论 文

基于离散随机技术的 Madaline 敏感性近似计算研究

钟水明*, 曾晓勤, 刘惠义, 徐彦

河海大学计算机与信息学院模式识别与智能系统研究所, 南京 210098

* 通信作者. E-mail: zhongyi@hhu.edu.cn

收稿日期: 2009-12-27; 接受日期: 2010-04-30

国家自然科学基金 (批准号: 60571048 和 60971088)、江苏省普通高校研究生科研创新计划 (批准号: CX09B_157Z) 和河海大学优秀博士学位论文培育计划 (批准号: 2010B19214) 资助项目

摘要 本文系统地讨论了离散型前向神经网络 Madaline 的输出对网络参数扰动的敏感性计算. 首先, 根据神经元 Adaline 的离散特性, 提出了一种离散随机技术方法, 推导了 Adaline 敏感性近似计算公式. 该方法从理论上解决了已有的连续随机技术方法无法处理的情况, 取消了过强的限制条件, 并使得从理论上对 Adaline 敏感性近似计算进行误差分析成为可能. 其次, 在 Adaline 敏感性基础上, 针对 Madaline 的结构特性以及处理层间信息传递面临的高复杂度组合计算问题, 提出了一种新的基于贡献度大小的取舍策略来计算 Madaline 敏感性. 该策略在计算精度上优于现有的简单平均方法, 并降低了计算复杂度. 本文给出的敏感性近似计算公式和算法具有形式简单、计算复杂度低、近似误差小、一般性强等优点, 大量的仿真模拟实验验证了公式和算法的准确性和有效性.

关键词 Adaline Madaline Feedforward Networks 敏感性 参数扰动 离散随机技术

1 引言

理论上, 前向神经网络是在一定的结构框架下, 根据给定训练样本数据, 通过自身的学习机制来调整权值, 从而在网络的输入和输出之间建立一种映射关系. 然而, 这个理论上的映射在面对现实世界中不可避免的干扰会表现出怎样的容错能力? 对非训练样本数据又会表现出怎样的泛化能力? 神经网络敏感性的研究就是要试图找出这些问题的答案, 同时也为探索解决神经网络的一些其他基本问题提供理论和技术上的支持.

从文献检索可知, 神经网络敏感性研究在国际上得到了越来越多的重视, 已有不少围绕该课题的研究成果见诸报道^[1-21]. 研究人员从不同的角度、针对不同的网络模型、采用不同的方法来进行敏感性研究^[1-14], 并应用敏感性研究的结果解决神经网络研究领域里的一些热点问题^[15-21], 例如, 网络容错性能度量^[16,17]、网络泛化性能评估^[17,20]、网络属性筛选^[15,18]、网络结构裁剪^[18,19,21]等.

Madaline 是一种离散的前向神经网络, 由神经元 Adaline 组成, 采用的是有监督的学习机制. 虽然离散问题可视为连续问题的特例, 但是对那些本质为离散且连续技术不直接适用的应用领域, 如逻辑运算、信号处理、分类与聚类、模式识别等, Madaline 比连续神经网络有明显优势, 主要表现为: (1) 更

引用格式: 钟水明, 曾晓勤, 刘惠义, 等. 基于离散随机技术的 Madaline 敏感性近似计算研究. 中国科学: 信息科学, 2011, 41: 157-172

易于对问题的描述和理解, 可避免连续神经网络处理离散问题时因其输出需额外离散化所带来的误差; (2) 网络设计和计算更简单; (3) 更便于网络硬件实现 (能利用现有的 VLSI 技术) 等. 由于 Madaline 的离散特性, 使得连续技术对它不适用, 必须另辟蹊径. 本文将着眼于 Madaline 敏感性的定性分析和定量计算, 打造一个通用的敏感性量化尺度, 为深入研究 Madaline 行为机理和解决相关问题奠定基础.

