AN ENGLISH NEURAL NETWORK THAT LEARNS TEXTS, FINDS HIDDEN KNOWLEDGE, AND ANSWERS QUESTIONS

Yuanzhi Ke, Masafumi Hagiwara

Graduate School of Science and Technology, Keio University,
3-14-1, Hiyoshi, Kohokuku, Yokohama, Japan

Abstract

In this paper, a novel neural network is proposed, which can automatically learn and recall contents from texts, and answer questions about the contents in either a large corpus or a short piece of text. The proposed neural network combines parse trees, semantic networks, and inference models. It contains layers corresponding to sentences, clauses, phrases, words and synonym sets. The neurons in the phrase-layer and the word-layer are labeled with their part-of-speeches and their semantic roles. The proposed neural network is automatically organized to represent the contents in a given text. Its carefully designed structure and algorithms make it able to take advantage of the labels and neurons of synonym sets to build the relationship between the sentences about similar things. The experiments show that the proposed neural network with the labels and the synonym sets has the better performance than the others that do not have the labels or the synonym sets while the other parts and the algorithms are the same. The proposed neural network also shows its ability to tolerate noise, to answer factoid questions, and to solve single-choice questions in an exercise book for non-native English learners in the experiments.

Keywords: natural language processing, neural network, question answering, natural language understanding

1 Introduction

Natural language processing is an important research area as a part of linguistic science valued in many fields such as human-machine interface or data-mining. An instinctive approach is writing rules to process natural language. This approach has a long-history being applied to many areas such as part-of-speech tagging [1]. However, it is extremely difficult to provide rules for well-formed and ill-formed utterances at the same time, whereas human beings often use ill-formed utterances in the daily-life for communicative needs [2].

Statistical approaches are based on an empiricist point of view assuming that a human does not have detailed principles of language when he or she is a baby, but learns them by association, pattern recognition and generalization upon what he or she hears and sees during years. Rather than building detailed grammatical rules manually, a statistical approach suggests specifying an statistical language model for a machine to process the natural languages. This kind of approach is able to deal with incomplete information, more loose and flexible than the rule-based approach, copes well with unseen utterances. They were firstly developed in the 1940s [3] and 1950s [4], have been used successfully and widely, and used for many tasks such as word sense disambiguation [5], part-of-speech tagging [6, 7, 8], machine translation [9],

and so on. However, the conventional statistical approaches heavily rely on the size and quality of the training data. Therefore, if there is no sufficient training data, we have to employ the other methods such as a rule-based approach.

Recently, neural networks become popular with the birth of deep learning. At the area of natural language processing, there have been researches about the application of neural networks on word vectorization [10], calculating similarity between texts [11], word sense disambiguation [12], text parsing [13], question answering [14, 15], machine translation [16] and so on. Not only they can deal with a mixture of formal and informal utterances, well-trained neural networks can also process short texts with appropriate training datasets.

In this paper, a novel neural network is proposed which is self-organized and works well with limited learning resources. The proposed neural network recalls lost parts of learned sentences, answers questions about the content in a document, and works well even when there is little training data to learn. It is organized on the basis of the structure of sentences in a text, able to connect neurons that correspond to similar contents based on the syntactic and semantic information. When the proposed neural network is learning, the learning methods and the algorithms make the neurons of similar contents bind closer.

This paper is organized as follows. Related works are introduced in Section 2. The structure and algorithms of the proposed neural network for learning and recalling are introduced in Section 3. The details of experiments are explained in Section 4. The discussion of the results and conclusions are given in Section 5.

2   Related Works

Hummel et al. proposed a theory and a model of inference and generalization based on the relational process of human thinking [17, 18, 19]. In their model, each proposition is represented by a four-tier hierarchy: semantic units, individual objects, sub-propositions and propositions. Individual objects share semantic units if they are semantically similar. The sub-propositions bind objects to relational roles, and are bound into complete proposi-

tions in the proposition tier. Because of the semantic units and objects connected by different propositions, the relations in each proposition, i.e. the "mappings", can be learned by extending the mappings via the connections and synchronizing the active states of the related propositions. In this way, the ability of associative thinking, which is an important aspect of human thinking, is implemented. They demonstrated that their model is able to simulate some of relational processes of human thinking, such as analogical inferences, schema induction, or interaction between schemas with inferences. However, their research did not pay much attention on problems about practical use, for example, it is not shown how to automatically recognize the propositions, sub-propositions, objects and semantic units and change them into the format in the model.

Saito et al. [20] also proposed a neural network for analogical inferences, which consists of a subject layer, a verb layer, a word layer, a connection layer and a dictionary layer. In their neural network, related neurons of subjects in the subject layer, verbs in the verb layer, and the other word neurons in the word layer are connected through neurons in the connection layer. At the same time, the word neurons are connected to the dictionary layer, in which there are neurons of the words in *Google N-gram dictionary* [21]. The neural network can perform inferences via the connection layer or by searching the dictionary layer for the word close to the activated words.

Fukuda et al. [22] proposed a neural network that consists of neurons of cases in a Japanese sentence with binary links labeled with the types of the relations between neurons, working with a module in their system that performs the transformation from the input sentence in Japanese to the neural network of cases. The neural network is designed to answer factoid questions by propagating the activated state of neurons. The neural network works well with a chronology corpus which contains only relatively simple sentences.

T. Sagara et al. [14], combined the ideas of the previous two researches. In their neural network, the neurons of cases are not connected to each other directly but connected to the neurons in the "knowledge layer" and the neurons in the "dictionary network". The neurons in the knowledge layer represent the propositions in the sentences. The neu-

rons in the dictionary network represent the head-words in a tree structured digital Japanese dictionary —*GoiTaike* [23] and *MeCab* [24]. The states of links and neurons are continuous variables instead of binaries, and a sigmoid function is employed to calculate the output of neurons. The neural network is able to deal with complicated sentences.

