A Contextual Dynamic Network Model for WSD Using Associative Concept Dictionary

Jun Okamoto †, Kiyoko Uchiyama §, and Shun Ishizaki §

† Keio Research Institute at SFC
§ Graduate School of Media and Governance, Keio University

5322 Endo, Fujisawa-shi, Kanagawa-ken, 252-8520, Japan
E-mail: juno@sfc.keio.ac.jp, kiyoko@sfc.keio.ac.jp, ishizaki@sfc.keio.ac.jp

Abstract

Many of the Japanese ideographs (Chinese characters) have a few meanings. Such ambiguities should be identified by using their contextual information. For example, we have an ideograph which has two pronunciations, /hitai/ and /gaku/, the former means a forehead of the human body and the latter has two meanings, an amount of money and a picture frame. Conventional methods for such a disambiguation problem have been using statistical methods with co-occurrence of words in their context. In this research, Contextual Dynamic Network Model is developed using the Associative Concept Dictionary which includes semantic relations among concepts/words and the relations can be represented with quantitative distances. In this model, an interactive activation method is used to identify a word's meaning on the Contextual Semantic Network where the activation on the network is calculated using the distances. The proposed method constructs dynamically the Contextual Semantic Network according to the input words sequentially that appear in the sentence including an ambiguous word.

1. Introduction

Word sense disambiguation is one of the difficult problems in natural language understanding research because it needs contextual meanings to be solved. A lot of previous researches for such disambiguation problems have been using co-occurrence of words in their context. Several machine learning algorithms have been used based on co-occurrence information among words, such as Naive Bayes methods or Support Vector Machine (Murata, et. al. 2003). Effectiveness of neural network approaches to the word sense disambiguation has been suggested (Woltz and Pollack 1985, Takahashi 1995). Not only the neural network architecture but also large-scale machine readable dictionaries were exploited (Veronis and Ide, 1990).

In this paper, we proposed a Contextual Dynamic Network Model (Okamoto 2007), where a structure of the contextual semantic network is organized dynamically by depending on an input sentence. By using the Associative Concept Dictionary (Okamoto 2001) which includes ontological information about the ambiguous words, we can disambiguate not only homographic ideogram but also various word sense disambiguations by using the Contextual Dynamic Network Model.

2. Associative Concept Dictionary

Background knowledge is crucial for computers to understand the contents of the text as well as its syntactic or shallow semantic information from input texts. The Associative Concept Dictionary (hereinafter referred to as ACD) has been built based on the results of large-scale online association experiments (Okamoto 2001). The association experiment system can be used by many human subjects simultaneously in a campus network at Keio University. In these experiments, the stimulus words are nouns and are obtained from Japanese elementary school textbooks. The subjects were requested to associate words from the stimulus words with a given set of semantic relations, hypernym, hyponym, part/material, attribute, synonym, action and situation.

Table 1: An example of association experiment (Stimulus word is "dictionary" in Japanese)

Semantic relations	{Associated word Response time}
Hypernym	{Publication 7} {Book 12}
Hyponym	{English dictionary 6} {Japanese dictionary 12}
Part / Material	{Entry word 18} {Word definition 33} {Page 38} {Cover page 44}
Attribute	{Difficult 6} {Easy to understand 11} {Pleasant 16}
Synonym	{Encyclopedia 17}
Action	{Read 5} {Investigate 11} {Consult 15} {Search 19} {Buy 29}
Situation	{Library 6} {Book store 27}

Table 1 is an example of association experiments by a subject where the stimulus word is "dictionary". The numbers in the table show the duration in second spent for the association by the subject. It is duration between the starting time of the association and the ending one. For example, "publication" is a hypernym of "dictionary" which duration is 7 seconds. And the next associated word is "book" with 5 seconds, where 12 is the summation of the two durations.

The ACD has been constructed by using all of the associated concepts, which are connected to the stimulus words with distances calculated by the following method. The distances are obtained using a linear programming method (Okamoto 2001). It combines the following three parameters linearly: *F*, *S* and *T*. Let *F* be an inverse of frequency of an associated concept which is normalized using δ to control its maximum value, *S* be an order of associated word, *T* be a duration spent for the association. Next, two boundary conditions are defined such that one is for the shortest distance and the other for a comparatively long distance. When using the Simplex method, we found the first two parameters significant for the distance calculation and the third parameter to be neglected. Then, the distance *D(x, y)* between concepts, *x* and *y*, is shown by the following formula:

$$D(x, y) = 0.81F + 0.27S, \qquad (1)$$

$$F = \frac{N_x}{n_y + \delta},$$

$$\delta = \frac{N_x}{10} - 1 \qquad (N_x \geq 10),$$

$$S = \frac{1}{n}\sum_{i=1}^{n} s_i,$$

$$T = \frac{1}{n}\sum_{i=1}^{n} t_i \times \frac{1}{60},$$

where N_x denotes a number of all the subjects for a given stimulus word *x*, n_y denotes a number of subjects who associated word *y* with the same semantic relationship for a given stimulus, δ denotes a factor introduced to limit the maximum value of *F* to 10, s_i is an order of the association by a subject and t_i is a response time spent for each association.

