

基于稀疏 RAM 的 N-tuple 神经网络模型^{*}

N-tuple Neural Network Model with Sparse RAM

彭宏京 陈松灿

(南京航空航天大学计算机科学与工程系 南京 210016)

Abstract WISARD system, which needs many hardware by virtue of every class having one discriminator device, is a N-tuple classifier with RAM without any ability of approximating continue functions. Kojcz. A et al. have presented a regression algorithm for the N-tuple neural network (NTNN) to be used to interpolate arbitrary functions by canceling any discriminator concept and derived a N-tuple kernel function to be applied to the general regression neural network (GRNN). However, when N is a very large number, many storages will be invalid and a hash technology has to be used. This paper presents a novel adaptive pattern recognition system-N-Tuple neural network model with sparse RAM which can be applied to pattern recognition as well as function approximation tasks. It is a general model to some extent in which both NTNN and SDM can be regarded as a special case. Finally, experiments have shown the ability of approximating functions for this novel model.

Keywords N-tuple, Sparse RAM, SDM, Function approximation, Neural network

1. 引言

N-tuple 结构首先由 Bledsoe & Browning (1959) 提出^[1], 并由英国帝国理工学院的 Alexander 教授等人通过用 RAM 式神经元代替常规的神经元, 利用 N 元 (N-tuple) 分解技术和可扩展的并行体系结构设计并实现了一个单层自适应模式识别系统 (WISARD)^[2], 使其具有大规模并行分布处理能力, 以及通用性和自适应性。适合于大维数输入模式和大样本集类识别问题, 并在工业零件和人脸识别方面获得了成功的应用。该系统的学习是通过样本的一次提交完成的, 因此速度快, 识别具有实时性且对已学样本具有完全的识别精度。但不足是: (1) 存在过泛化现象, 造成误识; (2) 因学习算法的限制, RAM 神经元的空间存在饱和问题; (3) 对每类模式分配一个分类器, 硬件成本高; (4) 对 N 元分解方法的敏感性; (5) 在模式识别中, 由于同类数据分布的凝聚性, 常规 RAM 中只有很小比例的存储单元被使用, 造成存储空间的很大浪费。

本文作者之一在文[3]中考虑到文[2]的不足, 用基于 SDM 思想的稀疏 RAM 代替常规 RAM, 设计了一个硬件可实现的双层自适应模式识别系统, 该系统缓解了前者的饱和问题, 扩充了 N 元值的选取范围,

并提高了系统的识别性能。但不足的是同 WISARD 一样仍然要为每类样本设置一个分类器, 且缺乏对连续函数的逼近能力。

文[4], [5]推广 WISARD 这种基于 RAM 的 N-tuple 分类器, 取消了分类器的设置, 用于函数内插。由于是基于常规 RAM 的, 同样存在上述 (1)、(2)、(4)、(5) 的缺点。

结合上述推广, 本文构造了一个一般性的新型自适应模式识别系统模型, 它既能用于模式识别也能用于函数逼近, 并且很大程度上可以克服上述模型的不足。

2. 相关工作的回顾

2.1 Kanerva 的稀疏分布存储器模型 (SDM)^[6]

一般地, 常规的 RAM 是存储位置的阵列, 其主要特征是在读写数据时, 对每一个输入地址, 仅有一个存储位置被选择。SDM 模型与常规的 RAM 主要区别在于它的稀疏性和分布性。如图 1 是 SDM 的用于联想记忆的模型。地址矩阵 A 包含 M 个地址, 这 M 个地址是从 2^N 全局地址中稀疏地选择得到的, 当输入地址向量 x 时, 在 A 中的所有与输入地址向量 x 邻近的地址均被选中, 如果视 N 位地址向量为 N 维地址空间中的

^{*} 国家自然科学基金资助项目 (编号: 69973021)。彭宏京 博士研究生, 研究领域为: 模式识别、图像处理和神经网络。陈松灿 博士生导师, 研究领域为: 神经网络、模式识别和 SVM。

点,则被选择的地址点稀疏分布于以输入地址点为中心,给定海明半径 r 的超球内。这些地址点选择向量 S 的相应位为1,其余位为0。数据就依 S 中的指示分布式地写入矩阵 C 中。读操作时,所有被选择的数据位置的内容按位相加,得到 sum 向量,最后,经过一个阈值运算得到输出数据。

