Predication of Meaning of Bisyllabic Chinese Compound Words Using Back Propagation Neural Network

Lua Kim Teng
Department Of Information Systems and Computer Science,
National University of Singapore
Kent Ridge, Singapore 119260
luakt@iscs.nus.sg http://www.iscs.nus.sg/~luakt

Abstract

A three layer back propagation neural net is set up to study the functional dependency between the semantic class of a bisyllabic Chinese word and that of its two constituent Chinese characters. Simulations were performed using a three-layer back-propagation neural net with various combination of inputs. The inputs are (1) semantic classes of the constituent characters, (2) Entropy of the characters and (3) semantic strength[1] of the characters. Our simulations show that we can obtain the meaning class of a bisyllabic word from the meaning classes of its two constituent characters to an accuracy of 81% by taking the semantic classes and semantic strength of the characters as input. This research establishes the dependency between the meaning class of a Chinese compound word and that of its two constituent characters.

1. Introduction

The processing of the meaning of a word is a difficult job. In all existing text books on semantics[2-4], only descriptive treatments are provided for explaning the meaning of a word. There have been some attempts to break meaning into more fundamental units (Chapter 6 of [2]) but without much success. The main obstacle is that we are still unable to quantify meaning. As there is no proper *representation* of meaning, it is also difficult to *process* meaning.

In some of our earlier papers [5-7], we have proposed a scheme to quantify the change in the meaning when a bisyllabic Chinese compound word is built from two Chinese characters. We take the meaning of a word as one of the meaning class given in a dictionary of meaning classes: 《同义词词林》 [8]. This dictionary classifies 70,000 Chinese words into

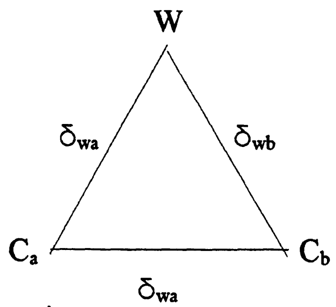

Figure 1 Semantic Distance Between Words and Characters

12 major, 94 medium and 1428 minor classes. A triangular figure is used to represent the change in meaning during a word formation process(See Figure 1). In Figure 1, W is the word, C_a, C_b are the characters forming the word, δ_{wa}, δ_{wb}, δ_{ab} are the so-called ***semantic distances*** between the word and the characters and between the two characters. The value of δ is computed according to Table 1.

Table 1: semantic distance δ

Type of distance	δ
same major, medium and minor class	0 (Hex 000)
same major class, medium class, different minor class	1 (Hex 001)
same major class, different medium class	2(Hex 010)
different major class	4(Hex(100)

We found that many Chinese bisyllabic words are formed as so-called *biased* type of words (these are the so-called 偏正、后补 type of words). Examples are(See Table 2):

Table 2 Semantic distances of some words

	W	C_a	C_b	δ_{wa}	δ_{wb}	δ_{ab}	δ_{wa}- δ_{wb}
微笑	Ic01	Ea03	Ic01	4	0	4	4
雪白	Ec04	Bf01	Ec04	4	0	4	4
立正	Fb03	Fb03	Ea09	0	4	4	-4
建成	Hc05	Hc05	If22	0	4	4	-4

Traditionally, 微笑、雪白 are classified as 偏正词组(biased compound word) and 立正、建成 are 后补词组(complement compound word).

Our question here is that: Is there any way to predict where the meaning of a word will fall, based on the meaning of the characters that it is made up? In [5], we derive a parameter called ***semantic strength*** to do such a prediction. The *semantic strengths* for the characters, a and b are computed as:

$$S_a = 4 - \delta_{wa}$$

$$S_b = 4 - \delta_{wb}$$

As a character will normally combine with a large number of other characters to form words, there are many values of S_a and S_b. We can therefore calculate the overall ***semantic strength*** by averaging all the S_a and S_b. This is:

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

We can consider s one of the basic semantic attribute of a character. We computed s for all characters and use it to predict the meaning of the word. In [5]. It was found that when $s_a > s_b$, the word formed (W= ab) will take a meaning class that is closer the meaning of the first character. This prediction is followed for 73.8% of the total cases. It is therefore obvious that the meanings of the characters, in one way or another, determine the meaning of the words that they form.