However, the neural networks based on the cases in the researches above by Saito, Fukuda, Sagara et al. are only for the Japanese language. Japanese has the obvious case markers. There are only nine case markers ("ga", "no", "wo", "ni", "he", "to", "de", "ka ra", "yo ri") in Japanese and all of them are particles separated from the words modified by them. It makes it easy to extract the cases in a Japanese sentence. But for the other languages such as English, extracting the cases is far more complicated to be done automatically, which makes the previous neural networks not as feasible for English as for Japanese.

3 The Proposed Neural Network

Our proposed neural network is constructed based on the syntactical structure of the sentences in a corpus. The cases are not extracted directly. The neurons of phrases and words are labeled with their semantic roles instead.

In the learning phase, the neural network is organized according to the parsed sentences. Then the active states of the neurons are propagated and the weights of the connections among the activated neurons are trained. In this way, the active states corresponding to the inputs are remembered. In the recalling phase, the neurons which correspond to the inputs are set to be activated, and the active states are propagated through the network layer by layer. After the propagation, the distribution of the active states of neurons in the neural network converges to a state at a past time when the sentences most related to the input were learned.

3.1 Architecture

As shown in Figure 1, the proposed neural network has five layers. Neurons in the sentence-layer, clause-layer, phrase-layer and word-layer correspond to the sentences, the clauses, the phrases, and

the words learned by the neural network respectively. Neurons in each of the top four layers are connected to those in the adjacent layers according to the parse tree. The neurons of the words are connected to the neurons of phrases containing them. Those of the phrases are connected to those of clauses, and those of clauses are connected to those of sentences similarly. Stop words are not involved in the network. The concept-layer contains a network of synonym sets (synsets), in which the synsets are connected as how they are in *WordNet* [25, 26]. Each neuron in the word-layer is connected to its corresponding synset in the concept-layer.
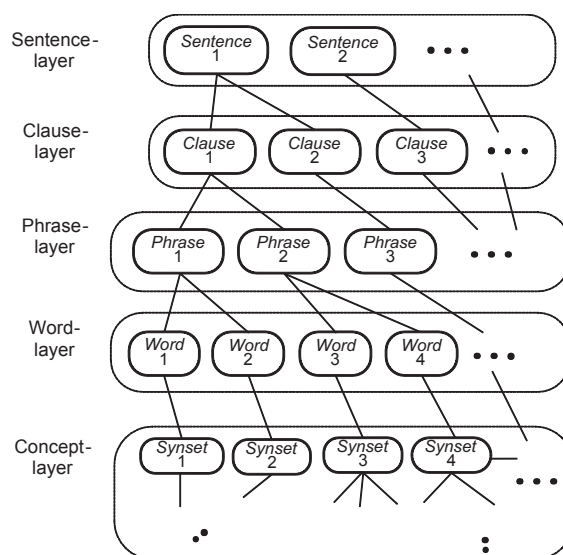

Figure 1. The overall view of the proposed neural network.

The same words share a single neuron and so are the phrases, clauses and sentences. Trees of sentences are connected by the neurons shared this way and the concept-layer.

Figure 2 shows an example of the neural network. In the example, the sentence neuron "every bird has wings and most bird species can fly" is connected to the two clause neurons, "every bird has wings" and "most bird species can fly", which are the 2 propositions in the sentence. Each phrase neuron is connected to the word neuron related to the phrase. For example, the phrase neuron "every bird" is connected to the word neuron "bird". Each word neuron is connected to the corresponding synset neuron in the concept-layer. The word neuron of "bird" is connected to the synset neuron of "bird", that of "have" is connected to the synset

neuron of "have", and that of "wing" is connected to the synset neuron of "wing".
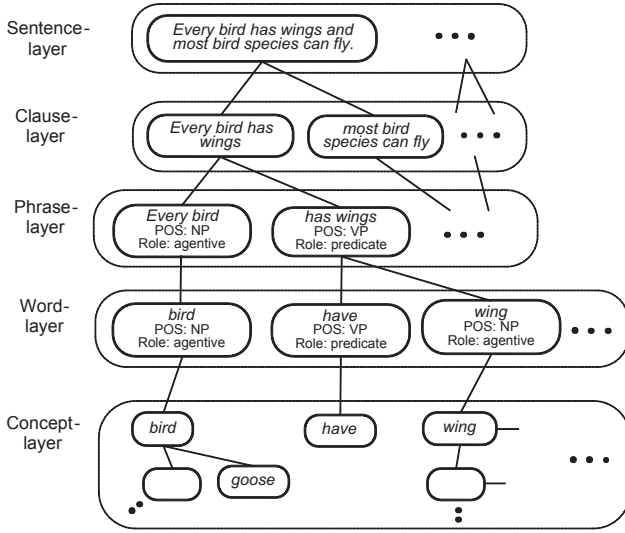

Figure 2. An example of the proposed neural network.

The concept-layer supports the proposed neural network to find hidden knowledge. For example, as shown in Figure 3, if the neuron of "bird" is activated, the corresponding neuron of "bird" in the concept-layer is also activated by propagation of the active states through their connection. The neurons of "goose" and "bird" are also connected in the concept layer. Therefore the neuron of "goose" in the concept-layer is activated similarly and so is the neuron of "goose" in the word-layer. Hence, if the proposition "birds have wings" is activated, "goose have wings" is activated as well analogically. This feature helps the neural network to respond correctly when there is no direct answer in the corpus and implicit knowledge is required to be extracted to give a correct response in some scenes. It also helps the proposed neural network tolerate noise because it makes the correct neurons easier to be activated.

