

一种用于复杂模式分类的电流型 Hamming 神经网络*

栗国星 石秉学 路 伟

(清华大学微电子学研究所 北京 100084)

摘要 本文结合手写体数字识别, 讨论了一种用 Hamming 神经网络实现复杂模式分类的方法。文中提出了模式分解, 分级模板匹配及模糊匹配技术及其电流型 VLSI 的实现, 并着重论述了一种高精度, 低复杂度, 结构灵活的电流型 k-WTA 的实现。

EEACC: 1280, 1285, 1295

1 引言

八十年代末神经网络热潮的兴起, 为其在各个领域的应用展开了新的局面^[1], 尤其是在模式信息处理方面更是令人注目, 但到目前为止, 神经网络的实现方式仍以软件实现为主, 用软件不能体现出神经网络本身固有的大规模并行性, 因此人们便开始探索用 VLSI 来实现神经网络, 在众多的神经网络模型中, Hamming 网络最易于 VLSI 的实现, 不少研究人员已提出或制作出了 Hamming 神经网络芯片, 如李斌桥提出并成功地制作了一种高速高精度的电压型 Hamming 神经网络芯片^[2,3], U. Cilingiroglu 提出了一种基于电荷型的 Hamming 网络实现方法^[4], 但时序复杂, 速度较慢。它们都难以实现对复杂模式的处理, 为此本文提出了一种用 Hamming 网络对复杂模式进行分类的方法及其电流型 VLSI 的实现。

2 基本原理

Hamming 网络较适合于二值图象或数据流的处理, 它可以从其记忆的 K 个标准模板中选择出与输入模式较接近的一个或多个模板, 较简单的 Hamming 神经网络可以只有两级, 其中第一级用来进行模板匹配, 第二级则用来求出与输入模式较匹配的模板, 也即求大网络。对于二值模板, 模板匹配的定义如下:

* 国家自然科学基金与高校博士点基金资助课题

栗国星 男, 1970 年出生, 博士生, 研究方向为字符识别、人工神经网络与模糊逻辑应用、模拟集成电路设计

石秉学 男, 1936 年出生, 教授, 博士生导师, 从事人工神经网络及模糊逻辑系统及其集成电路实现、模拟集成和数/模混合集成电路与系统研究

1997-05-05 收到, 1997-08-28 定稿

设输入模式 X 是一个 N 维的矢量, 共有 K 个标准模板 T_K , 则匹配度 $M d_i$ 定义为

$$M d_i = N - \sum_{j=1}^N X_j \oplus T_{ij} \quad i = 1, \dots, K$$

其中 T_{ij} 是第 i 个模板中的第 j 位值; $T_{ij} \in \{0, 1\}$, 有时只使用二值模板还不能满足要求, 这时可以使用实值模板, 即模板的值在 0 和 1 之间, 为了电路实现的方便, 可以使模板的值量化为 0 到 1 之间的几个离散值, 比如 0.2, 0.4, ..., 0.8, 1.0 等, 但输入仍是二值数值, 这时匹配度 $M d_i$ 可定义为

$$M d_i = \sum_{j=1}^N (X_j T_{ij} + \overline{X_j} (1 - T_{ij})) \quad i = 1, \dots, K$$

这里 $T_{ij} \in (0, 1)$, 它的选取是在片外学习得到的。这种匹配可以作为前一种匹配的推广, 对于某些问题, 它可能更具有实际意义。

前面两种匹配度的计算, 只考虑了输入模式的一维性质, 当输入模式是二维信息时, 为了获得更多的匹配信息, 可以采用加权匹配, 也可以加入相关项以增强模式的表达。

输入模式与 K 个标准模板的匹配度计算出之后, 由 WTA (Winner-Take-All) 选取出匹配度较大的一个或多个作为网络的最终输出。

这种结构的 Hamming 神经网络, 处理较复杂模式的能力不强, 比如要识别一个 20×20 点阵的手写体数字, 即使是用软件实现, 由于手写体字符的变化比较大, 一个类别很难用一个模板来表达, 尽管一个类别可以用较多的模板来容忍字符的变形等, 这势必会使网络的规模变得十分的庞大, 用 VLSI 来实现更是没有可能, 因为它还会受到 I/O 口数量的限制, 为了解决这一问题, 可以采用模式分解, 分级匹配的办法, 在我们的 VLSI Hamming 网络识别器中, 正是采用了这种思想。

