

一种可编程 Hamming 神经网络特征提取器*

栗国星 石秉学

(清华大学微电子学研究所, 北京 100084)

摘要: 根据 Hamming 神经网络的工作原理提出了一种可用于手写体数字识别的电流型可编程局部结构特征提取电路。该特征提取器的模板不仅是可编程的, 它可根据不同的需要来随时更改模板的内容以适应不同的情况, 而且其特征合并的方式也是可配置的, 它可根据不同的需要把所需的特征合并成不同的特征类别。对于该可编程特征的电路模拟以及用单层多晶、双层金属 $1.2\mu\text{m}$ 数字 CMOS 工艺所制作的实验芯片的测试表明该电路能很好地完成特征提取的功能。

关键词: 可编程; Hamming 神经网络; 特征提取; CMOS

EEACC: 1265B; 1295; 6140C; 2570D

中图分类号: TN 711.4 文献标识码: A 文章编号: 0251-4177(2001)03-0378-05

1 引言

Hamming 神经网络虽然在所有人工神经网络模型中结构是最简单的, 但其权重连接的有效性却非常高, 并且不管模板数量和长度有多大, 它总能收敛于其存贮的其中的一个模板^[1]。因此它能完成模式分类或特征提取^[2]以及误码校正等多种功能。所以对它进行深入的研究有着重要的意义。

Hamming 网络的一般结构如图 1 所示, 它是一个两层的网络, 其中第一层是模板匹配子网络。如果 Hamming 网络具有 M 个输入端, J 个模板, 则该层网络完成对一个未知输入模式的匹配度的计算, 其表达式为:

$$MS_j = M - \sum_{i=1}^M S_i \oplus P_i^j \quad j = 1, 2, \dots, J \quad (1)$$

其中 $S_i \in \{0, 1\}$ 是输入模式的一个元素; P_i^j 是第 j 个模板中的第 i 个元素; MS_j 是输入模式对于第 j 个模板的匹配度。第二层求大网络则从 J 个匹配度中挑选出最大者并使其增强为 1'而其它则抑制为 0'。

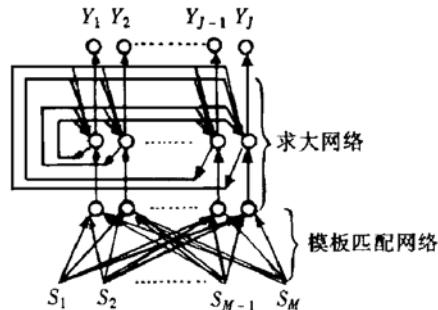


图 1 Hamming 神经网络的一般结构

FIG. 1 General Structure of Hamming Neural Network

Hamming 网络的 VLSI 实现有多种方式, 其中有电压型的^[3, 4]、电流型的^[5]、电荷型的^[6]以及开关电流型的^[7]等, 它们都是用模拟电路来实现的, 这是因为用模拟电路更容易实现网络中第二层的强非线性变换。以上几种实现 Hamming 网络的方法有一个共同的特点, 它们的权重都是固定不变的, 其规模也不能扩展。因此它们的灵活性和适应性等都相对较差。为了克服以上缺点, 须使其模板可以编程, 其规模可以扩展。常规的 Hamming 网络第二层子网

* 国家自然科学基金资助项目(69636030)。

栗国星 1970 年出生, 现为清华大学微电子所博士研究生, 研究方向为数/模混合电路和神经网络与模糊逻辑的 VLSI 实现。

石秉学 男, 1936 年出生, 1959 年毕业于清华大学。现为清华大学微电子所教授, 博士研究生导师, 从事人工神经网络和模糊逻辑系统及其 VLSI 实现和应用, 模拟集成和数/模混合集成电路及 BiCMOS 集成电路与系统的研究。