In this paper, we will like to further investigate into the various other factors that determine the meaning of the compound words. These include: (i) meaning classes of the characters, (ii) entropy values of characters, and (iii) semantic strength of characters. The simulations is performed on a three-layer back propagation neural net (BPNN). BPNN is being selected as it has the ability of producing fairly complicated functional parameters between its input and output neurons.

The following sections of this paper cover: In Section 2, we give a brief description on the BPNN that is built for the prediction of meaning of bisyllabic words. In Section 3, we present the results obtained for various sets of input. The training and testing if the BPNN is given in Section 4 and In Section 5, we present our conclusion and some proposals for further research in this area.

2. A Three Layer Back Propagation Neural Net For Simulation

The general structure of a 3 layer BPNN is shown in Figure 2. The number of neurons at the input layer depends on the type of simulation it is running(See Table 3). For example, there are 12 major semantic classes for each Chinese character. The total number of input neurons will be 24(as there are 2 characters). When a character belongs to semantic class A is presented at the input, the first neuron will be set to a value 1 and the other 11 neurons are all set to 0. For entropy and

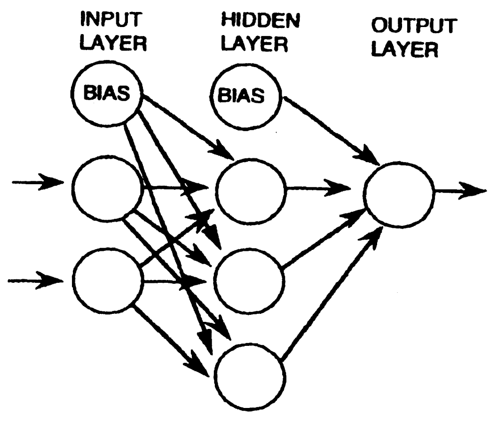

Figure 2 Three layer BPNN

strength, as these are single values, they are scaled to values between 0 and 1 before inputting to the input neuron. The scaling factors are 24 and 8 for entropy and strength respectively.

S = int (Strength + 4.5)/8
E = Entropy /24

For compatibility, we fix the number of neurons in the hidden layer to be either 12 or 2.

There is only one neuron at the output layer. During training, this is set to 1 when $\delta_{wa} > \delta_{wb}$ and to 0 when otherwise. At the testing stage, we will set $\delta_{wa} > \delta_{wb}$ if the final output is greater than 0.5 and $\delta_{wa} < \delta_{wb}$ if it is smaller than 0.5. These results of the BPNN simulations are shown in Table 3.

Table 3 Results of BPNN simulations

Type of Simulation	No of class/ value	No Input Neu-rons	No of Hid-den Neu-rons	Predi ction In-sider	Predi citon Ou-tsider
Semantic Class, Major	12	24	12	0.65	0.65
Semantic Class, medium	94	188	12	0.70	0.70
Entropy	1	2	2	0.59	0.59
Strength	1	2	2	0.67	0.67
Semantic Class, Major + Entropy	13	26	12	0.61	0.61
Semantic Class, Medium + Entropy	95	190	12	0.72	0.72
Semantic, Major+Strength	13	26	12	0.67	0.67
Semantic Class, Medium+Strength	95	206	12	0.81	0.81
Entropy+Strength	2	4	2	0.65	0.65

Selection of sample size

One problem with neuron net simulation is that if too few samples are presented for the training, the systems (through the adjustment of the link weightages) eventually memorize the whoe sample and there is no generalization at all. As a result, the predictability of the system will be greatly deteriorated. To prevent this, we have to select a sample size that is several times larger that the number of connections in the BPNN system. In this research, the largest net we build is for the simulation of medium semantic classes and strength as input. There are 190 inputs. The total number of connections in the BPNN system will be 190x12x1=2280. We thus select 5,000 samples from the dictionary for the training and testing. This same set of 5,000 samples provides input and output data for the so-called ***insider testing.*** Another 5,000 samples are again selected from the dictionary to test the predictability of the BPNN. This is the so-called ***outsider testing***. During an insider testing, the system performance is tested with data that used to derive the system parameter. The same set of data is used for training and testing. This provide an internal consistency checking. For outsider testing, the system is tested using data that are used in the training. The outsider testing provide more objective assessment to the system.