3 电流型 Hamming 神经网络识别器

我们设计的 Hamming 神经网络识别器结构框图如图 1 所示, 待识别模式是一个 20×20 的字符点阵, 该字符点阵先被分割为 3×3 个区域, 每个区域有 8×8 个格点, 每个区域与其相邻的区域有两层格点相重合以避免一个完整特征被分割开, 通过一个 5×5 的局部窗口逐点对待识别模式的每个区域进行扫描, 每一个局部窗口内的子模式与 23 个特征模板同时进行匹配操作, 所选取的特征包括端点, 十字叉, 拐角及其变形等, 由 23 端 WTA 选出匹配较好并大于给定阈值的作为对应类别特征的输出, 由于不同的模板具有不同的复杂度, 所以不同的模板应具有不同的阈值从而构成多阈值的模板匹配过程。由于模式的复杂性, 分解后得到的子模式也很难断定它属于哪一类特征, 通过实验证明, 匹配度最好的前两个输出结果正确的概率极大, 因此有必要使 WTA 网络选取出两个结果, 为此我们设计了一个 k -WTA 网络。随着扫描的不断进行, 可以把得到的特征按事先规定好的特征类别进行合并和压缩从而形成流水

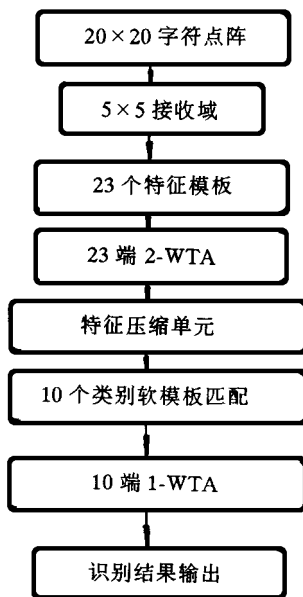


图 1 Hamming 网络识别器系统框图

作业, 压缩可由“或”门和触发器来实现。每个区域扫描结束后得到 8 类特征, 每一类由一位二进制数表示, 如其为“1”则表示在该区域内有这类特征, 为“0”则表示没有这类特征, 待 9 个区域全部扫描结束后, 便得到 8 个 3×3 的特征图, 这样共形成 72 位特征矢量, 由此可进行第二级的模板匹配过程, 为了减小电路的规模, 在尽可能多地保留有用信息的情况下, 可以利用自相关方法(另待发表)并借助于遗传算法把特征矢量长度进一步压缩为 48, 当然这种压缩是在片外完成后固定在片内的, 这时每个模式类别可用一个长度为 48 的模板表示, 对于阿拉伯数字, 共有十个类别, 因此共有十个模板。

