidum: evidence of a dual striato pallidofugal pathway. J. Comp. Neurol. **235**(3), 322–335 (1985)
34. Jones, E.G.: Correlation and revised nomenclature of ventral nuclei in the thalamus of human and monkey. Stereotact. Funct. Neurosurg. **54–55**, 1–20 (1990)
35. Jones, E.G.: The Thalamus of Primates. Elsevier, Amsterdam (1998)
36. Steriade, M.: Coherent oscillations and short-term plasticity in corticothalamic networks. Trends Neurosci. **22**, 337–345 (1999)
37. Alexander, G.E.: Basal ganglia-thalamocortical circuits: their role in control of movements. J. Clin. Neurophysiol. **11**(4), 420–431 (1994)
38. Macpherson, T., Hikida, T.: Role of basal ganglia neurocircuitry in the pathology of psychiatric disorders. Psychiatry Clin. Neurosci. **73**(6), 289–301 (2019)
39. Murdoch, B.E.: Speech and Language Disorders Associated with Subcortical Pathology. Wiley, Hoboken (2009)
40. Ullman, M.T.: The declarative/procedural model of lexicon and grammar. J. Psycholinguist. Res. **30**(1), 37–69 (2001)
41. Gibson, J.J.: The ecological approach to the visual perception of pictures. Leonardo **11**, 227–235 (1978)
42. Bar, M.: The proactive brain: using analogies and associations to generate predictions. Trends Cogn. Sci. **11**(7), 280–289 (2007)
43. Bubic, A., von Cramon, D.Y., Schubotz, R.I.: Prediction, cognition and the brain. Front. Hum. Neurosci. **4**, 25 (2010)
44. Clark, A.: Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. **36**, 181–204 (2013)
45. Colder, B.: Emulation as an integrating principle for cognition. Front. Hum. Neurosci. **5**, 54 (2011)
46. Friston, K.J., Stephan, K.E.: Free-energy and the brain. Synthese **159**(3), 417–458 (2007)
47. Grush, R.: The emulation theory of representation: motor control, imagery, and perception. Behav. Brain Sci. **27**, 377–396 (2004)
48. Hawkins, J., Blakeslee, S.: On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines. Macmillan, New York (2007)
49. Pezzulo, G., Butz, M.V., Castelfranchi, C., Falcone, R.: The Challenge of Anticipation: A Unifying Framework for the Analysis and Design of Artificial Cognitive Systems. Springer Science & Business Media, New York (2008)
50. Ito, M.: The Cerebellum and Neural Control. Raven Press, New York (1984)
51. Kawato, M., Furukawa, K., Suzuki, R.: A hierarchical neural-network model for control and learning of voluntary movement. Biol. Cybern. **57**(3), 169–185 (1987)
52. Gross, H.M., Heinze, A., Seiler, T., Stephan, V.: Generative character of perception: a neural architecture for sensorimotor anticipation. Neural Netw. **12**, 1101–1129 (1999)
53. Yamakawa, H., Miyamoto, Y., Baba, T., Okada, H.: Cognitive distance learning problem solver reduces search cost through learning processes. Trans. Jpn. Soc. Artif. Intell. **17**, 1–13 (2002)
54. Yamakawa, H., Okada, H., Watanabe, N., Matsuo, K.: Cooperation and negotiation mechanism through generalized state vectors. In: Multi-Agent and Cooperative Computation 1998 (1989)

55. Yamakawa, H., Arakawa, N., Takahashi, K.: Reinterpreting the cortical circuit. In: AGA 2017: IJCAI Workshop on Architectures for Generality & Autonomy. http://cadia.ru.is/workshops/aga2017/ (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Hiroshi Yamakawa is Chairperson of The Whole Brain Architecture Initiative (WBAI), a non-profit organization He is an AI researcher interested in the brain. His specialty includes brain-inspired artificial general intelligence, concept formation, neurocomputing, and opinion aggregation technology. He is a former Chief Editor of the Japanese Society for Artificial Intelligence. He is a co-organizer of the AEGAP workshop of IJCAI 2018 and 2019. He received an MS in physics and PhD in engineering from the University of Tokyo in 1989 and 1992, respectively. He joined Fujitsu Laboratories Ltd. in 1992. He founded Dwango AI Laboratory in 2014 and was a director since March 2019. In 2015, he co-founded WBAI and became a chairperson. He is a visiting professor of the University of Electro-Communications. He is a co-author of the Japanese books "Religion and life" and "The Constitutional Theory of the AI Era".