络只从 J 个匹配度挑选出一个最大者作为输出, 这对 Hamming 用作分类器时进一步提高性能不利。因为对于一个未知输入模式, 如果第一个匹配度最大者如果不是正确的结果, 则前几个最大匹配度中存在正确结果的几率却非常的大。如果 Hamming 网络能在第一级中选出这几个具有最大匹配度者, 再用一级较为简单的分类器利用一些局部特征从这几个类别中挑选出一个, 这样既提高了系统的识别率, 又使系统的复杂度增加不多从而增加分类器的性能代价比。兼于以上考虑, 本文提出了一种结构新颖的电流型 Hamming 神经网络, 它可以较好地完成对手写体字符点阵的特征提取。因为良好的特征是高性能识别的基础, 利用该特征提取器进行高质量的特征提取, 这样即精简了信号量, 又大大减轻了对后级识别系统的负担。本文的第二部分给出了可编程模板匹配电路的设计, 第三部分则给出了相应的模拟和测试结果, 最后则对本文加以总结。

2 特征提取器的总体结构

该 Hamming 神经网络特征提取器主要由可编程模板电路、可配置 WTA 电路、可配置特征合并电路以及译码及控制电路组成。该特征提取器的结构框图如图 2 所示。数据输入窗口是一个 5×5 的二值输入窗口, 通过该窗口在控制逻辑的作用下, 可对一个归一化为 20×20 的字符点阵分成 9 个区域进行扫描。可编程模板电路中最多可实现 24 个长度为 25 的可编程模板, 除此之外为了补偿不同模板之间因复杂度不同而对某些模板在匹配过程中造成某种歧视, 因此每个模板还引入两位编程码以构成 4 种局部阈值。每个模板电路中可存入事先学习得到的模板, 这样最多可形成 24 位具有不同局部阈值的模板。而且这种模板可根据需要随时进行更新。对于手写体数字识别, 我们只选择了具有代表性的 20 个模板, 它们分别用来从归一化和细化后的字符点阵中提取端点、十字叉、拐点等局部结构特征。对于剩余的模板电路可用作备用或构成牺牲模板等。对于较简单的特征只需一个模板表示既可, 而对于较复杂的特征需要用几个模板来表示, 这样 20 个模板共代表 8 类特征。当一个窗口的数据加载到模板电路模块后, 便得到 24 个匹配度电流, 这些电流被送往 WTA 电路中进行判断以决定输入的窗口数据最接近哪一个或哪两个模板, 从而决定输入的窗口数据

属于哪个特征。由 WTA 电路得到 24 位二值结果后, 到底哪些特征输出属于同一类别的不同变形则由可配置特征合并电路决定。特征合并形式的控制码是在模板编程完毕后的第一时钟周期从数据窗口存入到可编程模块的锁存电路中的。经过可配置特征合并电路后得到 8 位特征输出, 每位代表一个特征类别。在控制电路的作用下, 每扫描完一个区域便可得到对应该区域的最终的 8 位特征矢量, 这样对整个字符扫描完毕后便可得到长度为 72 位的一个特征矢量。该特征矢量送入模糊逻辑识别器或神经网络识别器中便可进行识别。