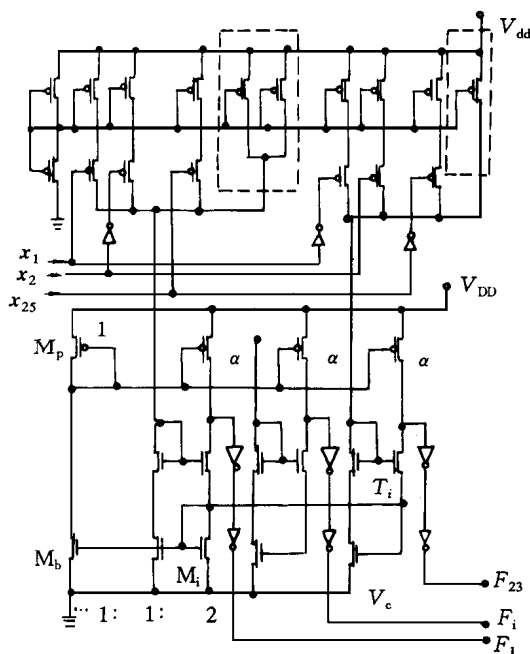


图 2 二值模板匹配和 2-WTA 电路

一位控制着一个开关管, 如果模板对应位为“1”, 则在输入端和相应的栅极之间加一个反相器, 否则直接连接到开关管的栅上。每个模板对应的阈值用虚线框内电流镜来实现, 它们没有开关管的控制而直接对整个匹配电流产生贡献, 其值的大小由其对应的模板所决定。图 3 是每一路电流为 $10\mu\text{A}$ 的情况下, 对于 25 个输入端匹配电流与匹配度的关系曲线, 由此可以看出, 匹配度和匹配电流成较好的线性关系。由图中也可以看出, 随着匹配度的增高, 匹配电流偏离理想值越多, 这是由于电流镜中 PMOS 管的沟道长度调制效应引起的, 因为在 WTA 输入阻抗一定的情况下, 随着匹配电流的增加, 构成电流镜的 PMOS 管的源漏之间的电压有所减小, 由于沟长调制效应而使电流镜中的电流有所减小, 但减小的电流在匹配电流的比重是很小的, 这个偏差可以通过进一步减小 WTA 的

为了提高识别率, 此时的模板不宜再用二值表示, 这时可采用实值模板, 这种模板也可叫软模板或模糊模板, 这类模板也可以方便地用电流镜来实现。经过该次匹配运算之后, 即可得到最终识别结果。

4 单元电路的设计

4.1 匹配电路与 2-WTA 网络电路

匹配电路有两种形式, 图 2 上半部分用来完成模板匹配运算, 下半部分是一个电流型 WTA 网络, 图中 $x_1, x_2, \dots, x_{25}$ 是 5×5 子模式的输入矢量, $F_1, F_2, \dots, F_{23}$ 是每个模板所对应的输出, $x_i, F_j \in \{0, 1\}$ ($25 \geq i \geq 1, 23 \geq j \geq 1$)。对于二值模板, 匹配电路中的电流镜的宽长都是一样的, 电流镜中的电流的大小可由左端的偏置电路加以调节, 在我们的设计中, 电流镜中的电流为 $10\mu\text{A}$ 。外部输入子模式的每

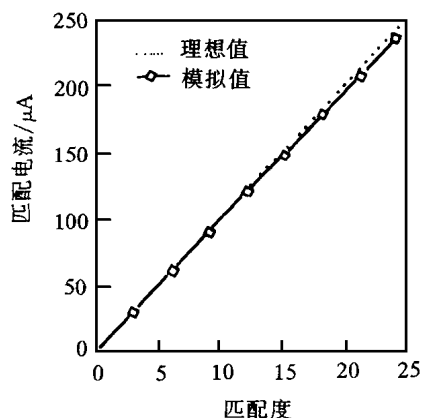


图 3 匹配度与匹配电流的关系

输入阻抗和(或)适当增大电流镜的沟道长度来减小,这需要在性能和芯片面积之间折中考虑

图 2 的下半部分是一个 $2 \times WTA$ 网络,其输入是模板匹配电路产生的匹配度电流,每一模板对应一个 WTA 输入,其输出对应两个最匹配的模板.该电路主要由多输入求大(MAX)网络^[5]和电流阈值单元构成,但构成 NMOS 电流镜的管子 M_i 和 M_b 以及提供阈值参考电流的 PMOS 电流镜的管子的尺寸的选择对整个 WTA 的功能产生重要的影响,(为一个小于 1 的比例因子,图中是一个 $2 \times WTA$ 的情况,.该电路不仅结构简单,而且精度较高,动态范围较大,可以工作在弱反型直至强反型区,改进的 Wilson 电流镜的使用,使 WTA 中的单元电路具有这种电流镜所具备的优点,其功耗也比较低,对于十一端的 $2 \times WTA$,在平均输入电流为 $50 \mu A$ 以上时,其功耗也只有 $0.5 mW$.图 4 是具有 4 个输入端的这种电流型 WTA 的在大电流和小电流情况下的精度模拟,图中 $V(out1), V(out2), V(out3), V(out4)$ 是输入端对应的输出, i_1, i_2, i_3, i_4 是输入电流,模拟是在固定 i_1, i_2, i_4 , 使 i_3 变化进行的.图 4(b) 的输出则分别是 $V(n11), V(n12), V(n13), V(n14)$. 在小电流情况下,其精度为 $1 \mu A$ 之内,在大电流情况下,其精度仍高于 $5 \mu A$,事实上其精度与所操作的电流的平方根的倒数成正比.其收敛的速度不仅与其操作的电流大小有关,而且还与各输入端电流的差值有关,操作的电流越小,收敛速度越慢,电流差值越小,收敛速度也越慢