The ACD is built using the quantified distances and is organized in a hierarchical structure in terms of the hypernym and hyponym. Attribute information is used to explain the features of the given word. In the association experiment, each stimulus word had 50 subjects who were students at SFC of Keio University. The number of stimulus words is currently 1100. Total number of associated words is about 280,000. And the number of associated words, when the overlapping words are not counted, is about 64,000 words. In Figure1, "chair" is a stimulus word for the association. "Furniture" is a higher-level concept of "chair". The numbers below <1> express frequencies of subjects who gave a same associated word, <2> an average of order of association, <3>an average of response time and <4> a conceptual distances.

(chair	<1>	<2>	<3>	<4>
(hypernym	↓	↓	↓	↓
(furniture	0.92	1.02	0.16	1.09)
(object	0.04	2.50	0.24	7.43))
(hyponym				
(sofa	0.48	1.92	0.42	1.96)
(rocking-chair	0.28	1.43	0.59	2.64))
(part/material				
(wood	0.60	1.20	0.14	1.52))
(attribute				
(hard	0.46	1.17	0.32	1.82))
(synonym				
(seat	0.02	1.00	0.15	8.37))
(action				
(sit down	0.70	1.03	0.15	8.37))
(situation				
(school	0.30	2.40	0.22	2.78)))

Figure1 Concept dictionary description for a stimulus word "chair" (a part of associated concepts are presented. The stimulus word and associated words are originally in Japanese)

3. Word Sense Disambiguation by Contextual Dynamic Network Model

This research proposes a Contextual Dynamic Network Model (hereinafter referred to as CDNM) to disambiguate word senses by using an interactive activation method in a contextual semantic network. The network is constructed by using the ACD which includes semantic relations and distance information among the words in the sentence. In addition, this network structure changes depending on the context of the words in the sentences which include a homographic ideograph. By using the dynamic feature, this method can disambiguate word senses based on the words located around the ambiguous words.

3.1 Construction of Contextual Semantic Network

The CDNM uses rich and dense information in the network with quantitative distances from the ACD for the word sense disambiguation as well as statistical co-occurrence information in their context. The following steps show a procedure in detail for this network construction.

- Part of speech information for words (nouns, adjectives, adverbs, verbs and so on) in an input sentence is obtained from morphological analysis by using ChaSen, most popular Japanese morphological analysis software.
- A contextual semantic network is constructed by extracting semantic relations among words in the input sentence from the ACD by using the information obtained from the morphological analysis.
  - ➢ When a noun in the input sentence is included in the ACD as a stimulus word, the noun is added to the network which starts from the

noun by tracing semantic relation paths until the accumulated distance becomes over a certain numerical level.

- When a word in the input sentence is found in associated words in the ACD and the stimulus word of the word is included in the nodes in the network, a semantic relation path from the associated word to the stimulus word is assigned to the network.
- When a stimulus word is a homographic ideograph and its associated words are associated from another homographic ideograph, an inhibitory link between the stimulus word and the associated word is added in the network
- Inhibitory links among stimulus words are added in the network when the stimulus words are homographic each other.

Let an input sentence be "the picture-frame of Picasso's picture dropped from the wall, and it struck my head and the forehead bled." In this sentence, "picture-frame" and "forehead" are English expressions which correspond to the homographic ideographs "額" in Japanese. This ideograph has two pronunciations /hitai/ and /gaku/, the former means a forehead of the human body and the latter has two meanings, an amount of money or a picture frame.
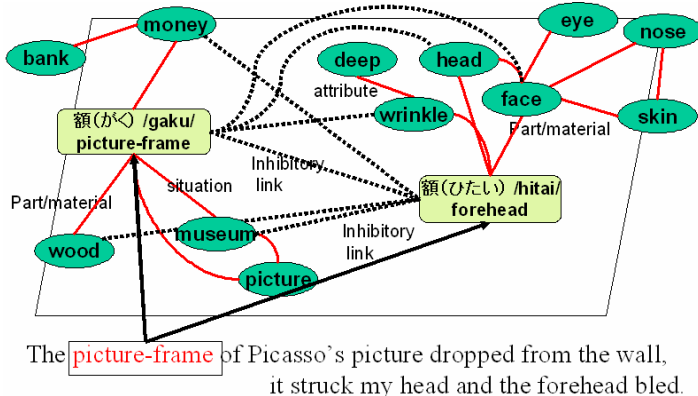

Figure 2 Example of Contextual Semantic Network including a homographic ideograph "picture-frame".