Every neuron has an active state label $E \in [0, 1]$ representing its active state. All of the connections are bidirectional and weighted by a parameter $W \in [0, 1)$ representing the relationship between the neurons. Using the continuous variable instead of binary variable for the active state makes it possible to alleviate the interference from noise by adjusting the updating algorithm for propagation.
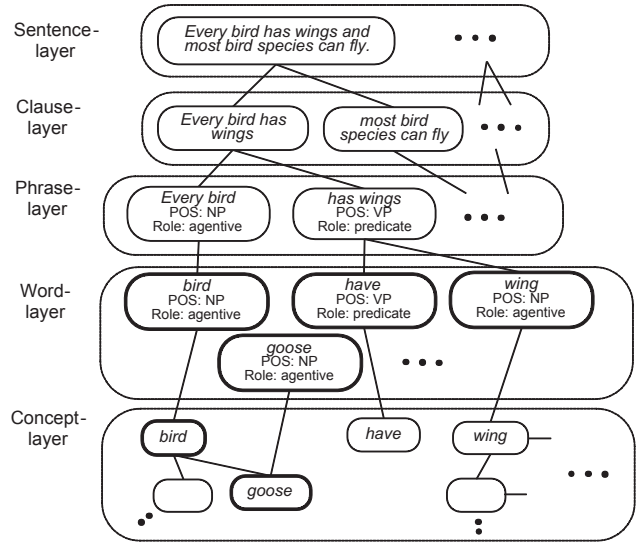

Figure 3. An example of how the proposed neural network finds hidden knowledge analogically by the neurons of synsets in the concept-layer.

3.2 Labels of Neurons

Each word neuron or synset neuron has a word name label, which is the position of the corresponding words in the vocabulary used for the neural network. Each phrase neuron has a serial number label which is the combination of the word name labels of the word neurons which are connected to it. Each clause neuron (or sentence neuron) has a serial number label which is the combination of those of the phrase neurons (or clause neurons) which are connected to the cluase (or the sentence) similarly.

Besides, each neuron in the phrase-layer and the word-layer has two labels, a label expressing the part-of-speech and a label expressing the semantic role. Each neuron in the phrase-layer and word-layer can be expressed as a vector $\vec{V} =$ *(word, part-of-speech, semantic,role)*. The same words with different part-of-speeches or different semantic roles are treated as the different neurons.

Like many other researches, the part-of-speech label set of *Penn Treebank* [27, 28] is used in the proposed neural network, and the semantic role labels are based on the label set of *PropBank* [29, 30, 31]. For a better recall rate avoiding overfitting, labels of all the semantic roles of arguments ("Arg0, Arg1, Arg2..." in PropBank) are tagged as "argument". Each semantic role is labeled together with the name label of the verb it belongs to.

Table 1. The part-of-speech labels used in the proposed neural network

Number	POS	Meaning	Number	POS	Meaning
0	CC	Coordinating Conjunctions	17	PRP$	Possessive Pronouns
1	CD	Cardinal Numbers	18	RB	Adverbs
2	DT	Determiners	19	RBR	Comparative Adverbs
3	EX	Existence There	20	RBS	Superlative Adverbs
4	IN	Prepositions and Subordinating Conjunctions	21	RP	Particles
5	JJ	Adjectives	22	TO	to
6	JJR	Comparative Adjectives	23	UH	Interjection
7	JJS	Superlative Adjectives	24	VB	Verbs (base form)
8	LS	List Item Markers	25	VBD	Verbs (past tense)
9	MD	Modal Verbs	26	VBG	Verbs (gerund or present participle)
10	NN	Common Nouns (Singular or Mass)	27	VBN	Verbs (past participle)
11	NNS	Common Nouns (Plural)	28	VBP	Verbs (non 3rd person singular present)
12	NNP	Proper Nouns (Singular)	29	VBZ	Verbs (3rd person singular present)
13	NNPS	Proper Nouns (Plural)	30	WDT	Wh-determiner
14	PDT	Predeterminers	31	WP	Wh-pronoun
15	POS	Possessive Endings 's	32	WP$	Possessive wh-pronoun
16	PRP	Personal Pronouns	33	WRB	Wh-adverb

Table 2. The semantic role labels used in the proposed neural network

Number	Label	Meaning	Number	Label	Meaning
0	Agent, Patient, Theme...	agentive	10	Secondary Predication	am-prd
1	Verb	verb	11	Purpose	am-prp
2	Comitative	am-com	12	Cause	am-cau
3	Locative	am-loc	13	Discourse	am-dis
4	Directional	am-dir	14	Adverbials	am-adv
5	Goal	am-gol	15	Adjectival	am-adj
6	Manner	am-mnr	16	Modal	am-mod
7	Temporal	am-tmp	17	Negation	am-neg
8	Extent	am-ext	18	Direct Speech	am-dsp
9	Reciprocals	am-rec	19	Causative Agents	aa

The labels of part-of-speeches and semantic roles are shown in Table 1 and Table 2. As shown later in the experiments, *SENNA* [13] is used to label the part-of-speech, semantic role and related verb of each word.

3.3 Learning Phase

This Section explains the learning phase of the proposed neural network. In this phase, the neurons and connections are organized and trained in an unsupervised way according to the input text.

At first, the input text is split into sentences and input to the neural network sentence by sentence. Because the architecture of the proposed neural network is based on the structure of a whole sentence, the input text needs to be processed sentence by sentence. The text is parsed and labeled at the beginning, and the input size of the neural network is not defined previously but set up according to the input text after the input text is parsed at the beginning of the learning phase, together with the number of neurons in each layer except for the concept-layer. The parser of *SENNA* [13] is used to find the beginning and the end of the sentences.

According to the parsing result, neurons are allocated in each layer for the sentences, clauses, phrases, and words. The neurons are connected as mentioned in Section 3.1. Then the initial active state E_0 of each neuron in the sentence layer is set to be 1.00, and the initial weight of each connection W_0 is set by the following equation

$$w_{ij}^{(0)} = \frac{w_{base}}{F_{ij}}. \tag{1}$$

Here, F_{ij} is the number of connections to the neu-

ron, set to make the active state propagated equally. w_{base} is a constant.

