Self-organizing Map Analysis of Conceptual and Semantic Relations for Noun

Minghu Jiang[1,2], Chengqing Zong[3], and Beixing Deng[4]

[1] Lab of Computational Linguistics, Tsinghua University, Beijing, 100084, China
[2] Creative Base of Cognitive Science, Tsinghua University, Beijing, 100084, China
jiang.mh@tsinghua.edu.cn
[3] State Key Lab of Pattern Recognition, CAS, Beijing, 100080, China
[4] Dept. of Electronic Eng., Tsinghua University, Beijing, 100084, China

Abstract. In this paper, we analyzed self-organizing map of conceptual and semantic relations for noun, discussing the semantic distinction between conceptual nouns for natural language processing and syntax acquisition, summarizing the lexical meaning and a detailed description of semantic lexical tagging of nouns. Our result reflects the noun-attribute associations and focuses on the conceptual relationships. By using several features, map models provide an operational definition of the conceptual nouns distinction.

1 Introduction

An important feature of human intelligence is the use of language consisting of symbols (words). Many researchers are exploring the lexical representation and operation in human brain and seeking their construction principle and computational theory. It is very useful for natural language understanding if there is semantic understanding of relevance. We need to know how the learning of concepts is simulated, how conceptual and semantic relations can be mapped with the aid of neural networks, and how the multi-disciplinary perspectives of cognitive neuroscience and experimental psycho-linguistics are used to construct neural network models to deal with the hierarchical structures based on the grammar and semantic rules. As a rule, the semantic features of conceptual nouns are described in high dimension vectors. For expedient comparison of neuroimaging localization of cognitive operations [1] and syntactic processing with pattern recognition technology, we adopted the feature compression technologies which map high dimension features to low dimension, and hold enough main information to distinguish the sorts among conceptual nouns. Multi-principal component analytical method is the classical statistical technology of data analysis and feature compression, which is of the function of extraction principal features, restraint noisy and drop dimension [2]; Kohonen's SOM (self-organizing map) nonlinear drop dimension processing enables weight vector to approximate the probability distribution of feature data and displays a topography structure ordinal logic diagram in a 2-dimension array plane [3], which displayable analysis for the feature data can reveal the similar extent among objects, relation structure and semantic topology relation.

J. Wang, X. Liao, and Z. Yi (Eds.): ISNN 2005, LNCS 3498, pp. 977–982, 2005.

2 Feature Data Compress and Map

2.1 The Self-organized Map [3]

Kohonen's SOM is based on research of physiology and brain science [3], the self-organized learning of which enables the similar kinds of nerve cells in function to be nearer and the different kinds of nerve cells in function to be more separate. SOM consists of input layer and output layer, the training of weight $W_{i,j}$ is finished by competitive learning. The black nerve cell r in the center is the winner, around r the nerve cells within the area $N(r)$ obtain excitement in different degree, the nerve cells besides $N(r)$ are restrained, as shown in Fig. 1(a).

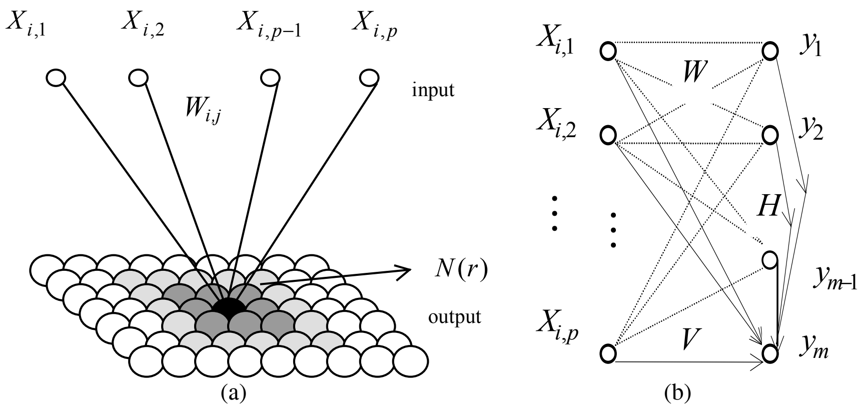

Fig. 1. (a) SOM Neural Network (b) APEX Neural Network.