Figure 2 shows an example of a Contextual Semantic Network based on the ACD. "Picture-frame" is the first input word to the network. Square nodes in the figure are input words and are homographic in the sentence. The two square nodes are those of homological ideographs and are connected with an inhibitory link. Because a word sense has not being assigned to the homographic word, two or more nodes of homographic ideogram are added to the network at the same time. Oval shape nodes are added by using ACD and are connected with excitatory links. The thick lines connect the oval nodes with the word in input sentence. The dotted lines mean inhibitory links. The associated word nodes of a homographic ideograph are connected with other homographic ideographs with an inhibitory link. "Wood"

is a part relation concept of "picture-frame". "Museum" is a situation concept of "picture frame". "Eye" and "nose" are part relation concepts of "face". "Deep" is an attribute for "wrinkle". "Picture" and "face" don't exist in the sentence but is obtained from ACD.

Figure 3 shows an example of Contextual Semantic Network constructed from input word "head". "Human-begin" and "blood" are links added in the network. The broken lines mean excitatory links and the dashed lines mean inhibitory ones connected with nodes added from previous input words. The distances of their links are set comparatively long.

3.2 Dynamic Reconstruction of Contextual Semantic Network

In this research, a Contextual Semantic Network based on the ACD is reconstructed for each word in an input sentence. Human-beings can disambiguate word senses based on two words on the left and two words on the right of the ambiguous word (Choueka and Lusifnan 1985). When some homographic ideographs are in the sentence, our proposed method can assign appropriate word senses to the given ambiguous word in the sentence depending on its contextual meaning.
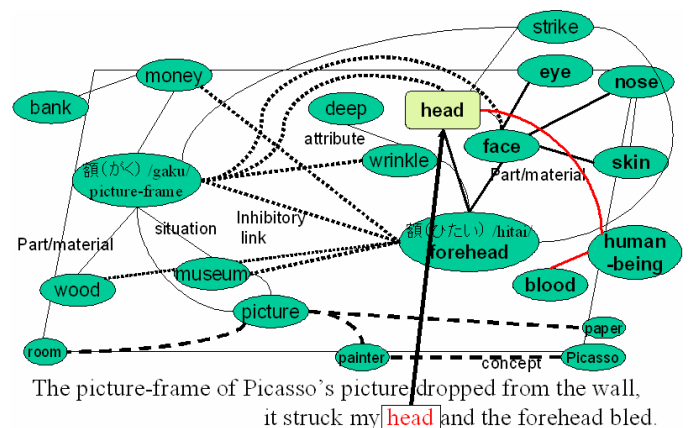

Figure 3 an example of a Contextual Semantic Network of "head"

3.3 Activation Calculation in the Network

The activation value of each node is calculated by an interactive activation model (Waltz and Pollack 1995) in the Contextual Semantic Network which reconstructed for each k-th word in an input sentence. We define the maximum activation level as 1.0. An initial value ($a_i(0)$) of k-th word in the input sentence is calculated by the following equation.

$$a_i(0) = (1.0 + S_{ki})/2 \ , \quad (2)$$

where 1.0 is a normalizing value. S_{ki} is an activation value of node i for k-th word that appears in the input sentence. The first activation value (S_{1i}) of word in the input sentence is 0. Next, the new activation value ($a_i(t+1)$) of each node N_i at time $t+1$ is calculated by

the following equation (3). The activation value is in the Contextual Semantic Network of k-th word in the input sentence.

$$a_i(t+1) = a_i(t) - \theta \cdot a_i(t) + \varepsilon_i(t) \quad , \quad (3)$$

where the decay parameter θ is assumed to be 0.1 and $\varepsilon_i(t)$ expresses influence of its neighbours at time t. When the neighbours of a node are active, they affect the activation value of the node by excitatory or inhibitory connections, depending on a link between two nodes. Those excitatory and inhibitory influences are combined by a simple equation (4) to yield a net input to the node. The net input is a weighted sum of its activation value of neighbour's nodes and weights assigned to the corresponding input connections. Thus, $n_i(t)$ represent the net input to the node by the following the equation.