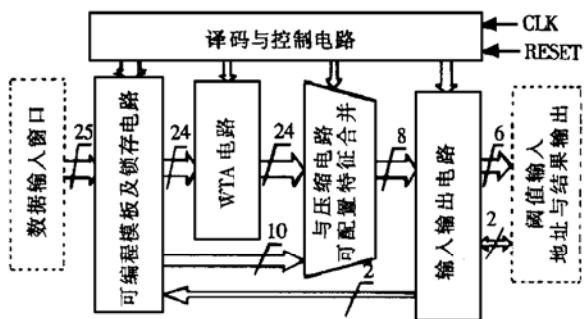


图 2 可编程特征提取器的总体结构框图

FIG. 2 General Structural Block Diagram for Programmable Features Extractor

3 可编程模板匹配电路和可配置特征合并电路的设计

模板匹配电路是对一个长度为 M 的输入矢量 S 与所存贮的 J 个长度也为 M 的标准模式矢量 P^j (模板) 进行相似性计算。这种计算是以全并行的方式进行的, 因此当未知模式输入后便可同时产生 J 个匹配度, 匹配度越大表明其相似性越强。本文所提出的可编程模板电路如图 3 所示, 模板的每一位由一个锁存器来存贮, SEL 是一个模板选择信号, 它来自译码器的输出, 译码地址由计数器产生。在 SEL 为高电平时, 来自模式输入端的数据便被锁存在 M 个锁存器中, 此后 SEL 便一直为低电平。这样 J 个模板通过 J 个时钟周期后便被编程完毕, 它是在系统起动后前 J 个时钟周期内完成的, 之后系统便可一直接受未知输入模式。图中 V_B 是一个偏置电压, 它使 NMOS 电流镜中每一路的导通电流为 I_{REF} 。这样当模板权重编程完毕后便对输入模式进行匹配运算, 其中第 j 个模板所计算出的匹配电流

为

$$I_{MSj} = M - \sum_{i=1}^M S_i \oplus P_i^j I_{REF} \quad j = 1, 2, \dots, J \quad (2)$$

该电流在 LD 为高电平时输出。图 4 是在 $I_{REF} = 10\mu A$, $M = 10$ 的情况下匹配电流随匹配度增加时的变化曲线。它是 HSPICE 对匹配度从 1 变化到 10 时的输出匹配电流。从图中可以看出这种电路可以较好地实现匹配度计算的功能。图中 NMOS 电流镜的管子大小取为 $10\mu m / 5\mu m$, 在匹配度的动态范围之内, 匹配误差都在 1% 之内。如果输入模式的矢量长度较长时, I_{REF} 的取值可适当减小以降低整体功耗, 同时也避免匹配电流超出后级 WTA 的动态范围。由于 I_{REF} 的减小, 最大匹配电流与次最大匹配电流之间的差也相应减小, 因此应能保证后级的 WTA 的分辨精度大于 I_{REF} 。

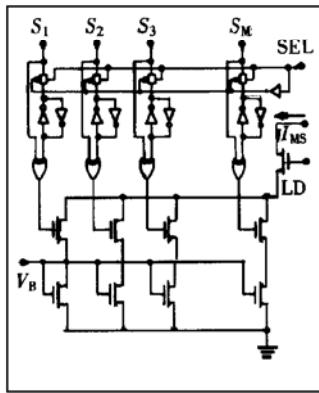


图 3 可编程模板匹配电路

FIG. 3 Matching Circuit Diagram for Programmable Template

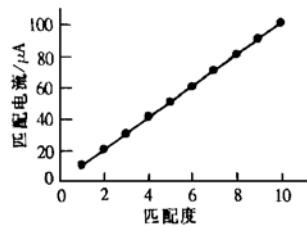


图 4 匹配电流与匹配度的关系

FIG. 4 Matching Current Versus Matching Level

得到各个模板所对应的匹配度电流后, 便可由 WTA 网络选出输入模式所属的特征类别, 我们所选取的 WTA 电路是一个即可配置为 1-WTA, 又可配置为 2-WTA 的电流型电路, 具体电路可参考文

献[8]。

由 WTA 网络得到 24 位特征矢量, 该特征矢量需要进行合并和压缩, 合并的目的是把代表相对较复杂的特征类别不同的特征模板输出合并为一个特征类别; 特征压缩的目的是检测某一类特征在某一区域是否存在。可配置特征合并和压缩的电路图如图 5 所示, 图中 $D_j (j = 1, 2, \dots, 24)$ 来自 WTA, 其中的每一位代表一种特征。 $C_i (i = 1, 2, \dots, 10)$ 是特征合并配置码, 如果 C_i 是高电平, 则第 j 种特征合并成第 i 类特征; 如果 C_i 为低电平, 则第 j 种特征合并为第 $i+1$ 类特征。特征合并配置码 $C_i (i = 1, 2, \dots, 10)$ 来自可编程模块中, 它是在系统启动时模板编程完毕后, 第一个时钟周期通过数据输入窗口送入到内部锁存器的, 之后它便一直保持不变。特征合并后, 通过 D 触发器和或非门, 在时钟信号 CLK 和复位信号 CLR 的控制下对一个区域内的特征进行压缩。