Then the active state of each neuron is updated layer by layer as shown in Figure 4. At first, the neurons in the sentence layer are updated, then those in the clause layer, the phrase layer, the word layer and the concept layer are updated. The active states are updated by the following equation

$$E_i^{(t+1)} = \frac{1}{1 + e^{-\alpha(E_i^{(t)} + \Delta E_i^{(t)} - \delta)}}. \tag{2}$$

Here, $\Delta E_i^{(t)}$ is the increment of the active state at time t of neuron i, which is the sum of the input to the neuron, defined as

$$\Delta E_i^{(t)} = \sum_{j}^{n} w_{ij}^{(t)} E_j^{(t)}. \tag{3}$$

Here, $w_{ij}^{(t)}$ is the weight of the connection between neuron i and neuron j at time t. The slop parameter α and the x-axis offset δ control the shape of the sigmoid function in the first quadrant and affect the performance of learning and recalling.

The active state defined this way is always between 0.00 and 1.00. If the active state is bigger than the threshold θ which is also set between 0.00 and 1.00, the neuron is considered activated. If neurons at both sides of a connection are activated, the weight of the connection is updated by Hebbian rule as the following

$$\begin{aligned} w_{ij}^{(t+1)} = w_{ij}^{(t)} + \gamma(w_{max} - w_{ij}^{(t)}), \\ if \; E_i^t > \theta \; and \; E_j^t > \theta. \end{aligned} \tag{4}$$

Here, γ is the parameter controlling the increment of the weight. By this method, the weights of the

connections between related neurons will be higher than those of the others, making the related neurons easy to be activated together. After learning, the active states of the neurons are reset to 0.00.

3.4 Recalling Phase

In the recalling phase, the network recalls the learned words, phrases, clauses or sentences. Instead of outputting something directly from the neural network, the neural network goes into a state which it used to be in the learning phase, whereby the recalled neurons are indicated by a high active state. At first, the input is parsed and labeled in the same way as in the learning phase. Then the stop words are filtered out and the active states of the neurons of the other words are set to be 0. The active states of all the neurons are updated by the Equation (2) and Equation (3), the same as in the learning phase. But the weights of connections are not updated and the update order is different. In the recalling phase, the neurons in the concept-layer which are connected to the active neurons in the word layer are updated at first, to infer hidden knowledge. Then the other neurons in the concept layer are updated. After that, the neurons in the word layer which are connected to the active neurons in the concept layer are updated. Then the other neurons in the neural network are updated in the following order: the phrase layer, the clause layer, the sentence layer, and then the clause layer, the phrase layer, and the word layer again. The overall update process order in the recalling phase is shown in Figure 5.
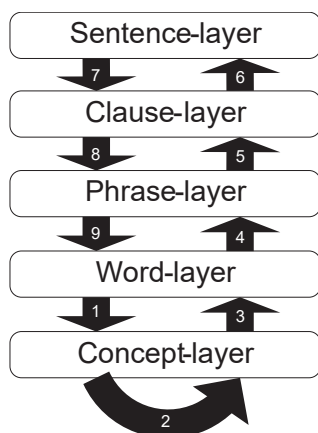
Figure 5. The update sequence in the recalling phase.

After the propagation of the active states, some neurons have higher active states. The higher the active state of a neuron is, the more related to the input in the past learning phase, whereby the neural network can be applied for such tasks as question answering. For example, supposed that the neural network is asked with a question "What date in 1989 did East Germany open the Berlin Wall?". After the propagation, the active states of the neural network will be like those shown in Figure 6.
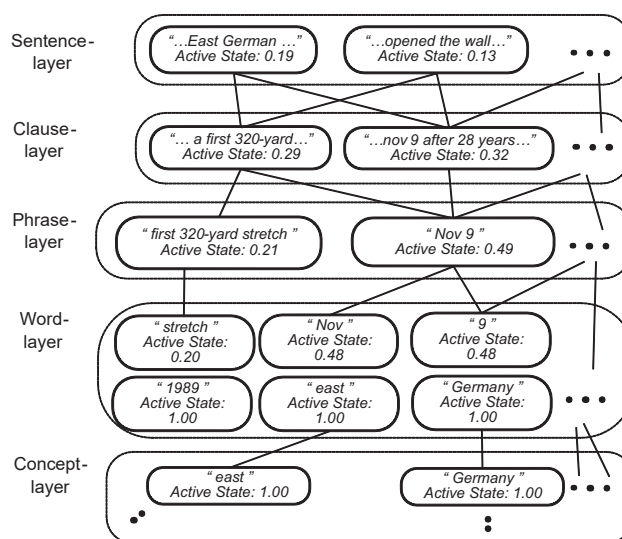
Figure 6. An example of the end state after the propagation in the recalling phase.

As shown in Figure 6, except for the neurons of the words in the input activated at the beginning, the phrase neuron "Nov 9" and the word neuron "Nov" and "9" have the highest active state. In this case, the phrase neuron "Nov 9" is regarded as the answer.

4 Experiments

Three kinds of experiments were performed to evaluate the proposed neural network. The first one was designed to evaluate the recalling performance and noise tolerance. The second was for the feasibility to apply the neural network for the question answering tasks. At last, the neural network was made challenge some reading comprehension questions in an exercise book for the Test of English for International Communication (TOEIC) [32]. The parameters of the neural network used in the experiments are shown in Table 3.

4.1 Evaluation Experiments of Recalling and Noise Tolerance

In these experiments, the task is to recall a lost phrase from each of the sentences in the dataset. In each one of these experiments, 100 sentences were randomly chosen from *NIST TREC Document Database: Disk 4 & Disk 5* [33, 34] for the experiment. Each sentence was chosen automatically and checked whether the sentence could be parsed and whether at least one clause and three phrases could be extracted from it by the parser used for the proposed neural network. If some of the sentences did not meet the requirements, they were removed from the dataset and that many sentences were randomly drawn from the database into the dataset to replace them. This process was repeated until all of the sentences in the dataset could be parsed and extracted at least one clause and three phrases out to make sure that the sentences were not too simple and the network could work correctly. Then the parser of *SENNA* was used to get the parse trees of the sentences and the labels of the phrases and the words. Parse trees of the sentences were saved as the training set. Then, for each parse tree in the training set, a new copy of the tree was created and a sub tree of a random phrase in the copy was deleted. The new copies that lost sub trees of phrases were saved as the test set.

