Association-based Natural Language Processing with Neural Networks

KIMURA Kazuhiro   SUZUOKA Takashi
AMANO Sin-ya
Information Systems Laboratory
Research and Development Center
TOSHIBA Corp.
1 Komukai-Tôsiba-tyô, Saiwai-ku, Kawasaki 210 Japan
kim@isl.rdc.toshiba.co.jp

Abstract

This paper describes a natural language processing system reinforced by the use of association of words and concepts, implemented as a neural network. Combining an associative network with a conventional system contributes to semantic disambiguation in the process of interpretation. The model is employed within a *kana-kanji conversion system* and the advantages over conventional ones are shown.

1 Introduction

Currently, most practical applications in natural language processing (NLP) have been realized via symbolic manipulation engines, such as grammar parsers. However, the current trend (and focus of research) is shifting to consider aspects of semantics and discourse as part of NLP. This can be seen in the emergence of new theories of language, such as Situation Theory [Barwise 83] and Discourse Representation Theory [Kamp 84]. While these theories provide an excellent theoretical framework for natural language understanding, the practical treatment of context dependency within the language can also be improved by enhancing underlying component technologies, such as knowledge based systems. In particular, alternate approaches to symbolic manipulation provided by connectionist models [Rumelhart 86] have emerged. Connectionist approaches enable the extraction of processing knowledge from examples, instead of building knowledge bases manually.

The model described here represents the unification of the connectionist approach and conventional symbolic manipulation; its most valuable feature is the use of word associations using neural network technology. Word and concept associations appear to be central in human cognition [Minsky 88]. Therefore, simulating word associations contributes to semantic disambiguation in the computational process of interpreting sentences by putting a strong preference to expected words(meanings).

The paper describes NLP reinforced by association of concepts and words via a connectionist network. The model is employed within a NLP application system for *kana-*

224

kanji conversion[1]. Finally, an evaluation of the system and advantages over conventional systems are presented.

2 A brief overview of kana-kanji conversion

Japanese has a several interesting feature in its variety of letters. Especially the existence of several thousand of *kanji* (based on Chinese characters; 海, 山,..) made typing task hard before the invention of kana-kanji conversion[Amano 79] . Now it has become a standard method in inputting Japanese to computers. It is also used in word processors and is familiar to those who are not computer experts. It comes from the simpleness of operations. By only typing sentences by phonetic expressions of Japanese (*kana*), the kana-kanji converter automatically converts *kana* into meaningful expressions(*kanji*). The simplified mechanism of kana-kanji conversion can be described as two stages of processing: morphological analysis and homonym selection.

- Morphological Analysis

  Kana-inputted (fragment of) sentences are morphologically analized through dictionary look up, both lexicons and grammars. There are many ambiguities in word division due to the agglutinative nature of Japanese (Japanese has no spaces in text). Each partitioning of the kana is then further open to being a possible interpretation of several alternate *kanji*. The spoken word *douki*, for example, can mean *motivation, pulsation, synchronization*, or *copperware*. All of them are spelt identically in kana(どうき), but have different kanji characters(動機, 動悸, 同期, 銅

  ---
  [1]Many commercial products use kana-kanji conversion technology in Japan, including the TOSHIBA Tosword-series of Japanese word processors.

器,respectively). Some kana words have 10 or more possible meanings. Therefore the stage of *Homonym Selection* is indispensable to kana-kanji conversion for the reduction of homonyms.

- Homonym Selection

  Preferable semantic homonyms are selected according to the co-occurrence restrictions and selectional restrictions. The frequency of use of each word is also taken into account. Usually, the selection is also reinforced by a simple context holding mechanism; when homonyms appear in previous discourse and one of them is chosen by a user, the word is automatically memorized in the system as in a cache technology. Then, when the same homonyms appear the memorized word is selected as the most preferred candidate and is shown to the user.

3 Association-based kana-kanji conversion

The above mechanisms are simple and effective in regarding kana-kanji converter as a typing aid. However, the abundance of homonyms in Japanese contributes to many of the ambiguities and a user is forced to choose the desired *kanji* from many candidates. To reduce homonym ambiguities a variety of techniques are available; however, these tend to be limited from a semantic disambiguation perspective. In using word co-occurrence restrictions, it is necessary to collect a large amount of co-occurrence phenomena, a practically impossible task. In the case of the use of selectional restrictions, an appropriate thesaurus is necessary but it is known that defining the conceptual hierarchy is difficult work [Lenat 89][EDR 90]. Techniques for storing previous kanji selections (cache)