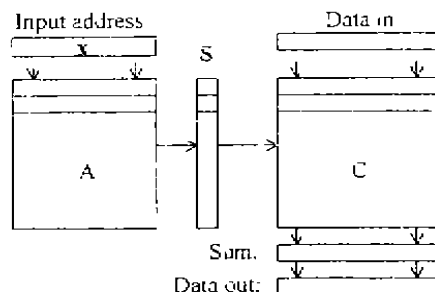


图1 SDM的自联想记忆模型

由于,联想记忆的SDM采用的是外积法,不能用于函数逼近,文[7]修改学习结构后推广了在函数逼近中的应用。

2.2 WISARD系统

该系统是一个由常规的RAM实现的单层自适应逻辑网络。输入模式 X 经二进制转换后,随机从中选取 N 个位形成地址,称为 N -tuple。这样,二值化后的模式空间被划分为若干个这样的 N -tuple,每个均与一个一位数据长的RAM相连,所有的RAM形成一个分类器。通过训练模式来设置RAM的数据位,被 N -tuple子模式选中(激活)的置1。未知模式激活分类器中的RAM,并将数据位求和输出,比较所有分类器的输出,具有最大输出响应的分类器所示的类别就是未知模式所属的类别。

文[3]用基于SDM思想的稀疏RAM代替常规RAM,降低了硬件开销,提高了识别精度,增大了 N 的选取范围。但是,还是为每个类保留了一个分类器。

文[4,5],取消分类器的设置后,修改系统的结构和学习算法使得可用于函数逼近。习惯上,称为逼近型 N -tuple网络。

3. 基于稀疏RAM的 N -tuple神经网络模型——SR-NTNN

3.1 SR-NTNN的体系结构的构造

本文提出的SR-NTNN系统结构如图2。它与文[4,5,7]逼近型NTNN不同的是用稀疏RAM代替常规RAM,而与文[3]的区别是取消了分类器的设置,因为它只能用于模式识别。

与WISARD系统一样,输入样本模式经二值化后

被映射为 n 个子模式,每个子模式大小 N 维,这个映射实际就是 N -tuple的采样过程。第 j 个 N 元子模式与 $A_{ij}(i=1, \dots, m)$ 均相连, $A_{ij}(j=1, \dots, n)$ 组成稀疏RAM神经元。每个 A_{ij} 与SDM中的稀疏地址矩阵相同,有 M 个单元, N 位字长。 W_{ij} 是与 A_{ij} 对应的权值矩阵, M 个单元,每个单元大小依问题而定。 A_{ij} 读写操作遵从SDM的概念,即当 N 元子模式提交给 A_{ij} 时,位于此子模式为中心、半径为 r 的海明球中的地址单元(A_{ij} 的行向量)处于激活状态,同时与 A_{ij} 相应的权值,读状态时,把稀疏RAM神经元的所有激活状态的权值经过某种运算后输出 Y 的一个分量。

3.2 SR-NTNN的学习算法及其特点

将要介绍的学习算法仍然采用一次提交完成的算法(ONE-SHOT),以保证学习的快速性。

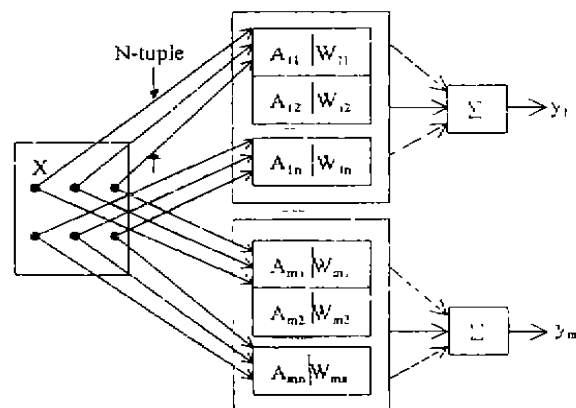


图2 基于稀疏RAM的 N -tuple神经网络模型的体系结构(SR-NTNN)

记训练模式空间为 D ,模式 $v \in D$ 的第 j 个 N -tuple二进制子模式记为 $v_j(j=1, \dots, n)$,记:

$$a_i(v_j, r) = \begin{cases} 1 & H(a_i^{(j)}, v_j) \leq r \\ 0 & H(a_i^{(j)}, v_j) > r \end{cases} \quad (1)$$