In the experiments, several trees of the sentences in the training set were learned together by the neural network in the learning phase, and that many corresponding trees in the test set were input into the neural network in the recalling phase to recall the lost phrases whose corresponding trees were deleted in the test set. At first, one sentence was input per time. Then the number of input sentences per time was increased. For each sentence, other sentences that processed together, could disturb the propagation in the learning phase and the recalling phase as noises because they were connected together in the neural network by the word-layer and the concept-layer.

Besides the proposed neural network, the same experiments were also repeated on two similar neural networks as a control group to see whether the labels and the concept layer improve the performance. Other parts of the neural networks in the control group were the same as the proposed neural network but the neurons of one of them were not labeled with part-of-speeches and semantic roles, and the other one did not include the concept layer. The mean reciprocal ranks of the active states of the phrase neurons removed in the test dataset were counted as the measure of the performance. All of the results are shown in Figure 7.

The results show that the proposed neural network can recall all of the lost phrases perfectly without noise. When noise is increased, the rank of the correct phrase decreases. However, the mean reciprocal ranks of the correct phrases never fall below 0.2 if the neural network has the concept layer and labeled with semantic roles and part-of-speeches. If the semantic roles and part-of-speeches are not involved, the performance deteriorates rapidly when noise is increased, and the mean reciprocal ranks of the correct phrases go down to 0.1 even though only 28 sentences are processed together. Until 20 sentences are processed together, there is no large difference between the performance of the neural network with the concept layer and that without the concept layer because there are not enough sentences to infer hidden knowledge on the basis of the synonyms of the words for the neural network. When more than 20 sentences are processed together, the concept layer observably improves the performance and the neural network with the concept layer usually recalls the lost phrases more correctly. While the neural network with the concept layer keeps the mean reciprocal ranks above 0.2, the neural network without the concept layer fails to do that when there are more than 66 sentences processed together. The results show that the semantic roles, the part-of-speeches, the concept layer involved and the inference ability based on them improve the recalling performance and keep it tolerant to noise.

4.2 Question Answering

In this experiment, the question answering (QA) track in TREC-8 was used to evaluate the proposed neural network when it is employed for information retrieval. In this track, each question was about one of the 555,000 documents in *NIST TREC Document Database: Disk 4 & Disk 5*. The documents come from *Congressional Record of the 103rd Congress, Federal Register (1994), Financial Times (1992-1994), the Foreign Broadcast Information Service and Los Angeles Times (randomly se-*
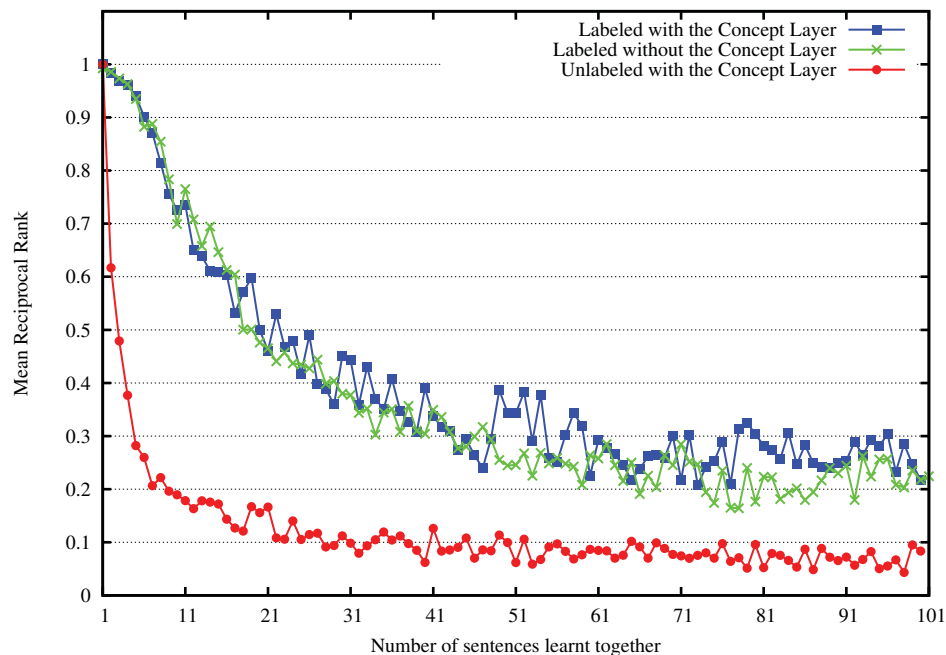
Figure 7. Mean reciprocal ranks in the recalling and noise tolerance test. The blue line with square points is the performance of the neural network with labeled neurons and the concept layer. The green line with cross points is the performance of the neural network with labeled neurons but without the concept layer. The red line with the circle points is the performance of the neural network with the concept layer but without labeled neurons.

lected articles from 1989 & 1990). An example of the questions in the QA track is the following:

"What date in 1989 did East Germany open the Berlin Wall?"

As stated in Section 3.4, to answer this question, at first the content words "date", "1989", "East", "Germany", "open", "Berlin" and "Wall" were input into the trained proposed neural network. Then after the propagation, the sentences with the highest active states were output as the candidates for the answer, ordered by their active states. In this way, the proposed neural network answered the questions.

The computer used in the experiment was a desktop personal computer with Intel Core i7-2600S CPU running on 2,794MHz and 12G DDR3 RAM Memory running on 1,333MHz. It took 57.5 hours for the computer to run the neural network to learn and recall 10000 sentences in the corpus in the previous experiments, while there were more than 1,000,000 sentences in the whole corpus for the QA track. It was thus impossible to build the neural network of the whole corpus in an acceptable

time. Therefore only the development question set of the QA track in TREC-8, provided with the name of the documents containing the answer, was used in this experiment.