クロックの立ち上がりにおいてこの信号はプロセッサに取り込まれる。
At the rising edge of the clock, the signal is fetched by the processor.
このようにどうきが
Like this, the douki

douki
どうき

signal
信号

clock
クロック

processor
プロセッサ

motivation
動機

synchronization
同期

suspicion
容疑

crime
犯行

police
警察

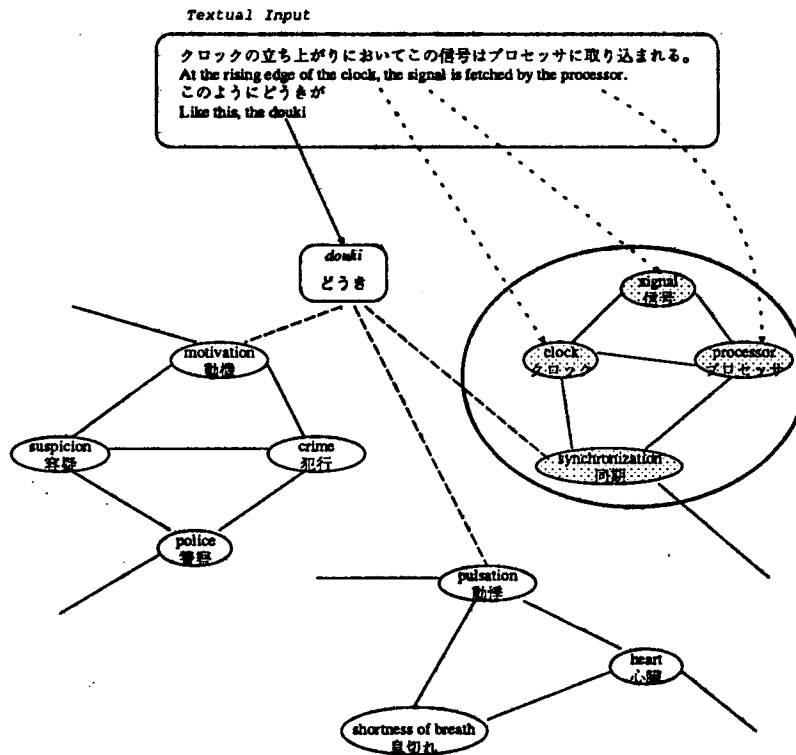pulsation
動悸

heart
心臓

shortness of breath
息切れ

Figure 1: Kana-Kanji Conversion with a Neural Network

are too simple to disambiguate between possible previous selections for the same homonym with respect to the context or between context switches.

To avoid these problems without increasing computational costs, we propose the use of the associative functionality of neural networks. The use of association is a natural extension to the conventional context holding mechanism. The idea is summarized as follows. There are two stages of processing: network generation and kana-kanji conversion.

A network representing the strength of word association is automatically generated from real documents. Real documents can be considered as training data because they are made

of correctly converted *kanji*. Each node in the network uniquely correspond to a word entry in the dictionary of kana-kanji conversion. Each node has an activation level. The link between nodes is a weighted link and represents the strength of association between words. The network is a Hopfield-type network[Hopfield 84]; links are bidirectional and a network is one layered.

When the user chooses a word from homonym candidates, a certain value is inputted to the node corresponding to the chosen word and the node will be activated. The activation level of nodes connected to the activated node will be then activated. In this manner, the activation spreads over the net-

226

work through the links and the active part of the network can be considered as the associative words in that context. In kana-kanji conversion, the converter decides the preference of word order for homonyms in the given context by comparing the node activation level of each node of homonyms. An example of the method is shown in Figure 1.

Assume the network is already built from certain documents. A user is inputting a text whose topic is related to computer hardware. In the example, words like *clock* (クロック) and *signal* (信号) already appear in the previous context, so their activation levels are relatively high. When the word *DOUKI* (どうき) is inputted in *kana* and the conversion starts, the activation level of *synchronization* (同期) is higher than that of other candidates due to its relationship to *clock* or *signal*. The input *douki* is then correctly converted into *synchronization* (同期).

The advantages of our method are:

- The method enables kanji to be given based on a preference related to the current context. Alternative kanji selections are not discarded but are just given a lower context weighing. Should the context switch, the other possible selections will obtain a stronger context preference; this strategy allows the system to capably handle context change.

- Word preferences of a user are reflected in the network.

- The correctness of the conversion is improved without high-cost computation such as semantic/discourse analyses.

4 Implementation

The system was built on Toshiba AS-4000 workstation (Sun4 compatible machine) using

C. The system configuration is shown in Figure 2.

The left-hand side of the dashed line represents an off-line network building process. The right-hand side represents a kana-kanji conversion process reinforced with a *neural network handler*. The network is used by the *neural network handler* and word associations are done in parallel with *kana-kanji* conversion. The *kana-kanji converter* receives *kana*-sequences from a user. It searches the *dictionary* for lexical and grammatical information and finally creates a list of possible homonym candidates. Then the *neural network handler* is requested for activation levels of homonyms. After the selection of preferred homonyms, it shows the candidates in *kanji* to a user. When the user chooses the desired one, the chosen word information is sent to the *neural network handler* through a *homonym choice interface* and the corresponding node is activated.