$$n_i(t) = \sum \frac{a_j(t)}{\alpha \beta D_{ij}} \quad , \quad (4)$$

where $a_j(t)$ denotes an activation value of the node Nj connected with node N_i. α is a constant weight given by total number of links of the Contextual Semantic Network. β is a constant weight given sequential order of the Contextual Semantic Network for each work in the input sentence. D_{ij} denotes a distance between two nodes, N_i and N_j. In this paper, the Contextual Semantic Network is constructed by tracing semantic relation paths with accumulating the distance before exceeding the value of

5.0. Therefore, the value of D_{ij} with an inhibitory link is assumed to be -5.0. When the net input from neighbours is excitatory, that is $n_i(t) > 0$, the effect on the node, $\varepsilon_i(t)$, is given by the following equation.

$$\varepsilon_i(t) = n_i(t)[M - a_i(t)] \quad , \quad (5)$$

where M is the maximum activation level of the node and is set to 1.0. When the net input is inhibitory, that is $n_i(t) \le 0$, the effect of the node is given by the following equation.

$$\varepsilon_i(t) = n_i(t)[a_i(t) - m] \quad , \quad (6)$$

where m is the minimum activation level of the node and is set to be 0.

4. Experiments for WSD by the Proposed Method

For word sense disambiguation, we use the CDNM. This method uses an interactive activation method in a Contextual Semantic Network constructed from each word in the input sentence. We can assign appropriate word senses to the given ambiguous word by comparing some activation value of homographic ideograms. Several of Japanese ideographs in the stimulus words in the ACD have a few pronunciations. In the association experiments, such ideographs were presented as stimulus words followed with their pronunciations to avoid ambiguities.

In this sentence, "picture-frame" and "forehead" are homographic ideographs in Japanese. This model required about 10 second to calculate the activation value for each word in the input sentence on a 800 MHz Pentium III. The activation values finally settle down completely after 20 cycles
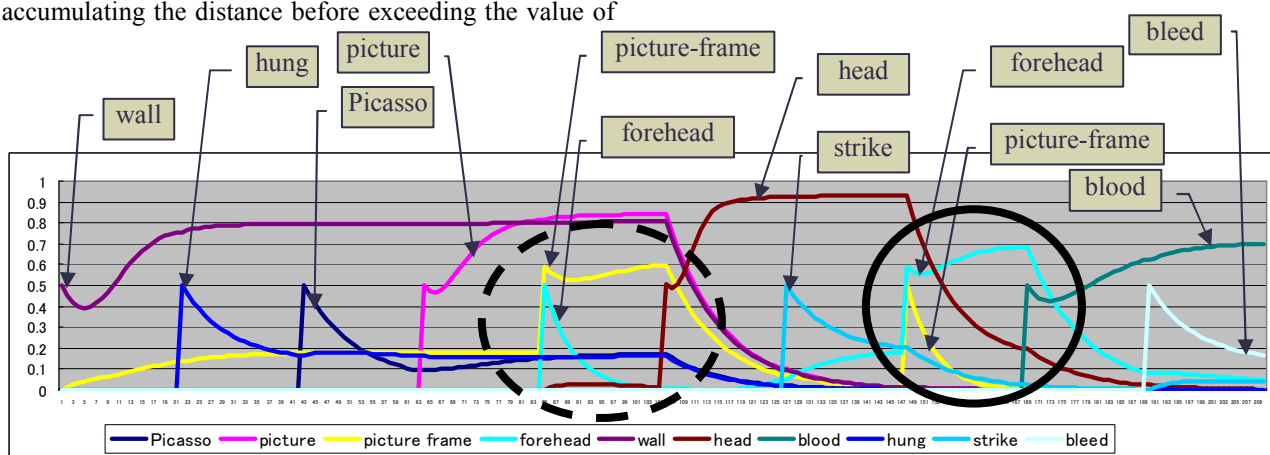

Figure 4 Activation Values for Selected Nodes in Sequential Input Word
The input sentence here is "The picture frame of Picasso's picture dropped from the wall, it struck my head and the forehead bled."