The results in the TREC-8 question answering track report [35] are split into two tables, one is the results that the output limit is 250 bytes, and the other is those that the output limit is 50 bytes. As the response of the system is a sentence that can be more than 50 bytes, the results are compared with the 250-byte responses.

Besides, two different ways to train the proposed neural network in the preprocess were compared in the experiments. One was to learn the whole document without checking if any words in the questions were contained. The other one was to only learn the related sentences that contained at least a content word in the questions.

As shown in Figure 8, without choosing what to learn, the MRR of the correct response of the proposed neural network is 0.170, between those of Scai8QnA and UIowaQA1; the percent of the questions for which correct responses are found is
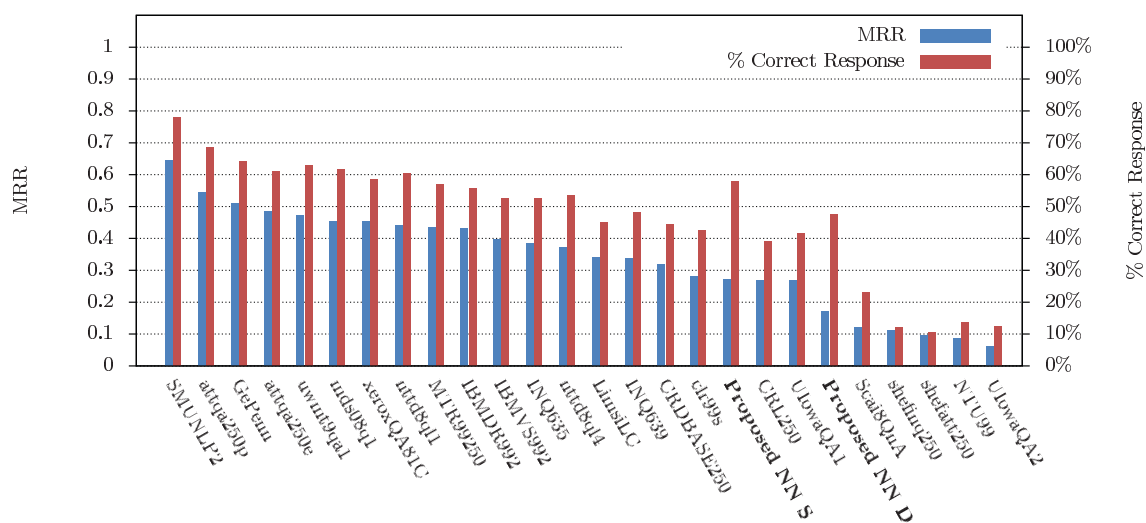
Figure 8. The comparison of results of the proposed neural network with the other systems stated in the TREC-8 question answering track report [35]. "MRR" means the mean reciprocal rank of the correct response among all of the responses of the system. "% correct response" means the percent of questions for which correct responses were found. "Proposed NN D" is the scores of the proposed neural network learning the whole documents without choosing. "Proposed NN S" is the scores if the neural network only chose the sentences in which at least a content word in the questions is detected to learn. The proposed neural network was only tested with the development set because it is impossible for our computer to complete the whole track in an acceptable processing time.

47.4%, between INQ639 and LimsiLC. If only related sentences which containing keywords in the questions are learned, the MRR is increased to 0.27, between CRL250 and clr99s; and the percent of the questions for which correct responses are found is increased to 57.9%, between MTR99250 and xeroxQA81C.

No matter whether the input sentences are filtered, the percent of the questions for which correct responses are found in the responses of the proposed neural network reaches the average (44.7%) of the other systems in the TREC-8 question answering track report [35], which indicates that the neural network could find the correct answer as well as the others. If it is made to choose what it is to learn, it works better. It can be also noticed that the MRR of the correct answer is near to the lower limit of those in the recalling test. It indicates that the performance of the proposed neural network is consistent in these two different tasks.

4.3 Solving Reading Comprehension Questions

The Test of English for International Communication (TOEIC) [32] is an English test for non-native English speakers to measure their daily English skills. The reading comprehension part of TOEIC contains passages of advertisements, mails or newspapers. For each of the passages, there are several questions about it. For each question, there are four choices to choose. As the question database of TOEIC is not public, the questions in an exercise book called *Longman Preparation Series for the New TOEIC Test: More Practice Tests* [36] was used in this experiments. In this exercise book, there is one correct answer in the choices for a single question. Most of the questions are about the content of the passages, but some ask examinees to choose a synonym of a word. The latter ones were skipped in the experiments. The following sentence is an example of the former kind of questions which was used in the experiments:

"According to the passage, who will be the most affected by this change?"

with four choices provided for it:

"A. postal employees; B. greeting card manufacturers; C. direct-marketing companies; D. stamp collectors."

The neural network was asked to choose the correct answer from the four choices like those at above. The process to answer a question was as the following:

1 At first, the passage which the question was for input into the neural network to learn.

2 After the passage was learned and the active states were reset, the active states of neurons in the sentence-layer and those corresponding to the content words in the questions were set to 1.00 if they exist. Then the neural network began to recall the content words in the questions.

3 At last, the active states of the corresponding neurons of the content words in each choice were summed as the score of the choice if the neurons exist. The choice that gained the most score was chosen as the answer.

The system with the proposed neural network was asked to answer 52 questions about the contents of 19 passages in the exercise book. 29 of them were correctly answered. The accuracy is 55.8%. According to *Examinee Handbook Listening & Reading* [38] for TOEIC, the standard error of measurement between the true scores and the test scores is 25 scaled score points. Therefore, the system is finally considered gaining a score between 251 and 301. According to the official program data and analysis of TOEIC [37], the overall average score of the public tests in the reading part in 2014 is 263, in which, the average score of businessmen is 273, and that of students is 251. As in Figure 9, the system has reached the average level of the student examinees of TOEIC.