When t=0, input classless sample $X = \{X_i \in \Re^p : i = 1,2,\cdots,n\}$, initial weight is put: $\{W_{i,j}, i, j=1,2,\ldots,m\}$. When t<$T_{\max}$, randomly select $X_i(t)$ in X set:

$$\text{Find out: } r = \arg\min_s \{\|X_i(t) - W_s(t)\|\} \tag{1}$$

$$\begin{aligned} \text{Iteration: } W_s(t+1) &= W_s(t) + \alpha_t \cdot e^{-dist(r,s)^2/\sigma_t^2}[X_i(t) - W_s(t)], \quad \forall s \in N_t(r), \\ W_s(t+1) &= W_s(t), \quad \forall s \notin N_t(r). \end{aligned} \tag{2}$$

$$\text{Update: } t{+}{=}1,\ N_t = N_0 - t(N_0 - 1)/T_{max},\ \alpha_t = \alpha_0(1 - t/T_{max}),\ \sigma_t = \sigma_0 - t(\sigma_0 - \sigma_f)/T_{max}. \tag{3}$$

Here, m is output array size, $T_{\max}$ is the max iterative number, N_0 is initial neighbor threshold, α_0 is initial learning rate, σ_0 and σ_f are the control parameter of step length, $dist(r, s)$ is a distance between node r and node s in the output array. $N(r)$ and α_t are iteration number t's monotony descend function.

2.2 Principal Component Feature Extraction

Adaptive principal component extraction (APEX) [2] remains eigen vectors which correspond to m maximum eigenvalues of covariance matrix of input data, canceling the eigen vectors which correspond to lesser eigenvalues, and maps n dimensions feature to m dimensions. Assumed that Mx is mean vector of input samples, the covariance (positive definite) matrix of samples is C_x, P order eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$, e_h is an orthogonal unit-eigen vector, $e_h^T e_h = 1$, $C_x e_h = \lambda_h e_h$, $h = 1, 2, \cdots, p$. APEX is its statistic property unknown by self-organnized learning to extract principal features and by weight of the former m-1 neural nodes to recursively calculate weight of m^{th} neural node, as shown in Fig.1 (b), which is shown as follows:

$$Y(t) = W(t)X(t) \tag{4}$$

$$y_m(t) = V(t)X(t) - H(t)Y(t) \tag{5}$$

here, input X is a p-dimension sample, $Y = [y_1, \cdots, y_{m-1}]^T$ is the former m-1 output nodes, $W = [e_1, \cdots, e_{m-1}]^T$ is a weight matrix which connects input X to the former m-1 output nodes, V is weight vector which connects input nodes to m^{th} neural node, H is weight vector which connects the former m-1 output nodes to m^{th} output node, $\lambda_i = e_i^T C_x e_i = E\{y_i^2\}$, when W converges, only V and H are trained, the t^{th} iteration is shown as follows [2]:

$$V(t+1) = V(t) + \beta(t)(y_m(t)X^T(t) - y_m^2(t)V(t)) \tag{6}$$

$$H(t+1) = H(t) + \beta(t)(y_m(t)Y^T(t) - y_m^2(t)H(t)) \tag{7}$$

Here, $\beta = 1/(M \cdot E\{y_{m-1}^2(t)\})$ (M is a constant). When W has converged to the former m-1 principal components $[e_1, e_2, \cdots, e_{m-1}]^T$, then V will converge to e_m. The training step is: m=1, V and H are iterated by Eqs(4)~(7) until the changes of V and H is less than a threshed, m is increased and continues iteration Eqs(6), (7) until the number of dimensions are satisfied. After convergence Y=[W, V]X maps p dimensions' vector to m dimensions' one.

2.3 Self-organizing Semantic Maps (SOSM) [5]

SOSM is self-organizing feature maps for the class-extended input feature [5]. Assumed that X is a p-dimension sample, after the classes are extended, the samples mode is: $\hat{X} = \{\hat{X}_i = \begin{bmatrix} \alpha X_{s,i} \\ X_i \end{bmatrix}\} \in \Re^{c+p}$. Here, X_i is the p-dimension mode vector, $X_{s,i}$ includes the class information (c classes), and $\alpha < 1$ is to weaken the importance of the class-extended information.

2.4 Self-organizing Maps for Detailed Feature (SOMDF)

Because the described noun attribute is quantized too rough for the training data of SOM, for example, although aircraft carrier (G) and frigate (I) are classified into one class, the difference of both volumes is very large. Therefore, their feature attributes should be described more detailed as shown in Table 1.

3 Experimental Results

Our feature description of 16 kinds of conceptual nouns in detail and experimental data for SOMDF are shown in Table 1, the other experiment data transform the data of Table 1 into binary values, and the parameters are set α_0 =0.9, σ_0 =4.0, σ_f =0.5, N_0 =9, m=10, $T_{\max}$ =10000, α =0.2.

During the learning process of SOM, weight updating is not only for the excited nerve cell, but also for those nerve cells within around N_r area at the same time, within which the more neighboring nerve cells can be reflected in the N_r area. Therefore, the network has a high ability to tolerate noisy and aberrance of the samples, which learning result enable the nearer samples of the input space to be mapped to nearer nodes in output layer. Our experimental results show that if the input samples have several classes, then according to their probability distribution, these samples are mapped to different area in output layer, one area representing one same class samples. We can learn from comparison as shown in Fig 2, APEX algorithm is similar to SOM for clustering ability, reflecting the high dimension features of the input samples are mapped to some low dimension spatial area and these similar features' samples in high dimension space are mapped to neighboring nodes in output layer. Because the conventional binary feature descriptions are too rough, many objects are mapped to the same output node, for example, aircraft carrier (G), chaser (H) and frigate (I), which features are described the same for the training data of conventional SOM, and output is mapped to the same node, as shown in Fig 2(a). The class extended information (16 classes) is added to the input features, as shown in Fig2(c), showing the same input features are also mapped to different output nodes. SOMDF can obtain fine classification for the described object as shown in Fig. 2(d).

Table 1. Detailed Feature Description for Output Objects (A: bicycle, B: motorcycle, C: microbus, D: coach, E: train, F: tank, G: aircraft carrier, H: chaser, I: frigate, J: huge passenger liner, K: passenger liner, L: lugger, M: helicopter, N: scout, O: huge airliner, P: airliner).

Feature Description \Output Object	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Small	0.8	1.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
Middle	0.0	0.0	0.0	1.0	0.0	0.9	0.0	0.0	0.0	0.0	1.2	0.0	0.9	0.9	0.0	0.0
(Very)Big	0.0	0.0	0.0	0.0	1.1	0.0	1.2	1.0	1.0	0.9	0.0	0.0	0.0	0.0	1.0	0.9
Seaway	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
Landway	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Airway	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0
Used in Army	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
Used in Civilian	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0
Drive	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0
Non-Drive	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
Cheap	0.8	0.9	1.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	00
Moderate	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.1	0.8	0.0	1.0	1.2	0.0	1.2
Expensive	0.0	0.0	0.0	0.0	0.9	0.0	1.2	1.0	0.9	0.0	0.0	0.0	0.0	0.0	1.1	0.0
High Speed	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.2	0.8
Middle Speed	0.0	1.2	1.1	1.0	1.1	1.0	0.8	0.9	1.1	0.8	0.8	0.0	1.1	0.0	0.0	0.0
Slow Speed	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	00