3. Training and Testing Data

The words and word classes used this research are directly derived from the《同义词词林》which contains 39,554 bisyballic words. Inside, 27,279 words or 69% are the *biased* type. We select them according to the criterion that $\delta_{wa} <> \delta_{wb}$. If $\delta_{wa} = \delta_{wb}$, there is no *bias* of the meaning class.

We found that for most of the time, there are more than one meaning class associated to each word and characters. There has to be a way of disambiguity. We adopt a simplest approach for this job. This is the so-called *closest meaning class* approach. In this apart, we compare three sets of meaning classes, $\{M1_w, M2_w, ...Mn_w\}$, $\{M1_a, M2_a, ...Mp_a\}$, $\{M1_b, M2_b, ...Mq_b\}$, where $M1_w$ is the first meaning class of word w and $M1_a$, $M1_b$ are the first meaning class of character a and b correspondingly. There are altogether npq possible ways of paring them up. For all possible pairings, we compute the semantic distances, i.e., δ_{wa} and δ_{wb} and select the one with minimum value of δ_{wa} and δ_{wb}.. This approach, though basically an *ah hoc* approach that based on intuition, works well for at least 90% of the time.

4 Discussion of the Results

First, we compare the prediction power between the semantic major class and the medium class. There are almost 8 times more input in medium class simulation. The prediction power is also higher, whether it is applied alone or combined with strength or entropy. So, it is obvious that a more detailed semnatic classification helps the predicition of the semantic class of the word formed.

Next, we look at the comparative merit of entropy and strength. It is obvious that strength has a much higher prediction power.

Combining the above two observations, we come to a conclusion that the best prediction could be obtained by feeding in the semantic medium class and strength value to the BPNN. The best result achieved so far is 81% of correct prediction.

This result is of important significance in the study of semantics for the following reasons:

(i) We have firmly established a functional relationship between the meaning of a words and those of its constituent characters. We should have a 50% success rate only if there is no functional dependency between the two. It looks a common sense from our every day's experience that the meaning of a word should be related to the meanings of its characters. But this is the first time that we establish a mathematical (functional) relationship in between the two. It thus open up a new channel for more research to be performed along this direction.

(ii) The way that the words are classified has a direct impact on the predictability of the BPNN. A more detailed classification helps improving the hit rate of the prediction. There is an overall increase of 5-10% hit rate when we increase the number of classes from 12 to 94. If the way of classification plays a significant role in the predictability of the meaning of the words formed, then we should do more study on the classification (and also representation) of meaning.

5. Conclusion

We simulate the prediction of meaning of a word using a three layer back propagation neural net. It is found that, using 94 semantic classes and a strength value, we can predict the meaning of a word with an accuracy of 81%. This finding is of significance as it has established the strong evidence that the meaning of a word can be predicted from the meanings of its constituent characters.

There are also some inadequacy in this research.

First, the current work focuses on the direction of movement of the meaning. We can only predict whether a word has a meaning closer to its first or second character. In the subsequent works, we are going to study how to predict the actual semantic class of a word.

Another research direction is to refine the meaning classes of the 《同义词词林》. This dictionary, as it is composed for the use of human language translation, may not have arranged all words in the best way suitable to our semantic processing. Also, as our simulation shows a strong correlation between the way that the words and characters are semantically classified and the accuracy of the prediction. More research should thus be conducted to look for an even better classification scheme for word senses.

References

[1] Lua K T, Chinese word semantics and its prediction, Computer Processing of Chinese & Oriental Languages, Vol 7, No 2, P167-190. (The terms 'semantic transformation vectors' and 'semantic strengths' are defined in this paper.)

[2]杰弗里。利奇，(Leech, G. N.), 《语义学》上海外语教育出版社, 1987.

[3] 刘叔新，《汉语描写词汇学》，商务印书馆，1990

[4] James Allen, Natural Language Understanding, Addison-Wesly, 1987

[5] Lua, K T, Chinese word semantics and its predication. Computer Processing of Chinese & Oriental Languages, vol.7, 2 (December 1993): 167-190.

[6] Lua, K T, A theory of semantic categorization of symbol language and its application to Chinese language. Computer Processing of Chinese & Oriental Languages, vol.7, 2 (December 1993): 273-284.

[7] Lua, K T, A study of Chinese word semantics. Computer Processing of Chinese & Oriental Languages, 7, no.1 (1993): 37-60.

[8]梅家驹、竺一鸣、高蕴奇、殷鸿翔，《同义词词林》，上海辞书出版社，1983。