Besides, this experiment result also indicates that the size of the corpus for training does not matter much for the proposed neural network, because the correct rate 55.8% is close to the percent of questions for which correct responds were found in the previous QA experiment (57.9%), while the

passages in TOEIC are much shorter than those in the corpus used for the QA experiment.

5 Conclusion and Discussion

A novel neural network is proposed in this paper. The proposed neural network is able to learn texts, find hidden knowledge and recall information in the learned texts. The neural network has the layers corresponding to the sentences, the clauses, the phrases, the words in the sentences and the synonym sets. The neurons in the sentence-layer, clause-layer, phrase-layer and word-layer are connected according to the parse trees. In the learning phase, the propagation methods and algorithms make the neural network organize itself for the input and build channels for inferences of the hidden knowledge. In the recalling phase, the neural network is designed to go into a state in which the neurons most related to the input are activated, instead of outputting something directly from the neural network.

Wordnet is involved and connected to the neurons of words as the concept-layer that supports the proposed neural network to find hidden knowledge. The concept-layer helps the proposed neural network find the analogical relationships between bodies of texts.

The evaluation experiments indicated that the proposed neural network which can infer hidden knowledge by the part-of-speech labels, the semantic role labels and the concept-layer has the better recalling performance and noise tolerance than the other neural networks that do not have the labels or the concept-layer while the other parts and the algorithms are the same. The experiments answering the factoid questions from TREC-8 QA track and single-choice questions from a TOEIC exercise book indicate that the proposed neural network can be applied to answering questions about a corpus no matter it is large or small. The correct rate is solid in the different experiments and positively correlated with the lower limit of the recalling ability of the neural network itself, which is related to the parameters and algorithms in the recalling phase. It indicates that an optimization of the algorithms or the parameters can improve the performance.
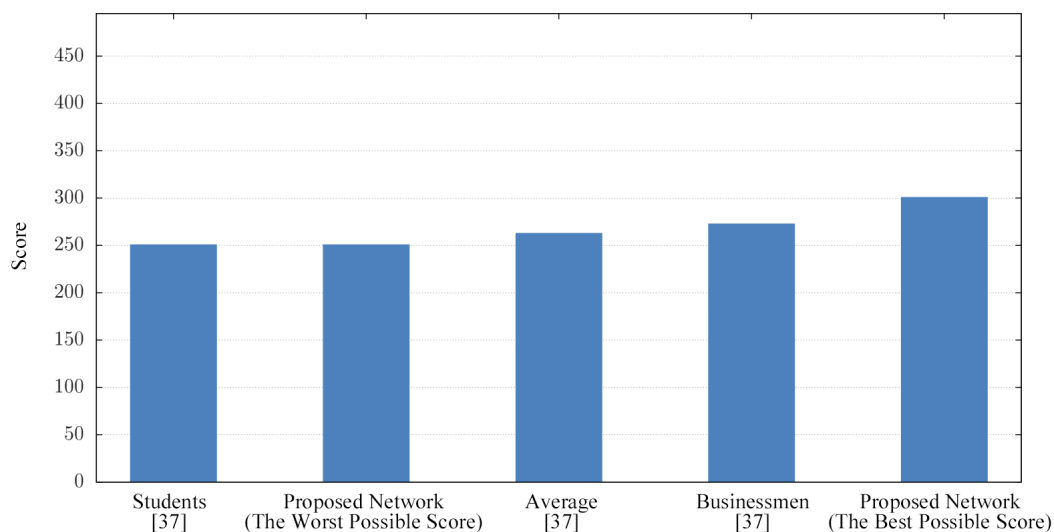
Figure 9. Comparison of the scores by the proposed neural network and the human examinees in 2014 [37]. "Students" is the average scores of the student examinees in 2014. "Average" is the average scores of all the examinees in public TOEIC test in 2014. "Businessmen" is the average scores of the businessman examinees in 2014. Because the TOEIC test score conversion table is unpublic and we only know that the standard error of measurement between the true scores and test scores is 25 [38]. Therefore "the worst possible score" and "the best possible score" are given, which are the worst and the best possible score of the proposed neural network for its accuracy of 55.8%.

References

[1] E. Brill, A simple rule-based part of speech tagger, In Proceedings of the Workshop on Speech and Natural Language, HLT'91, pp. 112–116, Association for Computational Linguistics, Stroudsburg, PA, USA, 1992

[2] C. D. Manning and H. Schütze, Foundations of statistical natural language processing, MIT press, 1999

[3] C. E. Shannon, A mathematical theory of communication, SIGMOBILE Mob. Comput. Commun. Rev., vol. 5, no. 1, pp. 3–55, 2001

[4] N. Chomsky, Three models for the description of language, Information Theory, IRE Transactions on, vol. 2, no. 3, pp.113–124, 1956

[5] W. A. Gale, K. W. Church, and D. Yarowsky, Work on statistical methods for word sense disambiguation, In Working Notes of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language, vol. 54, p. 60. 1992

[6] J. Kupiec, Robust part-of-speech tagging using a hidden markov model, Computer Speech & Language, vol. 6, no. 3, pp. 225 – 242, 1992

[7] H. Schmid, Probabilistic part-of-speech tagging using decision trees, in Proceedings of the international conference on new methods in language processing, vol. 12, pp. 44–49. Citeseer, 1994

[8] A. Ratnaparkhi et al, A maximum entropy model for part-of-speech tagging, in Proceedings of the conference on empirical methods in natural language processing, vol. 1, pp. 133–142. Philadelphia, USA, 1996

[9] P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer, The mathematics of statistical machine translation: Parameter estimation, Computational linguistics, vol. 19, no. 2, pp. 263–311, 1993

[10] H. Hotta, M. Kittaka, and M. Hagiwara, Word vectorization using relations among words for neural network, IEEJ Transactions on Electronics, Information and Systems, vol. 130, pp. 75–82, 2010