4.2 特征压缩电路

特征压缩根据不同的要求可采取不同的压缩方法.由于模式的输入是以分时方式进行的,在每一时钟周期内对所输入的子模式与 K 个标准模板进行匹配运算,求出匹配最好的两个模板作为该子模式的所对应的子特征类别,对较相似的特征类别进行或操作合并后得到 M ($M < K$) 个大的特征类别,在我们的识别系统中,是把 23 个特征合并为 8 类特征,这样每一个区域扫描结束后便得到得到 8 个规模为 8×8 的特征向量图,整个输入模式扫描完毕后便得到 9 组这样的特征图,显然这些特征还不能用来进行第二级的匹配,因此还必须进行更进一步的压缩,而这次压缩也是比较有意义的,通过这次压缩不仅可以大大减小电路所需的规模,而且也可以消除模式的局部歪曲变形等.事实上这种压缩是按区域随着特征的生成在控制电路的作用下流水式进行的,每一区域扫描完毕后,得到 8 位特征矢量,每一位代表一个特征类别.图 5 是在我们的设计中所采用的压缩单元电路^[6],其逻辑功能为当清零信号 $CLR = 1$ 时,压缩单元输出 $M_k = 0$,当 $CLR = 1$ 时,有 $M_k^j = F_k^j + M_k^{j-1}, j = 1, \dots, 64 \quad k = 1, \dots, 8$ 其中 F_k 是在某一区域内对一个 5×5 子模式扫描合并后得到的第 k 位特征, φ 是时钟信号,由于有 8 类特征,所以

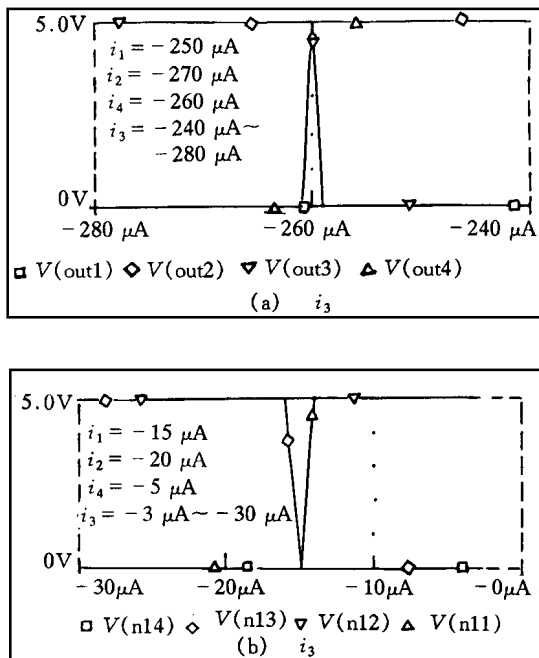


图 4 4 输入电流型 WTA 两种情况下的精度模拟

要有 8 个这样的单元并行工作, 这样待整个模式扫描完毕后便得到 8 个 3×3 的特征图 为了减小电路规模, 这个长度为 72 的特征矢量可用一定的方法作进一步的压缩, 从而最终得到长度为 48 的特征矢量, 该特征矢量可以用来进行第二级的模板匹配