早在 1962 年, 基于 n 维 Adaline 的所有输入模式点均匀分布于 n 维超球面上的事实, Widrow 的学生 Hoff^[1] 首次提出了 Adaline 输入空间的超球面几何模型, 并得出了 Adaline 任一输入模式点位于超球面某一特定区域的概率等于该区域面积占整个超球面面积的比值. 该模型为后续 Adaline 敏感性近似计算的研究从理论上开启了一个探索方向. 接着在 1965 年, 为了能充分利用超球面的对称几何特性, 以及考虑到神经元的偏置可作为一维特殊的权来对待, Glanz^[2] 将 Hoff 的方法进一步扩展到 $n+1$ 维空间, 提出了 Adaline 输入空间的半超球面模型. 随后于 1989–1990 年, 在 Hoff 和 Glanz 的研究基础上, Winter 和 Stevenson 等^[3,4] 采用以扰动角 θ 为随机变量的连续随机技术, 对 Adaline 敏感性进行近似计算, 并推导得出了在输入维数很大、Adaline 参数扰动前后大小的比值足够小 ($|\Delta W| \ll |W|$ 或 $|\Delta X| \ll |X|$) 前提条件下的 Adaline 敏感性近似计算表达式. 1995 年, Widrow 的学生 Piché^[6] 走了一条不同于以往的研究途径, 他以具有相同结构的整个一类 Madaline 为研究对象, 建立了一个随机模型来计算 Madaline 敏感性, 但是, Piche 的研究结论不适用于具体的个体网络.

除了上述 Widrow 团队的研究工作以外, 文献中还可见其他关于 Madaline 敏感性研究工作的发表. 例如, Alippi 等^[7] 推广了 Widrow 他们的超球模型, 研究了以离散多段函数为激活函数的广义 Madaline 敏感性. Oh 等^[8] 研究了特定结构 (单隐层) Madaline 敏感性. 近些年来, Zeng 和他的学生们对 Madaline 敏感性进行了深入的研究, 发表了一些敏感性计算和应用的研究结果^[13,14,21]. Zeng 等提出了超立方几何模型^[13], 用超立方顶点来表示 Adaline 输入空间, 并给出了 Adaline 敏感性的精确计算算法. 该算法虽然计算准确且降低了平均计算复杂度, 但是最坏计算复杂度仍然是指数级. 为了进一步降低计算复杂度, Wang 等^[14] 以输入的加权和为随机变量, 建立了随机模型, 在网络无偏置、输入维数足够大的前提条件下, 推导出了 Adaline 敏感性近似计算表达式并设计出了 Madaline 敏感性计算算法.

迄今为止, 已有的对 Adaline 敏感性处理的方法都还存在这样或那样的不足, 归纳起来有: (1) 模型不够准确. 例如文献 [1,2] 中采用超球面来近似表示 Adaline 的输入空间并不确切, 首先, Adaline 的输入空间并不构成一个超球面, 而只是超球的一个内接超立方的所有顶点; 其次, 可以证明随着输入维数的增加超立方的顶点在其外接超球面上将会越来越稀疏; 再就是, 虽然利用超球对称性大大简化了计算, 可是考虑超球表面某段弧面的面积时完全忽略了弧面的具体位置, 而面积相同但位置不同的弧面有可能包含不同个数的输入模式点. (2) 结论假设条件过强, 或计算复杂度过高, 难以实际应用. 例如文献 [3,4] 的研究结论只适用于高维输入、 $|\Delta W| \ll |W|$ 或 $|\Delta X| \ll |X|$ 的情况; 文献 [6] 中的随机模型要求诸参数均值为零, 这样的条件通常只适用于有大样本的一类网络, 而不适用于个体网络; 文献 [14] 所采用的技术和方法仅能用于推导无偏置、输入维数高的 Adaline 敏感性近似计算公式; 文献 [13] 的指数级复杂度难以适用于高输入维数的 Madaline. (3) 对 Adaline 近似计算误差缺乏理论解释和分析, 仅仅是根据对有限的实验结果观察来下结论. 仔细分析以上的不足, 不难发现它们有一个共同点, 即将 Adaline 敏感性计算问题从原本的离散问题转换为连续问题来近似处理. 尽管有许多成熟的连续技术可用, 但仍遇到求连续随机变量分布密度以及复杂积分计算等难题. 为了解决这些难题, 只好通过附加更多的约束条件来达到简化的目的, 这是导致以往研究结论带有过强假设条件、缺乏一般性的重要原因. 因此, 要全面、完整地解决 Adaline 敏感性近似计算问题, 势必要选择一条有别于以往的技术路线, 并使得由此得出的 Adaline 敏感性研究结果能从理论上合理地解释实验现象, Adaline 敏感

性计算公式能更具一般性地涵盖已有的研究成果.