[11] G. Tsatsaronis, I. Varlamis, and M. Vazirgiannis, Text relatedness based on a word thesaurus, Journal of Artificial Intelligence Research, vol. 37, no. 1, pp. 1–40, 2010

[12] G. Tsatsaronis, I. Varlamis, and M. Vazirgiannis, Word sense disambiguation with semantic networks, In Text, Speech and Dialogue, pp. 219–226. Springer, 2008

[13] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, Natural language processing (almost) from scratch, The Journal of

Machine Learning Research, vol. 12, pp. 2493–2537, 2011

[14] T. Sagara and M. Hagiwara, Natural language neural network and its application to question-answering system, Neurocomputing, vol. 142, pp. 201 – 208, 2014

[15] L. Dong, F. Wei, M. Zhou, and K. Xu, Question answering over Freebase with multi-column convolutional neural networks, in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pp. 260–269. Association for Computational Linguistics, Beijing, China, July 2015

[16] D. Bahdanau, K. Cho, and Y. Bengio, Neural machine translation by jointly learning to align and translate, CoRR, vol. abs/1409.0473, 2014

[17] J. E. Hummel and K. J. Holyoak, Distributed representations of structure: A theory of analogical access and mapping, Psychological Review, vol. 104, no. 3, p. 427, 1997

[18] J. E. Hummel and K. J. Holyoak, A symbolic-connectionist theory of relational inference and generalization, Psychological review, vol. 110, no. 2, p. 220, 2003

[19] J. E. Hummel and K. J. Holyoak, Relational reasoning in a neurally plausible cognitive architecture an overview of the LISA project, Current Directions in Psychological Science, vol. 14, no. 3, pp. 153–157, 2005

[20] M. Saito and M. Hagiwara, Natural language processing neural network for analogical inference, In The 2010 International Joint Conference on Neural Networks, pp.1–7, 2010

[21] T. Kudo and H. Kazawa, Web Japanese N-gram version 1, Gengo Shigen Kyokai, vol. 14, 2007

[22] M. Fukuda, S. Nobesawa, and I. Tahara, Knowledge representation for query-answer, In Forum on Information Technology, vol. 3, pp. 233–236, Information Processing Society of Japan, 2004

[23] S. Ikehara, M. Miyazaki, S. Shirai, A. Yokoo, H. Nakaiwa, K. Ogura, Y. Ooyama, and Y. Hayashi, GoiTaikei — A Japanese Lexicon, Iwanami Shoten, 1997

[24] T. Kudo, K. Yamamoto, and Y. Matsumoto, Applying conditional random fields to Japanese morphological analysis, in Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, vol. 2004, pp. 230–237. 2004

[25] G. A. Miller, WordNet: a lexical database for english, Communications of the ACM, vol. 38, no. 11, pp. 39–41, 1995

[26] G. Miller and C. Fellbaum, WordNet: An electronic lexical database, 1998

[27] M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini, Building a large annotated corpus of english: The Penn Treebank, Comput. Linguist., vol. 19, no. 2, pp. 313–330, June 1993

[28] M. Marcus, G. Kim, M. A. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger, The Penn Treebank: Annotating predicate argument structure, in Proceedings of the Workshop on Human Language Technology, HLT '94, pp. 114–119, Association for Computational Linguistics, Stroudsburg, PA, USA, 1994

[29] P. K. Martha and M. Palmer, From Treebank to Propbank, in Proceedings of the International Conference on Language Resources and Evaluation 2002, Las Palmas, Canary Islands, Spain, 2002

[30] P. Kingsbury, M. Palmer, and M. Marcus, Adding semantic annotation to the penn treebank, in Proceedings of the Human Language Technology Conference, pp. 252–256, Citeseer, 2002

[31] M. Palmer, D. Gildea, and P. Kingsbury, The proposition bank: An annotated corpus of semantic roles, Comput. Linguist., vol. 31, no. 1, pp. 71–106, March 2005

[32] P. E. Woodford, The test of english for international communication (TOEIC), 1980

[33] National Institute of Standards and Technology, NIST TREC Document Database: Disk 4, Retrieved June 25, 2016, from http://www.nist.gov/srd/nistsd22.cfm

[34] National Institute of Standards and Technology, NIST TREC Document Database: Disk 5, Retrieved June 25, 2016, from http://www.nist.gov/srd/nistsd23.cfm

[35] E. M. Voorhees et al. The TREC-8 question answering track report, in Proceedings of the 8th Text Retreval Conference, vol. 99, pp.77–82. NIST, Gaithersburg, MD, 1999

[36] L. Lougheed, Longman preparation series for the new TOEIC test: More practice tests, 2006

[37] T. S. Committee et al., TOEIC program data & analysis, 2014

[38] E. T. Service, Examinee handbook listening & reading, 2008

Yuanzhi Ke received his B.E. degree from Northeastern University (China), Shenyang, China, in 2011, and M.E. degree from Keio University, Yokohama, Japan in 2014. Currently, he is a Ph.D. student in Keio University. His research interests include neural networks, natural language processing, and machine learning.

Masafumi Hagiwara received his B.E., M.E. and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1982, 1984 and 1987, respectively. Since 1987 he has been with Keio University, where he is now a Professor. From 1991 to 1993, he was a visiting scholar at Stanford University. He received IEEE Consumer Electronics Society Chester Sall Award in 1990, Author Award from the Japan Society of Fuzzy Theory and Systems in 1996, Technical Award and Paper Awards from Japan Society of Kansei Engineering in 2003, 2004 and 2014, Best research award from Japanese Neural Network Society in 2013. He is a member of IEICE, IPSJ, JSAI, SOFT, IEE of Japan, Japan Society of Kansei Engineering, JNNS and IEEE (Senior member). His research interests include neural networks, fuzzy systems, and affective engineering. He is now the president of the Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT).