論遞迴神經群網與計算認知過程(1 及 2)

Collective Recurrent Neural Networks and Computational Mental Process

計畫編號：NSC 89-2213-E-002-066、NSC 89-2218-E-002-039

執行期限：88 年 8 月 1 日至 90 年 7 月 31 日

主持人：劉長遠　　　台大資訊系教授

研究助理：蘇恩民、許惠玲、廖建凱、林書賢、賴治權

一、中文摘要

本計畫主要目的在運用類神經網路的計算能力，探討人類在語言、影像、及聲音的認知過程。

在語言方面，我們旨在設計一個 corpus-based 的方法，針對文學家的作品，根據前後文資訊，將每個字編碼，進而提供"語意檢索"的功能。語意檢索並不是利用關鍵字的統計來決定相似度，系統找出的文章段落中所包含的文字與使用者輸入的文字甚至可能沒有交集，但是兩者蘊含了最相近的語意結構。

在影像認知方面，我們研究了 Darwin V 系統，並根據這個系統的視覺架構發展的一個視覺系統，將之運用到文字辨識中，觀察他的學習能力。我們同時也採用了 Deborah Walters 教授所提出的 ρ -space 架構作為擷取視覺特徵的部分。

在聲音認知上，我們針對人類歌聲做分析及合成，利用類神經網路抓取人聲中的特質，以做為合成時的參數決定。此計畫已經執行兩年完成。

關鍵詞：遞迴類神經網路、Elman Network、語意編碼、語意檢索、達爾文學習法、歌聲合成、文字辨識。

Abstract

The purpose of this project is to investigate the mental cognition process in language, image, and voice using the computational ability of recurrent neural networks.

On language cognition, the main idea is to design a system for semantic search. We encode words according to contextual information. Search results may have nothing in common with words of query, but they share similar semantic structures.

On image cognition, we have studied the Darwin V system. Then we developed a visual system, based on the visual part of Darwin V, to emulate the visual neuron activity of brain, and apply it to the pattern recognition task. We've used the ρ -space structure developed by Deborah Walters in 1987 as feature extraction part in our system.

On cognition of voice, we use neural networks to obtain the features of the voice in songs. The features will be used in synthetic singing.

Keywords: Recurrent Neural Networks, Elman Network, Semantic Coding, Semantic Indexing, Darwin Learning, Synthetic singing, Character recognition.

二、結果與討論

The information in literary works is rich. The task to read between the lines is challenging even for the most sophisticated system such as human brains. The main idea of this project is to design a corpus-based method to find semantic structures. We develop the "semantic encoding" of words based on their contextual information. A system for "semantic search" is further designed to transform the words or sentences users input into semantic encoding, and find the passages with similar semantic structures--even though the passages found may have no word in common with the input wording.

We investigate two semantic encoding approaches. The first encoding scheme is “semantic bit encoding”, which means the semantic information is implicitly represented by the bits of the encoding. We use three algorithms and their combinations to develop the semantic encoding, including Non-negative Matrix Factorization [1], Semantic Associative Search [2], and Elman Network [5].

We use the works of Bertrand Russell to train the Elman Networks and form the data matrix in NMF and Semantic Associative Search. When the training completes, we can analyze the results to get the encoding.

The second encoding scheme is “semantic word encoding”. Here we use the works of Mark Twain, who was famous for his extraordinary sense of humor and social concern. Self-organizing Map (SOM) is applied to investigate the relations among words in Mark Twain's vocabulary. Clustering according to their context vectors, words are encoded based on their unique context structures.

In this project we are trying to emulate the visual brain function and trying to build a system that use this technique to recognize characters or letters.

The techniques about pattern recognition are well developed. Most of the techniques are developed using the skills of graphic analysis, such as statistics and synthesis. But we are using neural network instead. Inspired by the idea of emulating the human brain, we think that using neural network to emulate visual neuron pathway is a fundamental for emulating the whole brain, a truly artificial intelligence.