The roles and the functions of main components are described as follows.

- Neural Network Generator

  Several real documents are analyzed and the network nodes and the weights of links are automatically decided. The documents consist of the mixture of *kana* and *kanji*; homonyms for the *kanji* within the given context are also provided. The documents, therefore, can be seen as training data for the neural network. The analysis proceeds through the following steps.

  1. Analyze the documents morphologically and convert into a sequence of words. Note that particles and demonstratives are ignored because they have no characteristics in word association.

  2. Count up the frequency of the all combination of co-appeared word-pair in a paragraph and memorize
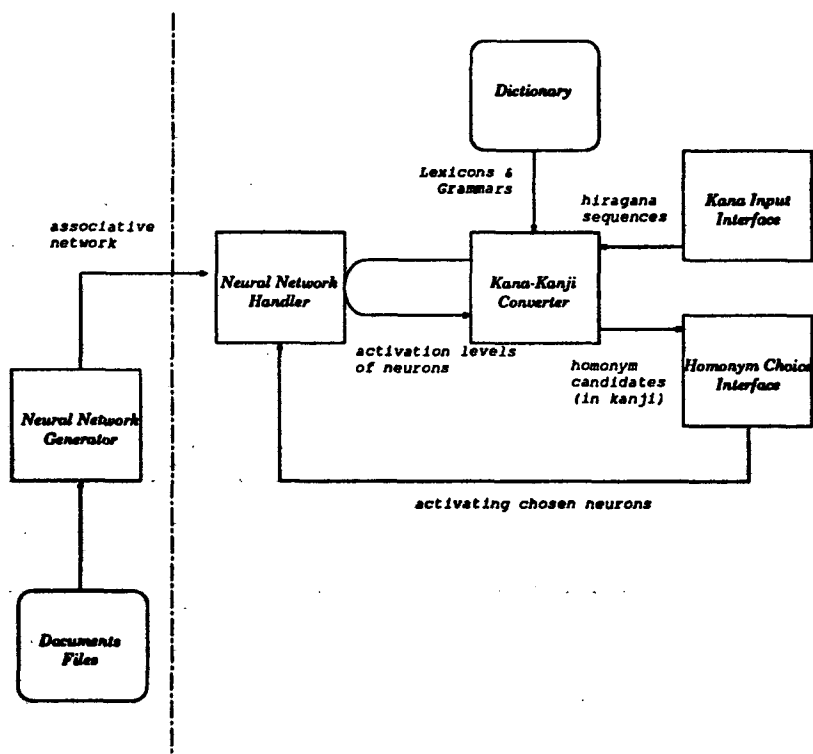
227

Figure 2: System Configuration

them as the strength of connection. A paragraph is recognized only by a format information of documents.

3. Sum up the strength of connection for each word-pair.

4. Regularize the training data; this involves removing low occurrences (noise) and partitioning the frequency range in order to obtain a monotonically decreasing (in frequency) training set.

Although the network data have only positive links and not all nodes are connected, non-connected nodes are assumed to be connected by negative weights so that the Hopfield conditions [Hopfield 84] are satisfied.

As described above, the technique used here is a morphological and statistical analysis. Actually this module is a pattern learning of co-appearing words in a paragraph.

The idea behind of this approach is that words that appear together in a paragraph have some sort of associative connection. By accumulating them, pairs without such relationships will be statistically rejected.

From a practical point of view, automated network generation is inevitable. Since human word association differ by individ-

228

ual, creation of a general purpose associative network is not realistic. Because the training data for the network is supposed to be supplied by users' documents in our system, automatic network generation mechanism is necessary even if the generated network is somewhat inaccurate.

- Neural Network Handler

The role of the module is to recall the total patterns of co-appearing words in a paragraph from the partial patterns of the current paragraph given by a user.

The output value O_j for each node j is calculated by following equations.

$$O_j = f(n_j)$$

$$n_j = (1 - \delta)n_j + \delta(\sum_i w_{ji}O_i + I_j)$$

where

f : a sigmoidal function

δ : a real number representing the inertia of the network($0 < \delta < 1$).

n_j : input value to node j.

I_j : external input value to node j.

w_{ji} : weight of a link from node i to node j; $w_{ji} = w_{ij}, w_{ii} = 0$.

The external input value I_j takes a certain positive value when the word corresponding to node j is chosen by a user. Otherwise zero.

Although the module is software implemented, it is fast enough to follow the typing speed of a user. [2]

- Kana-Kanji Converter

---

[2]A certain optimization technique is used respecting for the sparseness of the network.