对于 Madaline 敏感性计算的研究, 一般都采用自底向上的方法, 即从前往后逐层计算各 Adaline 敏感性, 最后得到的输出层上各 Adaline 的敏感性就构成 Madaline 的敏感性. 然而, 不同于单独的 Adaline, 作为网络节点的 Adaline 有着更为复杂的计算环境. 在单独的 Adaline 敏感性研究中, 为了讨论的简单方便, 输入参数的扰动是被视为确定事件来处理的, 而事实上, 作为 Madaline 网络节点的 Adaline 所面对的输入参数扰动更多的是一随机事件. 这种输入参数扰动的随机性反映在两个方面: (1) 外界对 Madaline 输入的干扰通常会表现出一定的随机性, 作为网络首层节点的 Adaline 其敏感性计算理应要考虑输入参数扰动的这种随机性; (2) Madaline 的结构特点决定了其前层节点输出的变化是后继层节点相应输入的扰动, 该输入扰动就是以概率形式给出的前层节点的敏感性. 输入参数扰动的随机性使得作为网络节点的 Adaline 敏感性计算面临高复杂度的组合计算难题. 面对这种组合计算问题, 文献 [3, 4, 14] 都采用简单平均方法来达到简化计算的目的, 即用平均扰动概率取代所有输入分量的实际扰动概率, 显然这种不加区别的近似处理必将增大计算误差, 且这种误差还将随着逐层计算不断传递和扩大, 最终影响 Madaline 敏感性的计算精度. 为此, 如何降低因组合计算问题带来的复杂性, 而又尽可能降低精度损失, 是研究 Madaline 敏感性近似计算的关键问题之一.

本文工作的主要贡献在于: (1) 全面分析总结了已有代表性的 Adaline 和 Madaline 敏感性研究方法和结果, 找出了存在的问题和导致的原因; (2) 进一步完善了 Adaline 敏感性计算理论, 采用了一种更加适合 Adaline 离散特性的离散随机技术, 推导出了无约束条件、低计算复杂度和低计算误差的 Adaline 敏感性近似计算公式; (3) 对近似计算误差展开了理论分析, 给出了一些定性的分析结果; (4) 为了降低 Madaline 敏感性计算中面临的组合计算的高复杂度, 同时尽可能地避免计算精度的损失, 提出了一种基于贡献度大小的取舍策略方法; (5) 设计了更加全面的实验方案, 用大量的仿真模拟实验验证了所给公式和算法的准确性和有效性.

2 Adaline 与 Madaline 模型及符号约定

神经元 Adaline 是神经网络 Madaline 的基本构造元素, 构成网络中的一个节点, 也是 Madaline 的最简单形式. 结构上, Adaline 一般由输入、连接权、阈值、激活函数以及输出等几个组成部分. 在本文中, 输入表示为向量 $X = (x_1, x_2, \dots, x_n)^T \in \{-1, 1\}^n$, 与之相关联的权记为 $W = (w_1, w_2, \dots, w_n)^T \in R^n$, $w_0 \in R$ 是一阈值, 激活函数 $f(\cdot)$ 为对称硬限幅函数, $y \in \{-1, 1\}$ 是输出. 就功能而言, Adaline 实现了一种从输入空间到输出空间的逻辑映射关系, 即 $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$, 可表示如下:

$$y = f\left(\sum_{j=1}^n x_j w_j + w_0\right) = \begin{cases} -1, & \sum_{j=1}^n x_j w_j + w_0 < 0; \\ +1, & \sum_{j=1}^n x_j w_j + w_0 \geq 0. \end{cases} \quad (1)$$

图 1 给出了 Adaline 实现映射功能的几何示意图, 可直观地解释 Adaline 的工作原理. 由于输入、输出的二值 $(-1, 1)$ 特性, Adaline 的输入空间是以坐标原点为中心的 n 维单位超立方顶点的集合, 这些顶点均匀地分布在同样以坐标原点为球心的 n 维外接超球面上. Adaline 的输入空间可被 n 维超平面 $P: \sum_{i=1}^n w_i x_i + w_0 = 0$ 一分为二, 分别对应 Adaline 的两个不同输出. 图 1(a) 为 Adaline 的超立方/超球面模型的平面示意图. 为了更好地利用超球面关于原点的对称性, 可将 Adaline 的权及输入加以扩展, 各自加入偏置 w_0 分量及其所对应的输入 $x_0 (\equiv 1)$ 分量, 即 $W = (w_0, w_1, w_2, \dots, w_n)^T \in R^{n+1}$,

图 1 Adaline 实现映射功能的示意图

Figure 1 The mapping function of Adaline

(a) The hypercube and hypersphere model; (b) the semi-hypercube and semi-hypersphere model

$X = (x_0, x_1, x_2, \dots, x_n)^T \in \{-1, 1\}^{n+1}$. 这样, 从 x_0 轴看, Adaline 的输入空间将集中在 x_0 轴的正半轴部分, 相应地 Adaline 的几何模型将调整为 $n+1$ 维半超立方/半超球面模型. 图 1(b) 是 Adaline 半超立方/半超球面模型平面示意图. 扩充后的 $n+1$ 维超平面 $P: \sum_{i=0}^n w_i x_i = 0$ 将总是通过坐标原点.