At first we have studied the Darwin V system developed by Neuroscience Institute in 1998. An environment, a robot, and a control system compose this system. They use the key idea “emulating brain function” to build the control system, and train it. Their result was really amazing: the system learns how to distinguish objects by the different pattern on it after a relative short period of training. Our system is based on its visual part.

Our system uses two-layer

structure: the elementary feature extraction layer and the higher-order recognition layer.

In the elementary feature extraction layer, we use ρ-space structure that is developed by Deborah Walters in 1987. ρ-space is excellent in finding line sections with different orientation.

The higher-order recognition layer is composed by so called BCM neurons. BCM neuron was brought up by Bienenstock, Cooper, and Munro in 1982. It's a neuron model describe the modulation of synapses. This model is based on temporal competition or pattern competition.

The output data of ρ-space layer is connected to higher-order recognition layer directly. We've used the standard nglish letter 24x24 bitmap font as training sample, and use handwriting letters to test if the system could recognition the handwrites as the right corresponding standard letters.

In the latter part, we have also introduced a theory of our brain (cerebral cortex), which is by Robert Hecht-Nielsen. The theory is not so certain, but it may provide us a different view of our brain.

In this project we use self-organization neural networks to analyze and extract the delicate features in the voice in songs. Once we have classified the voice features, we can use fewer data to express the songs. Compared with the conventional approach which is database-oriented, neural networks can recognize and classify the phonemes automatically.

Once we have the available data, we can train a nonlinear learning system. The input samples are the extracted features and the phonemes. The corresponding output is the frequency. When the training is complete, we have the relationships between the features and the frequency and we can use the relationships to automatically synthesize the voice.

三、參考文獻

[1] D. D. Lee and H. S. Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401, 788-791 (1999).

[2] Naofumi Yoshida, Yasushi Kiyoki and Takashi Kitagawa, ``An Associative Search Method Based on Symbolic Filtering and Semantic Ordering for Database Systems," Data Mining and Reverse Engeenering: Searching for Semantics, Part Two, Chapter 6, Spaccapietra Stefanno (Editor), Fred Maryanski (Editor), Kluwer Academic Publishers, pp.105-128, 1999.

[3] J. Shavlik, S. Calcari, T. Eliassi-Rad, & J. Solock (1999). An Instructable, Adaptive Interface for Discovering and Monitoring Information on the World-Wide Web. Proceedings of the 1999 International Conference on Intelligent User Interfaces, Redondo Beach, CA.

[4] G. G. Towell & J. W. Shavlik

(1994). Knowledge-Based Artificial Neural Networks. Artificial Intelligence, 70, pp. 119-165.
[5] Elman, J.L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.
[6] Exercises in Rethinking Innateness. A Handbook for Connectionist Simulations by Kim Plunkett and Jeffrey L. Elman.
[7] Elman, J.L. (1998). “Generalization, simple recurrent networks, and the emergence of structure,” In M.A. Gernsbacher and S.J. Derry (Eds.) Proceedings of the Twentieth Annual Conference of the Cognitive Science Society.
[8] Guido Deboeck, Teuvo Kohnen (eds). Visual Exploration in Finance: with Self-Organizing Maps. Springer-Verlag London 1998.
[9] Krista Lagus, Timo Honkela, Samuel Kaski, Teuvo Kohonen. WEBSOM for Textual Data Mining. Artificial Intelligence Review 13: 345-364, 1999.
[10] Teuvo Kohonen (ed). Self-Organizing Maps. Springer-Verlag Berlin Heidelberg 1995.
[11] R.F. Hadley, V.C. Cardei. Language acquisition from sparse input without error feedback. Neural Networks 12, 217-235, 1999.
[12] M.F. Porter. An Algorithm for suffix stripping. Program. Vol 14, no. 3. pp 130-137, july 1980
[13] W. B. Frakes. Stemming Algorithm. Data Structures and Algorithms. Prentice-Hall. 1992
[14] Frakes, W. B., "Stemming Algorithms", in Frakes, William B. and Baeza-Yates, Ricardo (Eds.), Information Retrieval: Data Structures and Algorithms, Englewood Cliffs, NJ: Prentice-Hall, 1992. pp. 131-160.
[15] Tony McEnery, Michael Oakes. Authorship Identification and Computational Stylometry. Handbook of Natural Language Processing. 545-562. Marcel Dekker, Inc. 2000
[16] Ido Dagan. Contextual Word Similarity. NLP Handbook 459-475. Marcel Dekker, Inc. 2000
[17] Dieter Merkl. Text Data Mining. Handbook of Natural Language Processing. 889-903. Marcel Dekker, Inc. 2000
[18] P.N. Johnson-Laird, Paolo Legrenzi, Vittorio Girotto, Maria S. Legrenzi. Illusions in Reasoning About Consistency. SCIENCE vol. 288, 21 APRIL 2000.
[19] Noam Chomsky. New Horizons in the Study of Language and Mind. Cambrage University Press UK. 2000.
[20] H.B. Barlow, T.P. Kaushal, G.J. Mitchison. Finding Minumum Entropy Codes. Neural Computation 1, 412-423. MIT, 1989.
[21] Cheng-Yuan Liou, Wen-Pin Tai, Maximum A Posteriori Restoration of Blurred Images Using Self-Organization, Journal of Electronic Imaging, vol 7(1), Jan, 1998.
[22] K. Lomax The Analysis and Synthesis of the Singing Voice, Oxford