The basic algorithm is almost same as the conventional one. The difference is that homonym candidates are sorted by the activation levels of the corresponding nodes in the network, except when local constraints such as word co-occurrence restrictions are applicable to the candidates. The associative information also affects the preference decision of grammatical ambiguities.

5 Evaluation

To evaluate the method, we tested the implemented sytem by doing kana-kanji conversion for real documents. The training data and tested data were taken from four types of documents: business letters, personal letters, news articles, and technical articles. The amount of training data and tested data was over 100,000 phrases and 10,000 phrases respectively, for each type of document. The measure for accuracy of conversion was a reduction ratio(RR) of the homonym choice operations of a user. For comparison, we also evaluated the reduction ratio(RR') of the kana-kanji conversion with a conventional context holding mechanism.

$$RR = (A - B)/A$$

$$RR' = (A - C)/A$$

where

A : number of choice operations required when an untrained kana-kanji converter was used.

B : number of choice operations required when a NN-trained kana-kanji converter was used.

C : number of choice operations required when a kana-kanji converter with a conventional context holding mechanism was used.

The result is shown in Table 1. The advantages of our method is clear for each type

Table 1: Result of the Evaluation

document-type	$RR(\%)$	$RR'(\%)$
business letters	41.8	32.6
personal letters	20.7	12.7
news articles	23.4	12.2
technical articles	45.6	40.7

of documents. Especially, it is notable that the advantages in business letter field is prominent, because more than 80% of word processor users write business letters.

6 Discussion

Although the result of conversion test is satisfactory, word associations by neural network are not human-like ones yet. Following is a list of improvements that many further enhance the system:

- Improvements for generating a network

  The quality of the network depends on how to reduce noisy word occurrence in the network from the point of view of association. The existence of noisy words is inevitable in automatic generation but plays a role to make unwanted associations. One approach to reducing noisy words is to identify those words which are context independent and remove them from the network generation stage. The identification can be based on word categories and meanings. In most cases, words representing very abstract concepts are noisy because they force unwanted activations in unrelated contexts. Therefore they should be detected through experiments. Another problem arises because of the ambiguity of morphological analysis. Word extraction from real documents is not always correct because of

the agglutinative nature of the Japanese language. Other possibility for network improvement is to consider a syntactic relationship or co-occurrence relationship while deciding link weights. In addition, there are keywords in a document in general which play a central role in association. They will be reflected in a network more in consideration of technical terms .

- Preference decision in kana-kanji conversion

  The reinforcement of associative information complicates the decision of homonym preference in kana-kanji conversion. We already have several means of semantic disambiguation of homonyms: co-occurrence restrictions and selectional restrictions. As building a complete thesaurus is very difficult, our thesaurus is still not enough to select the correct meaning(kanji-conversion) of kana-written word. So selectional restrictions should be weak constraints in homonym selection. In the same vein, associative information should be considered a weak constraint because associations by neural networks are not always reliable. Possible conflict between selectional restrictions and associative information, added to the grammatical ambiguities remaining in the stage of homonym selection, make kanji selection very complex. The problem of multiply and weakly constrained

homonyms is one to which we have not yet found the best solution.

7 Conclusion

This paper described an association based natural language processing and its application to kana-kanji conversion. We showed advantages of the method over the conventional one through the experiments. After the improvements discussed above, we are planning to develop a neuro-word processor available in commercial use. We are also planning the application of the method to other fields including machine translations and discourse analyses for natural language interface to computers.

References

[Amano 79] Kawada, T. and Amano, S., "Japanese Word Processor," *Proc. IJCAI-79*, pp. 466–468, 1979.

[Barwise 83] Barwise, J. and Perry, J., "*Situations and Attitudes*," MIT Press, 1983.

[EDR 90] Japan Electronic Dictionary Research Institute, "Concept Dictionary," Tech. Rep. No.027, 1990.

[Hopfield 84] Hopfield, J., "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons," *Proc. Natl. Acad. Sci. USA 81*, pp. 3088–3092, 1984.

[Kamp 84] Kamp, H., "A Theory of Truth and Semantic Representation," in Groenendijk et al(eds.) "*Truth, Interpretation and Information*", 1984.

[Lenat 89] Lenat, D. and Guha, R., "*Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project,*" Addison-Wesley, 1989.

[Minsky 88] Minsky, M., "*The Society Of Mind,*", Simon & Schuster Inc., 1988.

[Rumelhart 86] Rumelhart, D., McClelland, J., and the PDP Research Group, "*Parallel Distributed Processing: Explorations in the Microstructure of Cognition,*" MIT Press, 1986.

[Waltz 85] Waltz, D. and Pollack, J., "Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation," *Cognitive Science*, pp. 51–74, 1985.