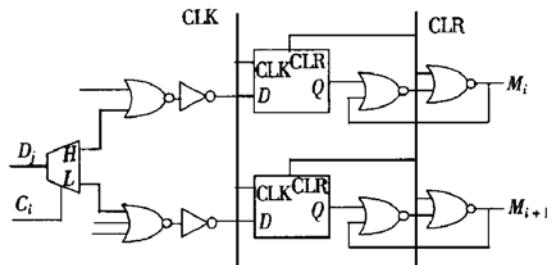


图 5 可配置特征合并与压缩电路

FIG. 5 Reconfigurable Feature-Merging and Extracting Circuits

4 系统电路的模拟与测试

我们在 CADENCE 环境下利用 HSPICE 对该特征提取器的单元电路以及全电路进行了最好、最差以及典型的工艺参数情况下的模拟, 从模拟结果可以看出, 该电路能够较好地完成特征提取的功能。

图 6 是该特征提取器在一种输入情况下的模拟, 图中 CLR 为复位信号, 它为高时对触发器进行置零操作, CP 信号为结果输出时钟, 在编程阶段, CP 信号用来从数据输入窗口加载编程数据。在特征提取阶段, 每经过 64 个时钟周期, 它输出一个高电平, 该高电平用于把一个区域内的特征数据送出。 M_0-M_7 是特征提取结果, 图中只给出前 4 个。从图可以看出, 在编程阶段, M_0-M_3 都有输出, 此时它们输出的是对外部 EEPROM 存贮器进行寻址的地

址。编程完毕后只有 M_0 在特征提取阶段始终输出为高电平, 而其它输出端为低电平。这与预期的完全一致。对于其它情况下的模拟也说明其功能的正确性。

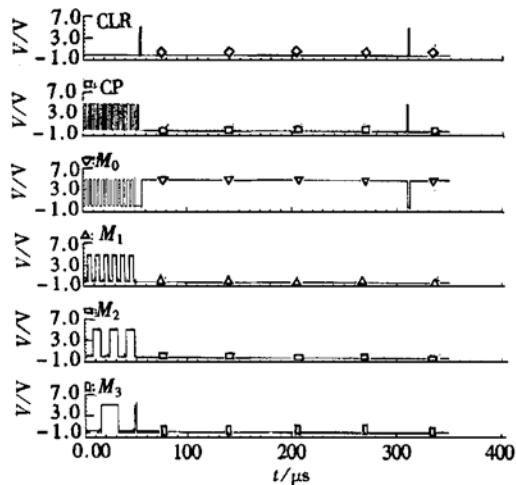


图 6 Hamming 神经网络特征提取器全电路的模拟

FIG. 6 All-Circuit Simulation for Hamming Neural Network

我们用单层多晶、双层金属的 $1.2\mu\text{m}$ 数字 CMOS 工艺对该可编程 Hamming 神经网络特征提取器进行了流水试验。图 7 是该特征提取器的芯片显微照片。从图中也可以看出, 24 个长度为 25 位的可编程模板电路占据了芯片的绝大部分面积。它的芯片总面积为 $4.3\text{mm} \times 4.6\text{mm}$, 核心电路面积为