Madaline(multi-Adalines) 是由多个 Adaline 神经元按一定规则连接成的前向神经网络. 在网络中, 神经元分层组织, 每层由若干个具有相同输入的神经元组成, 相邻两层上的神经元彼此互相连接, 同层和非相邻层上的神经元彼此没有连接. 各层神经元的输出构成该层的输出, 前层输出作为相邻后继层每个神经元的输入, 第一层神经元的输入为网络输入, 最后一层输出为网络输出. 为表达方便, 本文下面用记号 $n^0 - n^1 - \dots - n^L$ 来表示一个有固定结构的 Madaline, 其中, $n^l (1 \leq l \leq L)$ 既代表一层又指出该层上 Adaline 的个数, n^0 代表的是不含神经元的输入层并指出输入维数, n^L 代表的是输出层. 此外, 用 $X^j, Y^j (1 \leq j \leq L)$ 分别表示 Madaline 第 j 层的输入、输出, 且 $Y^{k-1} = X^k (2 \leq k \leq L)$, 即第 $k-1$ 层的输出是第 k 层各 Adaline 神经元的输入, 特别地, X^1 为网络第 1 层的输入, 也是 Madaline 的输入, 网络最后一层 (第 L 层) 的输出 Y^L 是 Madaline 的输出. 功能上, Madaline 是通过自身的学习机制调整权参数值来实现在给定的训练数据中蕴涵的输入输出之间的对应关系, 从而建立从 n^0 维输入空间到 n^L 维输出空间某种特定的逻辑函数的映射 $f: \{-1, 1\}^{n^0} \rightarrow \{-1, 1\}^{n^L}$.

3 Adaline 敏感性

如何定义合适的敏感性使其能反映出 Adaline 的输出受参数扰动影响的大小? 实际上, 敏感性研究就是要找出输出变化与参数扰动之间的函数依赖关系, 并据此关系来分析 Adaline 的一些行为特性和构造一个可评测 Adaline 输出变化的尺度. 影响 Adaline 输出的主要参数有权和输入. 由于对输入扰动的处理可转化为对权扰动的处理, 为了简化起见且又不失一般性, 首先仅就权有扰动的情况来讨论. 对于 Adaline, 最自然和直接地表示因权扰动而导致输出产生变化的方法可由下式给出.

$$\Delta y = f(X^T(W + \Delta W)) - f(X^T W). \quad (2)$$

显然, 式 (2) 中输出变化是输入、权和权扰动的函数, 给定这些自变量的值就能轻易地计算出输出变化的量值. 但是在实际应用中, 这样直接算得的结果有时并没有多大的实用价值, 例如不能仅根据对一个输入的结果来衡量给定 Adaline 的有关性能. 敏感性应该在既能反映输出变化与权扰动之间依

赖关系的同时也能从整体上反映 Adaline 的行为特性. 另外, 由于 Adaline 输出是个二值的开关量, 没有必要细究其变化的具体量值. 基于这些考虑, 我们将输入参数作为统计变量来处理, 用输出发生变化的个数来取代输出变化的量值, 从而给出下面的敏感性定义.

3.1 基于权扰动的敏感性定义

定义 1 对于输入 $X \in \{-1, 1\}^n$, 给定权值 $W \in R^{n+1}$, 和权的扰动值 $\Delta W \in R^{n+1}$, Adaline 的敏感性定义为对于所有可能的输入模式点因权参数的扰动而引起的输出变化的概率, 记成

$$s(\Delta W) = \frac{V_{\text{var}}}{V_n}, \quad (3)$$

其中, V_n 是所有输入模式点的个数, V_{var} 是所有输出发生变化的模式点个数, 可用解析式表示为

$$V_{\text{var}} = \frac{1}{2} \sum_{q=1}^{V_n} |f(X_q^T(W + \Delta W)) - f(X_q^T W)|. \quad (4)$$

上述敏感性的定义仅涉及了权参数的扰动, 而现实中输入参数的扰动也是一个不容忽视的因素. 理论上, 输入扰动可以转换成等价的权扰动来考虑. 因为输入元素是二值量, 对任一元素 $x_i (1 \leq i \leq n)$, 扰动结果只可能有 $x'_i = -x_i$. 因而有如下等价变换关系:

$$x'_i w_i = (-x_i) w_i = x_i (-w_i) = x_i w'_i, \quad (5)$$

其中 $w'_i = -w_i = w_i - 2w_i$, 从而 $\Delta w_i = -2w_i$. 这就证明了输入参数的扰动可以归结为等价的权扰动来考虑. 对于输入和权都有扰动的情况, 类似地也可归为仅对权扰动的考虑. 因此, 针对权扰动的敏感性定义和计算方法可以方便地推广到有输入扰动的情况.