这里 $a_i^{(j)}$ 是 A_{ij} 中第 i 个二值地址向量, $H(\dots)$ 为海明距离, r 为存取半径。

在训练阶段,第 j 个 N -tuple对应的权值矩阵 W_{ij} 的第 i 个位置的权重设置为:

$$w_i^{(j)} = \sum_{v \in D} Y(v) \cdot a_i(v_j, r) / \sum_{v \in D} a_i(v_j, r) \quad (2)$$

则对测试模式 u ,网络输出为:

$$Y(u) = \frac{1}{n} \sum_{j=1}^n \left(\sum_{i=1}^M w_i^{(j)} \cdot a_i(u_j, r) / \sum_{i=1}^M a_i(u_j, r) \right) \quad (3)$$

从SR-NTNN的体系结构和学习算法分析,它有几个特点:

I. 当(2)式中的 $Y(v)$ 是类别指示器时(网络输出需要作必要的阈值运算),SR-NTNN用于模式分类

问题。

Ⅰ. 当取 $n=1$ 时, 即没有 N 元分解阶段, 整个模式样本 v 作稀疏分布存储, 得到 SDM 的函数内插模型。此时, $v_i = v_i(2)$ 式变为:

$$w_i = \sum_{v \in D} Y(v) \cdot a_i(v, r) / \sum_{v \in D} a_i(v, r) \quad (4)$$

测试样本 u 的输出为:

$$Y(u) = \sum_{i=1}^M w_i \cdot a_i(u, r) / \sum_{i=1}^M a_i(u, r) \quad (5)$$

对于模式识别问题, $Y(u)$ 是一个类指示器; 对于自联想记忆问题, 则有 $Y(v) = v$ 。

Ⅱ. 当 $M=2^N, r=0$ 时, 即是基于 RAM 的 N-tuple(NTNN)的函数逼近模型, 则(1)式变为

$$a_i(v_i, 0) = \begin{cases} 1 & H(a_i'', v_i) = 0 \\ 0 & \text{否则} \end{cases} \quad (6)$$

记

$$\eta_i(v) = \sum_{t=1}^{N-1} v_{a_{j,t}} \cdot 2^t \quad (7)$$

其中 $v_{a_{j,t}}$ 指模式 v 被划分到第 j 个 N-tuple 中的第 t 位的值。

式(7)实际上是将二进制子模式 v_i 化为十进制, a_i'' 也化为十进制数记为 η , 则式(6)即为:

$$\delta_{\eta, \eta_i(v)} = \begin{cases} 1 & \eta = \eta_i(v) \\ 0 & \text{否则} \end{cases} \quad (8)$$

则 N-tuple 的函数逼近算法为:

$$w_i'' = \sum_{v \in D} Y(v) \cdot \delta_{\eta, \eta_i(v)} / \sum_{v \in D} \delta_{\eta, \eta_i(v)} \quad (9)$$

$$Y(u) = \frac{1}{n} \sum_{j=1}^n w_{\eta_j''} u_{\eta_j}$$

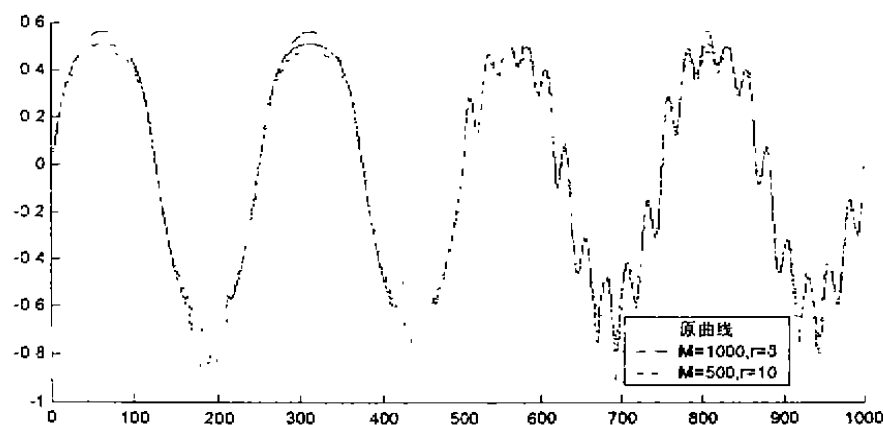


图3 SR-NTNN在稀疏RAM的地址个数不同、选取半径不同下的逼近效果

于稀疏RAM的N-tuple神经网络模型(SR-NTNN)。通过实验证实了它的逼近能力。优于现有结构的是大大减少了存储空间, 并通过调整选取半径 r 、预置地址数量 M , 可以达到理想的逼近效果。以后进一步的工作是继续探讨 SR-NTNN 在模式识别方面的应用及