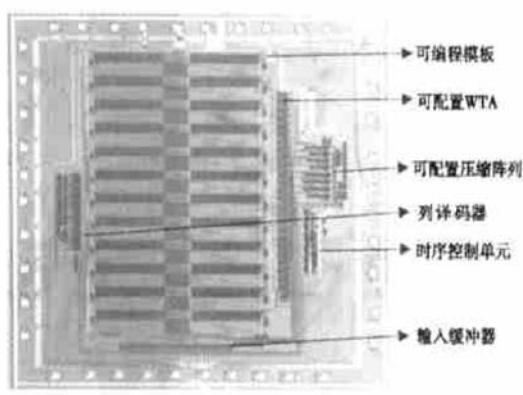


图 7 可编程可配置 Hamming 网络特征提取器显微照片

FIG. 7 Microscope Photograph of Features-Extractor for Programmable and Reconfigurable Hamming Neural Network

$3.4\text{mm} \times 3.7\text{mm}$ 。我们以 40 脚双列直插(DIP40)陶瓷封装, 可编程模板中电流镜的参考电流通过调

节外部的电阻使其恒定为 $3\mu\text{A}$, 因为这样可使输入到 WTA 的最大匹配电流刚好为 WTA 分辨度最高的范围($50\text{--}100\mu\text{A}$)。输入到该芯片的时钟和复位信号以及数据信号都是通过并口得到的。我们可以利用示波器分别测试芯片的输出, 以检验其功能的正确性。该特征提取器的输出主要有两部分, 一部分是特征提取结果输出, 它们分别为 $M_0\text{--}M_7$ 共 8 个信号, 另一部分是控制和状态信号输出, 它们分别是 CP 和 EA。在 EA 为低电平时表明提取器正处于模板编程阶段, 此时 CP 的输出为编程周期, 而 $M_0\text{--}M_4$ 的输出为正在被编程的模板地址, 它们可以用来对外部存贮器进行寻址, 从而把其内容加载到对应的模板中。在 EA 为高电平时表明对模板编程已经结束, 提取器正处于特征提取状态, 此时 $M_0\text{--}M_7$ 便为提取结果, 在经过 64 个时钟周期后 CP 为高电平表明对一个区域扫描完毕。之后便对输出结果进行清除以进行下一个区域的扫描。图 8 是对输出信号 EA 和 M_2 及 M_4 的测试结果。由于模板编程完毕之后该 Hamming 网络特征提取器的输入是固定的, 所以其输出也与输入对应而固定不变, 这里我们选

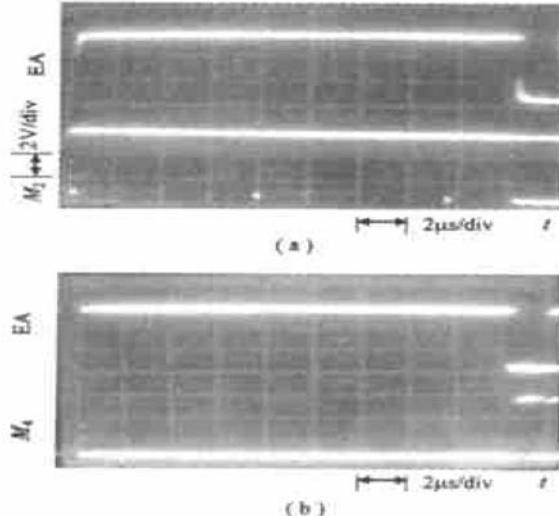


图 8 Hamming 网络特征提取器测试结果

FIG. 8 Measuring Result of Feature-Extractor for Hamming Network

择了与 M_2 对应的模板作为固定输入。从图 8(a)可以看出, 在 EA 为低电平时 M_2 输出为几个方波, 它是正在被编程的模板所对应的一个地址。在 EA 为高时, 它输出也一直为高电平, 只是每经过 64 个时钟周期它被置为低电平一次。而图 8(b)则为 M_4 的输出结果, 在 EA 为低时它输出的是模板的一个地

址,而在 EA 为高时它输出一直为低电平,表明输入数据中没有这类特征.对于其它几个特征输出端的测试也是同样的结果,即除 M_2 输出为高电平外,其它输出端在 EA 为高时都输出低电平.