3.2 权扰动下的敏感性近似计算

从 Adaline 的半超球面模型可以看出, 权参数的扰动将导致其超平面的偏转, 形成一新的扰动超平面, 扰动前、后的两个超平面将在半超球面上围出一段弧面 (该弧面称为扰动弧面, 记为 G 其所对应的夹角称为扰动角, 记为 θ), 而正是位于该弧面上的所有输入模式点使得 Adaline 的输出在权扰动后发生变化 ($-1 \rightleftharpoons 1$).

图 2 为权扰动下的 Adaline 半超球模型平面示意图. 图中, 扰动前、后的超平面 P 、 P' 将半超球面分为三段弧面, 在半超球面上夹角 θ 所对应的弧面便是扰动弧面 G (这里, 假定 $w_0 w'_0 > 0$, 否则, 为另外两个弧面), 位于扰动弧面 G 上的输入模式点便是使 Adaline 的输出发生变化的模式点.

依据式 (3) 敏感性的定义, 计算权扰动下 Adaline 的敏感性关键是统计出扰动弧面 G 上所含的输入模式点个数, 即 V_{var} . 由于对 V_{var} 作出精确统计的时间复杂度太高^[13], 因此, 对 V_{var} 进行合理的近似计算是一种在降低计算复杂度与保持计算精度之间取得一定平衡的解决方法. 显然, V_{var} 值的大小与扰动弧面 G 的面积大小成正比关系, 然而, 由于输入模式点在半超球面上仅仅是均匀分布而非紧密分布, 对于给定的输入空间而言, 扰动弧面 G 所含输入模式点个数 V_{var} 与该扰动弧面 G 在半超球面上所处的位置也是有关的, 换言之, 相同大小的扰动弧面 G 所含输入模式点个数有可能随着其所处位置不同而呈现一定程度的差异性, 这表明对应于整个一类大小相同的扰动弧面 G , 其所含输入模式点个数 V_{var} 是一个随机变量. 换个角度看, 相对于扰动弧面 G 而言, 对于任意输入模式点 X_i , X_i 要么

图 2 扰动前后的 Adaline 半超球面模型

Figure 2 The semi-hypersphere model before and after the weight perturbation

位于扰动弧面 G 上, 要么不位于扰动弧面 G 上. 为此引入随机变量 ξ_i 表示第 $i (1 \leq i \leq 2^n)$ 个输入模式点 X_i 所处的区域 (G 区域或非 G 区域), 并令

$$\xi_i = \begin{cases} 1, & X_i \in G, \\ 0, & X_i \notin G, \end{cases} \quad (6)$$

则有随机变量 ξ_i 服从 (0-1) 分布, 记为 $\xi_i \sim (0-1)$. 由于所有输入模式点均匀分布在 $n+1$ 维半超球面上, 因而 ξ_i 服从分布律为 $2A_G/A_{n+1}$ 的 (0-1) 分布 (这里, A_G 表示半超球面上的扰动弧面 G 的面积, A_{n+1} 为 $n+1$ 维超球面面积).

由式 (6) 可得, $V_{\text{var}} = \sum_{i=1}^{2^n} \xi_i$. 由统计学知识可知, 随机变量 V_{var} 将服从参数为 $2^n, 2A_G/A_{n+1}$ 的二项分布, 记为 $V_{\text{var}} \sim B(2^n, 2A_G/A_{n+1})$. 另外, 注意到扰动弧面 G 的面积与两超平面夹角 θ 的大小成正比关系. 因此可求得 V_{var} 的均值作为它的近似值, 记为 $\hat{V}_{\text{var}}$, 从而有

$$\hat{V}_{\text{var}} = E(V_{\text{var}}) = 2^n \left(\frac{2A_G}{A_{n+1}} \right) = 2^n \frac{\theta}{\pi}. \quad (7)$$

由式 (7) 及 Adaline 敏感性定义式 (3) 可进一步求得 Adaline 相对于权扰动下的敏感性近似计算式:

$$\hat{s}(\Delta W) = E(s) = \frac{E(V_{\text{var}})}{V_n} = \left(2^n \frac{\theta}{\pi} \right) \frac{1}{2^n} = \frac{\theta}{\pi}. \quad (8)$$