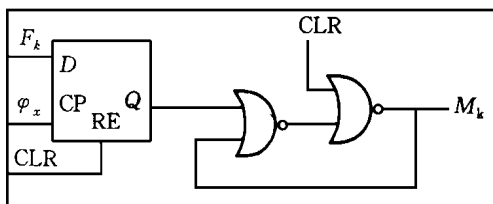


图 5 一种特征压缩单元电路

与此类似 其实现方法是先选定一个参考电流镜来实现模板中的“1”, 其它数值的元素便由电流镜 PMOS 管的宽度之比确定 (长度相同), 在具体实现时, 对于那些为“0”或“1”的模板元素, 可以省去单元中左边或右边中的相应支路中的管子. 特征码为 M_i ($48 \geq i \geq 1$), 它是二值的 该电路中的 WTA 是可配置的, 根据控制码 CTL 的不同, 它可实现 1-WTA 或 2-WTA, 这在实际应用中是很有意义的, 如果一级识别达到要求, 则可控制它使其完成 1-WTA 的功能, 否则可使其成为 2-WTA 以便实现系统之间的级联, 由于匹配最好的前两个结果正确的概率极大, 这样后级便可由实现两两类别对之间分类的分类器来实现, 因为两个类别之间的分类总是要容易一些, 它可只使用一些局部特征便可达到分类的目的 为了实现高的可靠率, 在该 WTA 网络中加入了阈值电流 I_{th} 输入端, 调节该阈值电流 I_{th} 便可实现识别率和可靠率之间的权衡 R_i ($i = 0, 1, \dots, 9, J$) 为识别结果, 其中 J 为拒识信号, 当 J 为高电平时便表示拒识

4.3 软模板匹配电路和 1-WTA 电路

图 6 是软模板匹配电路和一可配置的 WTA, 图中只画出了一个模板, 其它模板结构

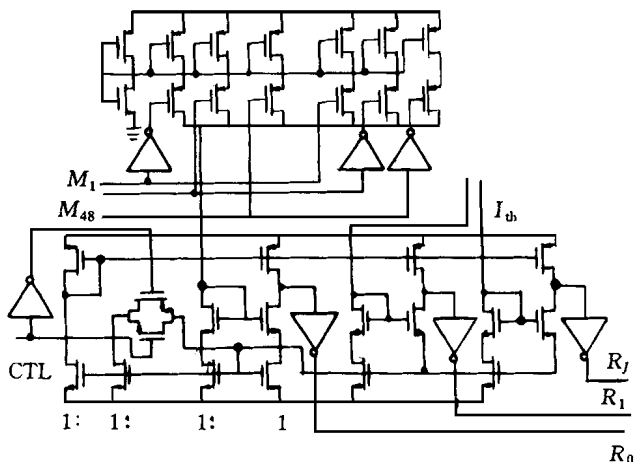


图 6 第二级模板匹配电路及可配置 WTA

5 结论和进一步的工作

本文基于手写体数字的识别问题, 提出了一种用于复杂模式分类的 Hamming 神经网络结构及其电流型 VLSI 的实现, 各个模块电路不仅结构简单, 而且还具有灵活性高, 高精度低功耗的特点, 目前部分模块的电路已经画完版图, 预计能于十月份进行工艺流水

由于 E^2 PROM 工艺的成熟, 为了使制作的芯片更具有灵活性, 模板电路可以用 E^2 PROM 来控制, 这样所选用的模板以后如果不能满足要求时可以在片上进行调整, 压缩电路通过引入一些由 E^2 PROM 实现的控制单元以使压缩电路具有一定的可配置性, 这样就可以达到实现不同的合并和压缩方案以适应不同的应用要求

参 考 文 献

- [1] A. K. Jain *et al* , Computer, 1996, Mar , 31~ 44
- [2] Binqiao Li, Zhijian Li and Bingxue Shi, IJCNN '93, Nagoya, Japan, 879~ 882
- [3] 李斌桥,李志坚,石秉学, 半导体学报, 1996, 17(3): 217~ 222
- [4] Ugur Cilingiroglu, IEEE J. Solid-State Circuits, 1993, 28(1): 59~ 67.
- [5] C.-Y. Huang and B.-D. Liu, Electronics Letters, 1994, 30(23): 1924~ 1925
- [6] 路伟, 清华大学博士论文, 1997, 5

A Kind of Current Mode Hamming Neural Network to Classify Complex Pattern

Li Guoxing, Shi Bingxue, Lu Wei

(Institute of Microelectronics, Tsinghua University Beijing, 100084)

Received 5 May 1997, revised manuscript received 28 August 1997

Abstract A method to classify complex pattern with Hamming neural network is proposed on the base of handwritten digit recognition in this paper. Some techniques such as pattern decomposition, multi-level template matching and fuzzy template matching and their current mode VLSI realization are put forward, and a current mode k-WTA with high precision, low complexity, smart structure is deeply discussed.

EEACC: 1280, 1285, 1295