3.3 逼近能力实验

不失一般性, 我们考察 $R^D \rightarrow R$ 的函数逼近情况, 例如: 非线性函数

$$g(x_1, \dots, x_5) = \frac{x_1 x_2 x_3 x_4 (x_5 - 1) + x_4}{1 + x_2^2 + x_3^2} \quad (11)$$

再生方程:

$$y_p(k+1) = g(y_p(k), y_p(k-1), y_p(k-2), u(k), u(k-1)) \quad (12)$$

这里 $y_p(k)$ 和 $u(k)$ 分别表示第 k 次的输出和输入。本网络选取 1000 个样本对来训练, 控制输入由下式给出:

$$u(k) = \begin{cases} \sin(2\pi k/250) & k \leq 500 \\ 0.8 \cdot \sin(2\pi k/250) + 0.2 \cdot \sin(2\pi k/25) & k > 500 \end{cases} \quad (13)$$

输入变量到二进制转换使用 thermometer 编码, 取 $N=31, n=16$, 各输入变量均量化为 $Q=512$ 等级, 稀疏RAM的地址分布采用均匀预置。如图3给出了 $M=1000, r=8$ 和 $M=500, r=10$ 时的逼近结果。我们看到 $M \ll 2^N$, 表明原 NTNN 的 RAM 存储单元浪费很大, 由于数据的凝聚性, 实际只需要部分存储单元。另外, 在新模型中, 可以适当增加存储单元数量来提高逼近的精度, 同时 r 也可以适当小些。在我们的实验中, 因为预置的地址过于稀疏, 所以 r 不能取太小 ($r > 5$), 而且我们看到随着 M 的成倍增加, 逼近精度不会显著提高。

结论 结合文[3~5]的改进, 我们提出了一个一般性的既可用于函数逼近, 也可用于模式识别的基

相关性能的研究。

参考文献

- 1 Bledsoe W, Browning I Pattern recognition and reading by machine. IRE Joint Computer Conference, 1959 225~232

(下转第16页)