式 (8) 中, 与扰动弧面 G 相对应的扰动角 θ 的大小可由扰动前后的超平面 P, P' 的法向量, 即扰动前后的权 W, W' , 进一步求得. 因此, Adaline 权扰动下的敏感性近似计算式可进一步表达如下:

$$s(\Delta W) \approx \hat{s}(\Delta W) = \frac{\theta}{\pi} = \frac{1}{\pi} \arccos \left(\frac{WW'}{|W||W'|} \right) = \frac{1}{\pi} \arccos \left(\frac{\sum_{i=0}^n w_i w'_i}{\sqrt{\sum_{i=0}^n w_i^2} \sqrt{\sum_{i=0}^n w'^2_i}} \right). \quad (9)$$

因为当 $|\Delta W| \ll |W|$ 时, 有 $\tan \theta \approx \frac{|\Delta W|}{|W|} \approx \theta$ 成立. 因而在该约束条件下, Adaline 权扰动下的敏感性近似计算表达式可进一步简化为

$$s(\Delta W) \approx \hat{s}(\Delta W) = \frac{\theta}{\pi} \approx \frac{1}{\pi} \frac{|\Delta W|}{|W|} = \frac{1}{\pi} \frac{\sqrt{\sum_{i=0}^n \Delta w_i^2}}{\sqrt{\sum_{i=0}^n w_i^2}}, \quad \text{当 } |\Delta W| \ll |W|. \quad (10)$$

与式 (9) 相比, 式 (10) 形式更简洁, 但这种形式上的简洁是以部分计算精度的损失和强约束条件为代价的. 式 (10) 与 Stevenson 等得出的权扰动下 Adaline 近似计算式^[3,4] 完全一致. 这也从一个侧面反映了本文方法的正确性以及理论结论的一般性. 从式 (9), (10) 可以看出, Adaline 敏感性近似计算的时间复杂度仅仅是其输入维数的多项式级别, 即 $O(n)$.

3.3 输入扰动下的敏感性近似计算

为了统计因输入扰动导致 Adaline 输出发生变化的模式点个数, 去遍历 Adaline 输入空间中的所有模式点显然是件很耗时的工作. 庆幸的是, 由式 (5) 可知, Adaline 的输入扰动可以方便地转化为相应权扰动来处理, 即输入分量 $x_i (1 \leq i \leq n)$ 的扰动等价于相应权分量 $w_i (1 \leq i \leq n)$ 的扰动, 且扰动值为 $\Delta w_i = -2w_i$. 这样, 由权扰动下 Adaline 敏感性近似计算式 (9) 可以得到存在 k 个输入分量扰动下 Adaline 敏感性近似计算式如下:

$$s(\Delta X) \approx \hat{s}(\Delta X) = \frac{1}{\pi} \arccos \left(\frac{\sum_{i=0}^n w_i w'_i}{\sqrt{\sum_{i=0}^n w_i^2} \sqrt{\sum_{i=0}^n w'^2_i}} \right) = \frac{1}{\pi} \arccos \left(\frac{\sum_{i=0}^n w_i w'_i}{\sum_{i=0}^n w_i^2} \right), \quad (11)$$

其中,

$$w'_i = \begin{cases} -w_i, & i \in \{j_1, j_2, \dots, j_k\}, \\ w_i, & \text{其他}, \end{cases}$$

$j_t (1 \leq t \leq k)$ 表示 Adaline 第 j_t 个输入分量有扰动.

由于当 Adaline 的输入扰动分量个数远少于输入维数 ($k \ll n$) 时, $|\Delta W| \ll |W|$ 成立. 同时, 与 $\sum_{i=0}^n w_i^2$ 相比, $w_i^2 (0 \leq i \leq n)$ 足够小. 这时, 可合理地认为 $w_{j_1}^2 / \sum_{i=0}^n w_i^2 \approx \dots \approx w_{j_k}^2 / \sum_{i=0}^n w_i^2 \approx 1/(n+1) (1 \leq j_k \leq n)$ 成立. 这样, 依据权扰动下 Adaline 敏感性近似计算式 (10) 可以进一步得到输入扰动下 Adaline 敏感性近似计算简约式如下:

$$\begin{aligned} s(\Delta X) \approx \hat{s}(\Delta X) &\approx \frac{1}{\pi} \frac{|\Delta W|}{|W|} = \frac{1}{\pi} \frac{\sqrt{\sum_{i=0}^n \Delta w_i^2}}{\sqrt{\sum_{i=0}^n w_i^2}} = \frac{1}{\pi} \sqrt{\frac{4 \sum_{h=1}^k w_{j_h}^2}{\sum_{i=0}^n w_i^2}} \\ &\approx \frac{1}{\pi} \sqrt{\frac{4k}{n+1}} = \frac{1}{\pi} \frac{|\Delta X|}{|X|}, \quad \text{当 } k \ll n, \end{aligned} \quad (12)$$