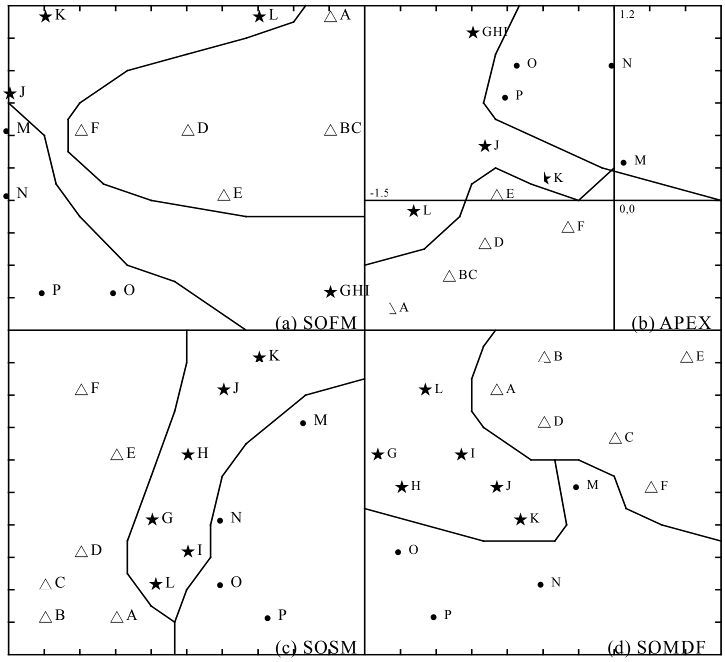

Fig. 2. Several SOMs of Conceptual and Semantic Relations for noun.

4 Conclusion and Discussion

Although the APEX algorithm is a non-linear processing, principal component feature extraction belongs to a linear operation, which only adopts a non-linear neural network method to realize its linear algorithm. The class extended feature of SOSM is a supervisory learning [5] which based on brain learning rules. A supervisory learning of class extended mode can improve systemic performance and increase systemic memory ability. Above experiments show that if the feature description includes its feature content, symbolic class information and its attribute relations, then after feature map the semantic relation can certainly be responded. Even if two objects' features are the same, but the class extended SOSM distinguishes them into different classes, and forces map to different output nodes. However, SOMDF is different from SOSM, the dimensions of the former input features are not increased, the objects can also be fine classified. Neural networks can imitate the organize structure and function mechanism of brain nerve in some extent, by learning to obtain the cognition of some impersonal objects. Because human brain can save plentiful information including speech, image, figure and text etc., and is good at integrating all information for judgment of fuzzy, misshapen and aberrance objects, to reach the aim of different objects correctly recognized. High-level cerebral running depends mainly on the extraction of concept which is expressed by sign and language. During learning process, sign is expressed by its content and similitude degree between signs which is expressed by similitude degree of their content. The feature description used in neural network models is based on digital information, needs to translate all kind of information from real world into a digital style. Human brain can deal with some feature descriptions by fuzzy or simulative mode, for example, big and small, gentle and simple etc. However, neural networks can only deal with this information by using a digital expression. Learning methods of attribute-based description's neural networks have a limited expression of the background knowledge and make the concept description language appropriate to natural language understanding in some degree. At present about research of feature expression is very insufficiency. The other way about research of all kind of algorithms is very plenty, we remedy the feature expressive shortcoming by algorithms' predominance. When there is a lack of the content tag, class information of signs and their attribute relation in feature expression, it is difficult to reach an expectant aim if we only depend on the algorithms' predominance.

References

1. Li, P., Jin, Z., Tan, L. H.: Neural Representations of Nouns and Verbs in Chinese: An FMRI Study, **21** (2004) 1533-1541
2. Kung, S. Y., et al.: Adaptive Principal Component Extraction (APEX) and Application. IEEE Transactions on Signal Processing, **42** (1994) 1202-1217
3. Kohonen T.: The Self-Organized Map. Proceedings of the IEEE, **78** (1990) 1464-1480
4. Dong, Z., Dong, Q. Hownet. Kim, P.: Theories of Early Language Acquisition. Trends in Cognitive Sciences, **1** (1997) 146-1535
5. Bezdek, J. C., Pal, N. R.: A Note on Self-Organizing Semantic Maps. IEEE Transactions on Neural Networks, **6** (1995) 1029-1036