Figure 4 shows activation values of the homographic ideographs in the simulation where each input word spans 20 time cycles. The horizontal axis represents time, and the vertical axis represents activation values. The words in the rectangles are major words selected from the input sentence. At first, input words are "wall", "picture-frame"

and are connected in the contextual semantic network. The word order in the system follows the Japanese language one. The figure shows that activation values of "picture-frame" and "wall" increase slightly at the same time. For the homographic ideogram, "picture-frame" has bigger values than those of "forehead" (see broken-line circle in the figure 4). We can assign its appropriate pronunciation and meaning to "picture-frame" as those of Japanese ideograph. The homographic ideogram "forehead" has bigger values than those of "picture frame" (see thick-line circle in the figure 4). We can assign "forehead" to the Japanese ideograph.

The test data sets from large corpora are used to evaluate the interactive activation model based on the Contextual Semantic Network. The test data sets which include homographic ideograms "額" are extracted from a Japanese newspaper, Mainichi News Paper (from 1993 to 1995). These test sets have 24 sentences which include many stimulus words as much as possible. We use the CDNM to assign appropriate word senses to the given ambiguous homographic ideogram "額". We can assign the meaning by comparing two activation value of homographic ideogram node (額 /gaku/) or one of its node (額 /hitai/).

Table 2 The accuracy rate of homographic ideogram "額" in test data sets

	Number of sentences	Accuracy rate
Appropriate pronunciation	19	79.2%
Inappropriate pronunciation	5	
Total number of sentences	24	

Table 2 shows the accuracy rate of homographic ideogram's correct pronunciation in all the test data. In this CDNM, the activation values of nodes are calculated to each input word in the sentence sequentially. Concerning 5 sentences of inappropriate pronunciation, there are no modifier around the homographic ideogram "額". Therefore we cannot construct close network centred on this word in input sentence.

5. Future Work

We will use the framework of multinomial Naive Bayes text classification for word sense disambiguation. A pair of training and test data from large corpora is used to evaluate the interactive activation model based on the Contextual Semantic Network. The accuracy rate of homographic ideogram's correct pronunciation in all the test data are compared among the interactive activation model, the Naive Bayes method.

In this research, we proposed a method for disambiguation of pronunciations of homographic ideographs as well as the meanings by using the CDNM. The ideograph "額" has two major senses when its pronunciation is /gaku/. One is an amount of money and the other is a picture frame. It is necessary to carry out more experiments to disambiguate the two meanings. In the association experiments, stimulus words have to be presented unambiguously so that homonyms are presented with their meanings to avoid ambiguities. In the near future, we will conduct such association experiments to assign appropriate word senses with CDNM.

The ACD is a relatively small dictionary. We will extend it to a large-scale dictionary by extracting concepts from corpora automatically. This extension will be useful for higher level contextual understanding system such as WSD system, document summarization system or analyzer for free-answer questions of the questionnaires.

References

Choueka, Y. and Lusifnan, S. (1985) Disambiguation by short contexts, *Computers and the humanities*, Vol. 19, pp 147--157.

Matsumoto, Y., Kitauchi, A., Yamashita, T., Hirano, Y., Matsuda, H. and Asahara, M. (1999) Japanese Morphological Analysis System ChaSen Manual version 2.0 Manual 2nd edition (in Japanese)., *NAIST Technical Report*, NAIST-IS-TR99009 Nara, Institute of Science and Technology.

McCelland J.L. and Rumelhart D.E.(1981) An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings, *Psychological Rev.*, Vol.88, No.5, pp375--407.

Murata, M., Utiyama, M., Uchimoto, K., Ma, Q. and Isahara, H. (2003) CRL at Japanese dictionary based task of SENSAVAL-2 -- Comparison of various types of machine learning methods and features in Japanese word sense disambiguation -- (in Japanese)., *Journal of NLP* Vol.10 No.3.

Okamoto, J. and Ishizaki, S. (2001) Construction of Associative Concept Dictionary with Distance Information, and Comparison with Electronic Concept Dictionary (in Japanese)., *Journal of NLP* Vol.8 No.4, pp37--54.

Okamoto, J. and Ishizaki, S. (2005) Word Sense Disambiguation Using Contextual Semantic Network Using Associative Concept Dictionary, *SNLP 2005*, pp205--210.

Okamoto, J. and Ishizaki, S. (2007) Word Sense Disambiguation on Contextual Dynamic Network Using Associative Concept Dictionary, *PACLING 2007*, pp.93--100.

Veronis, J. and Ide, N. M. (1990). Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries, *Coling '90*.

Waltz, D. L. and Pollack, J. B. (1985). Massively parallel parsing: A strongly interactive model of natural language interpretation, *Cognitive Science*, Vol.9, pp.51--74.