其中,

$$\Delta w_i = \begin{cases} -2w_i, & i \in \{j_1, j_2, \dots, j_k\}, \\ 0, & \text{其他}, \end{cases}$$

$j_t (1 \leq t \leq k)$ 表示 Adaline 第 j_t 个输入分量有扰动.

式 (12) 表明, 当输入扰动满足 $k \ll n$ 时, Adaline 敏感性的大小仅与输入扰动分量个数 k 以及输入维数 n 有关, 而与具体的扰动分量无关. 这对于估算任意 k 个输入分量扰动 ($k \ll n$) 下的 Adaline 敏感性提供了更为便捷的计算形式. 值得一提的是, 式 (12) 与 Stevenson 等^[3,4] 得出的输入扰动下 Adaline 敏感性近似计算式具有相同的形式. 式 (10), (12) 的结论表明, 本文方法在推导出一般化的计算公式的同时, 还可以涵盖已有的研究结论.

3.4 权和输入都有扰动的敏感性近似计算

当 Adaline 输入扰动、权扰动同时存在时, 可以将两种不同形式的扰动统一为单一形式的权扰动. 为区别起见, 在权、输入扰动并存时, 将权扰动后的权向量记为 W' , 而将统一后的等价权向量记为 W'' . 这样, 统一后的权 W'' 可表示为

$$W'' = (w''_0, \dots, w''_i, \dots, w''_n)^T, \text{ 其中, } w''_i = \begin{cases} w_i + w'_i, & x_i \text{ 无扰动}, \\ -(w_i + w'_i), & x_i \text{ 有扰动}. \end{cases} \quad (13)$$

图 3 针对不同权扰动, Adaline 的真实敏感性与近似敏感性的比较

Figure 3 Comparison between the actual sensitivity and the approximate sensitivity under different weight perturbations

(a) 10-dimension Adaline; (b) 15-dimension Adaline; (c) 20-dimension Adaline

这样, 依据权扰动下 Adaline 敏感性近似计算式 (9), 可以方便地处理输入、权扰动同时存在的 Adaline 敏感性近似计算问题.

3.5 Adaline 敏感性近似计算公式的实验验证

鉴于输入扰动能转化为相应等价的权扰动, 因此下面只针对权扰动下的 Adaline 敏感性近似计算式, 即式 (9) 和 (10), 进行了仿真验证.

3.5.1 对式 (9) 的实验验证

为了验证公式 (9) 的正确性, 特地设计了下列一组实验. 实验结果分别以图 3 的形式给出. 实验设计如下: 针对 10 维、15 维、20 维输入, 各任取一 Adaline 作为实验对象, 分别选取它们在权分量扰动值为 0.1, 0.2, ..., 1.0 时的敏感性真实值和近似值作为观察数据. 实验结果如图 3 所示.

从图 3 可以看出, 随着输入维数的增加, Adaline 敏感性的近似计算误差越来越小的变化趋势相当明显, 当输入维数增加到 10 维或 10 维以上时, Adaline 的敏感性近似值和真实值的接近程度相当完美.

3.5.2 对式 (10) 的实验验证

鉴于式 (10) 是对式 (9) 加约束条件和简化所得, 且与文献 [4] 给出的近似公式相一致, 因此该实

图 4 针对不同 $|\Delta W|/|W|$ 扰动比值, Adaline 敏感性近似值与真实敏感性的比较

Figure 4 Comparison between the actual sensitivity and the approximate sensitivity under different weight perturbation ratios

(a) 10-dimension Adaline; (b) 15-dimension Adaline; (c) 20-dimension Adaline

验既是对式 (10) 的有效性和适用性的验证, 即观察它的近似精度和约束条件是否实用, 也是对式 (9) 更具一般性的验证. 为了与式 (9) 实验结果作比较, 本节沿用 3.5.1 小节实验的设计, 差别仅仅在于用式 (10) 取代原来式 (9) 来计算 Adaline 敏感性近似值, 同时为了便于观察权扰动比值 $|\Delta W|/|W|$ 对式 (10) 计算得到的 Adaline 敏感性近似值的影响, 将实验观测视角由原来的权扰动量替换为相应的 $|\Delta W|/|W|$ 扰动比值. 实验结果如图 4 所示.