由于该特征提取器中构成模板的电流镜中的电流设定为 $3\mu A$,所以一个模板全部匹配时其匹配电流也只有 $75\mu A$,而该电流只有在 LD 信号为高电平时才构成通路,第二级的 WTA 又是一个自偏置电路,只有当有电流输入时它才工作,而后级的特征压缩电路都是数字电路,因此,尽管该特征提取器近一万两千支管子,而其功耗只有 $16.0mW$. 编程完毕之后它可以在 $2MHz$ 以上工作,其最低工作电压为在 $3.3V$.

5 结论

本文提出的模板可编程、特征合并可配置 Hamming 神经网络特征提取器,具有结构简单灵活,精度较高、功耗较低的特点. 它可有效地以流水的方式按区域对手写体数字字符点阵进行局部结构特征的提取,每扫描完一个区域便输出其结果并进

行下一区域的扫描. 对于该提取器的电路模拟以及实验芯片的测试结果也证明了以上优点.

参考文献

- [1] R. P. Lippmann, IEEE ASSP Mag., 1987, **4**: 4—22.
- [2] Lu Wei, Shi Bingxue and Li Zhijian, Acta Electronica Sinica, 1997, **25**(2) : 25—28 (in Chinese) [路伟, 石秉学, 李志坚, 电子学报, 1997, **25**(2) : 25—28].
- [3] Li Binqiao, Li Zhijian and Shi Bingxue, Chinese Journal of Semiconductors, 1996, **17**(3) : 217—222 (in Chinese) [李斌桥, 李志坚, 石秉学, 半导体学报, 1996, **17**(3) : 217—222].
- [4] M. Robinson, H. Yoneda and E. S-Sinencio, IEEE Trans. Neural Networks, 1992, **3**(3) : 444—456.
- [5] Lu Wei, Shi Bingxue and Li Zhijian, Acta Electronica Sinica, 1997, **25**(5) : 29—34 (in Chinese) [路伟, 石秉学, 李志坚, 电子学报, 1997, **25**(5) : 29—34].
- [6] U. Cilingiroglu, IEEE J. Solid-State Circuits, 1993, **28**(1) : 59—67.
- [7] D. Grant, J. Taylor and P. Houslander, IEEE J. Solid-State Circuits, 1994, **29**(9) : 1154—1157.
- [8] Li Guoxing, Shi Bingxue and Lu Wei, Chinese Journal of Semiconductors, 1998, **19**(7) : 532—537 (in Chinese) [栗国星, 石秉学, 路伟, 半导体学报, 1998, **19**(7) : 532—537].

A Programmable Hamming Neural Network Feature Extractor*

LI Guo-xing and SHI Bing-xue

(Institute of Microelectronics, Tsinghua University, Beijing 100084, China)

Abstract: A current-mode programmable local feature extractor for handwritten digit recognition is put forward based on the principle of Hamming neural network. The templates of the extractor is programmable and their contents can be refreshed according to different requirement. The way of feature merging is also reconfigurable so as to merge different features into different feature classes. A prototype chip of the feature extractor has been implemented in single poly, double metals $1.2\mu m$ digital CMOS technology and both the simulated and measured results prove that the extractor can realize the feature extraction well.

Key words: programmable; Hamming neural network; feature extracting; CMOS

EEACC: 1265B; 1295; 6140C; 2570D

Article: 0253-4177(2001)03-0378-05

* Project Supported by National Natural Science Foundation of China Under Grant No. 69636030.

LI Guo-xing was born in 1970. He received the Ph. D degree from Tsinghua University in 1999. His research interests include artificial neural networks, fuzzy logic and their VLSI implementation, analog and digital/analog mixed integrated circuits.

SHI Bing-xue was born in 1936. He is now a professor and an advisor of Ph. D students. His research interests include analog, mixed digital/analog ICs, BiCMOS DC-DC convertors, artificial neural networks, fuzzy logic and their VLSI implementation.