University Ph.D, 1997.
[23] N.Gershenfeld, B.Schoner & E.Metois, Cluster-Weighted Modeling for Time-Series Analysis Nature,Vol.397 Jan.1999.
[24] Perry R. Cook, Singing Voice Synthesis : History, Current Work, and Future Directions, Computer Music Journal, 1996.
[25] Prame, E. Measurements of the Vibrato Rate of Ten Singers. JASA 96(4), 1979-1984., 1994.
[26] R. J. Mcaulay, T. F. Quatieri, Speech Analysis/Synthesis Based on a Sinusoidal Representation, IEEE trans. on ASSP-34 no.4, Aug 1986.
[27] R. Maher, J. Beauchamp, An Investigation of Vocal Vibrato for Synthesis, Applied Acoustics (30), 1990.
[28] X. Serra, J, Smith, Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition, Computer Music Journal, 1990.
[29] Cheng-Yuan Liou, and Hsin-Chang Yang, "Handprinted Character Recognition Based on Spatial Topology Distance Measurement," IEEE trans. Pattern Analysis and Machine Intelligence, vol. 16, 1994, pp.394-404
[30] Cheng-Yuan Liou, and Hsin-Chang Yung, "Selective Feature-to-Feature Adhesion for Recognition of Cursive Handprinted Characters," IEEE trans. Pattern Analysis and Machine Intelligence, vol.21, 1999, pp.184-191
[31] Gerald M. Edelman, George N. Reeke, Jr,, W. Einar Gall, Giulio Tononi, Douglas Williams, and Olaf Sporns, "Synthetic Neural Modeling Applied to a Real-world Artifact," Proc. Natl. Acad. Sci. USA, vol. 89, 1992, pp7267-7271
[32] Nikolaus Almassy, Gerald M. Edelman, and Olaf Sporns, "Behavioral Constraints in the Development of Neuronal Properties: A Cortical Model Embedded in a Real-world Device," Cerebral Cortex, vol. 8, 1998, pp. 346-361
[33] Deborah Walters, "Selective of Image Primitives for General-Purpose Visual Processing," Computer Vision, Graphics, and Image Processing, vol 37, 1987, pp. 261-298
[34] Elie L. Bienenstock, Leon N. Cooper, and Paul W. Munro, "Theory for the Development of Neuron Selectivity: Orientation Specificity and Binocular Interaction in Visual Cortex," The Journal of Neuroscience, vol. 2, 1982, pp. 32-48
[35] Robert Hecht-Nielsen, "A Theory of the Cerebral Cortex," Proceeding of the Fifth ICNIP, 1998, pp 1459-1464