- IEEE Intelligent Systems, 1999, 14(4): 44~54
- 3 Rennie J, McCallum A. Using reinforcement learning to spider the web efficiently. In: Proc. of the 16th Int. Conf. on Machine Learning ICML-99, 1999
- 4 Madria S K, et al. Research issue in web data mining. In: Proc. of Data Warehousing and Knowledge Discovery, First Intel. Conf., DaWak'99, 1999 303~312
- 5 Zainane O R, et al. Multimediaminer: a system prototype for multimedia data mining. In: Proc. ACM SIGMOD Int. Conf. On Management of Data, 1998. 581~583
- 6 Feldman R, Dagan I. Knowledge discovery in textual databases. In: Proc. of the first Int. Conf. On Knowledge Discovery and Data mining, Montreal, Canada, 1995. 112~117
- 7 Kosala R, Blockeel H. Web mining research: a survey
- 8 Kargupta H, Hamzaoglu I. Distributed data mining using an agent based architecture. In: Proc. of Knowledge Discovery and Data Mining: AAAI Press, 1997. 211~214
- 9 Dumais S, Platt D J. Heckerman Inductive learning algorithms and representations for text categorization. In: Proc. of the 1998 ACM 7th Int. Conf. On information and knowledge management, Washington United States, 1998. 148~155
- 10 Yang Y, et al. Pierce Learning approaches for detecting and tracking news events. IEEE Intelligent System, 1999, 14(4): 32~43
- 11 Billsus D, Pazzani M. A hybrid user model for news story classification. In: Proc. of the 7th Int. Conf. On User Modeling(UM'99), Banff, Canada, 1999
- 12 Hofmann T. The cluster abstraction model: Unsupervised learning of topic hierarchies from text data. In: Proc. of 16th Int. Joint Conf. On Artificial Intelligences IJCAI-99, 1999. 682~687
- 13 Nahm U Y, Mooney R J. A mutually beneficial integration of data mining and information extraction. In: Proc. of the 17th National Conf. On AI, 2000
- 14 Nigam K, Lafferty J, McCallum A. Using maximum entropy for text classification. In: Proc. of the IJCAI-99 Workshop on Machine Learning for information filtering. 61~67
- 15 Furnkranz J. Exploiting structural information for text classification on the www. In Advances in Int. Data Analysis, Third Int. Symposium, IDA-99, 1999. 487~498
- 16 Crimmins F, Smeaton A. Information discovery on the internet. IEEE Intelligent Systems, 1999, 14(4): 55~62
- 17 Singth L, Chen B, Height R. A robust system architecture for mining semistructured data. In: Proc. of the Second Int. Conf. On Knowledge Discovery and Data Mining, 1998. 329~333
- 18 Soderland S. Learning information extraction rules for semistructured and free text. Machine Learning, 1996, 34: 233~272
- 19 Florescu D, Levy A Y. Database techniques for the world wide web: A survey. SIGMOD Record, 1998, 27(3): 59~74
- 20 Abiteboul S, et al. The Lorel query language for semistructured data. Int. J. on Digital Libraries, 1997, 1(1): 68~88
- 21 Zauane O R, Han J. Resource and knowledge discovery in global information systems: A preliminary design and experiment. In: Proc. of the First Int. Conf. On Knowledge Discovery and Data Mining, Montreal, Quebec, 1995 331~336
- 22 Khosla I, Kuhn B, Soparkar N. Database search using information mining. In: Proc. of 1998 ACM-SIGMOD Int. Conf. On Management of Data, 1998
- 23 Meraldo P, Atzeni P, Mecca G. Semistructured and structured data in the web: Going back and forth. In: Proc. of Workshop on the management of Semistructured Data (in conjunction with ACM SIGMOD) 1997
- 24 Craven M, Slattery S, Nigam K. First-order learning for web mining. In: Proc. of 10th European Conf On Machine Learning, Chemnitz, 1998
- 25 Brun S, Page L. The anatomy of a large-scale hypertextual web search engine. In: 7th Int. World Wide Web Conf., Brisbane, Australia, 1998
- 26 Kleinberg J M. Authoritative source in a hyperlinked environment. In: Proc. of ACM-SIAM Symposium on Discrete Algorithms, 1998. 668~677
- 27 Chakrabarti S, et al. Mining the link structure of the world wide web. IEEE Computer, 1999, 32(8): 60~67
- 28 Spertus E. Mining structural information on the web. In: Proc. of the sixth Int. World Wide Web Conf. 1997. <http://decweb.ethz.ch/www6/technical/paper206/paper206.htm>
- 29 DiPasquo D. Using HTML formatting to aid in natural language processing on the World Wide Web. School of Computer Science, Carnegie-Mellon University, 1998
- 30 Luotonen. The common log file format. <http://www.w3.org/pub/www/>. 1995
- 31 Cooley R, Mobasher B, Suvastava J. Data preparation for mining World Wide Web browsing patterns
- 32 Elo-Dean S, Viveros M. Data mining the IBM official 1996 Olympics Web site: [Technical report] IBM T. J. Watson Research Center
- 33 Shahabi C, et al. Knowledge discovery from user Web-page navigation. In Workshop on Research Issues in Data Engineering Birmingham, England, 1997
- 34 Spiliopoulou M. Data mining for the web. In Principale of Data Mining and Knowledge Discovery, Second European Symposium, PKDD'99, 1999. 588~589

(上接第 39 页)

- 2 Aleksander I, Thomas W V, Bowden P A. WISARD—a radical step forward in image recognition, Sensor Review, July, 1984. 120~124
- 3 Yang Guoqing, Chen Songcan, Lu Jun. Multilayer Parallel Distributed Pattern Recognition System Model Using Sparse RAM Nets. IEE PROCEEDINGS-E, 1992, 139(2): 144~146
- 4 Kolcz A, Nigel M. Allinson. N-tuple Regression Network. Neural Networks, 1996, 9(5): 855~869
- 5 Tattersall G D, Foster S, Johnston R D. Single-layer lookup perceptrons. IEE PROCEEDINGS-F, 1991, 138(1): 46~54
- 6 Kanerva P. Sparse Distributed Memory. MIT Press, Cambridge, Massachusetts, 1988
- 7 Rohwer R, Morciniec M. The Theoretical and Experimental Status of the n-tuple Classifier. Neural Networks, 1998, 11(1): 1~14