从图 4 可以看出, 随着权扰动比值 $|\Delta W|/|W|$ 的增大, 式 (10) 计算得到的 Adaline 敏感性近似值越来越偏离其真实值, 仅在 $|\Delta W|/|W|$ 很小时, Adaline 敏感性近似值与其真实值具有令人满意的逼近程度. 更多的实验进一步表明, 式 (10) 在 $|\Delta W|/|W| \leq 30\%$ 时具有良好的适用性, 这与文献 [3] 观察得到的经验值是一致的. 该实验结果也反映了式 (10) 仅是式 (9) 的一个特例的事实, 其所带的约束条件是一个强约束条件, 这也是阻碍文献 [3, 4] 理论结果推广应用的主要因素.

4 Adaline 敏感性近似误差分析

对于任何近似计算而言, 总是以牺牲一定的计算精度为代价的. 因而分析近似计算误差是近似计算研究的重要一环.

4.1 近似误差的理论分析

由统计学知识可知, 均值描述了随机变量的取值中心, 而方差可进一步描述随机变量的可能取值

与均值的偏离程度. 因而, 作为均值意义下的 Adaline 敏感性近似值, 其近似误差情况可以借助于方差来加以分析.

由 3.2 小节可知, 随机变量 V_{var} 服从二项分布, 即 $V_{\text{var}} \sim B(2^n, 2A_G/A_{n+1})$, 因而可求得随机变量 V_{var} 的方差为

$$D(V_{\text{var}}) = 2^n \frac{\theta}{\pi} \left(1 - \frac{\theta}{\pi}\right). \quad (14)$$

由 Adaline 敏感性定义式 (3) 和 (14) 可进一步得到 Adaline 敏感性 s 的方差 $D(s)$,

$$D(s) = \frac{D(V_{\text{var}})}{2^{2n}} = \frac{1}{2^n} \frac{\theta}{\pi} \left(1 - \frac{\theta}{\pi}\right) = \frac{1}{2^n \pi^2} (\pi\theta - \theta^2). \quad (15)$$

令 $f(\theta) = \pi\theta - \theta^2$, 显然 $f(\theta)$ 是个二次凸函数, 在 $\theta = \pi/2$ 取得最大值. 因此, 通过式 (15), 可以得出如下有关本文 Adaline 近似计算误差的理论分析结论:

- (1) Adaline 敏感性近似计算误差, 直接取决于扰动角 θ 和输入维数 n ;
- (2) 当 Adaline 输入维数 n 一定时, s 在区间 $[0, \pi]$ 的两端近似计算误差相对小些;
- (3) 当 Adaline 扰动角 θ 一定时, s 近似计算误差将随着输入维数 n 的增大而减少;
- (4) 当 Adaline 输入维数 n 大到一定值时, 维数 n 将成为影响 s 近似计算误差的主要因素, 而扰动角 θ 的影响越来越微.

上述分析的结论, 从理论上合理地解释了已有的基于实验观察得出的结论^[3,4,13,14], 即对 Adaline 敏感性的近似计算, 输入维数越大, 近似精度越高. 同样, 在 3.5.1, 3.5.2 小节实验中得到的结论这里也得到了合理的理论解释.

4.2 近似误差的实验验证

为了进一步验证上述关于 Adaline 敏感性近似计算误差的理论分析结论的正确性, 我们进行了大量的实验验证工作. 不同于 3.5.1, 3.5.2 小节的实验中以一个 Adaline 为实验对象, 下面的实验更突显统计特性, 实验中的每个结果数据是取多个 Adaline 实验结果的均值. 结果数据以绝对误差的形式用图加以直观表示, 绝对误差定义如下:

$$\text{绝对误差} = |\text{真实敏感性} - \text{近似敏感性}|. \quad (16)$$

显然, 绝对误差反映了 Adaline 近似敏感性与真实敏感性之间的实际偏差大小.

从式 (9) 可以看出, 虽然 Adaline 的敏感性大小是由其扰动前后的权参数所决定, 但与扰动前后的权参数大小之间不存在有规律的数量关系, 而与扰动前后的权向量所形成的扰动角大小呈线性关系. 因此, 与权扰动量相比, 扰动前后的权向量所形成的扰动角将是一个更为理想的观测视角. 在本节实验中, 针对每一输入维数 (5, 10, 15, 20), 随机选择该输入维数下的 Adaline, 在 $[-1, 1]$ 区间内做随机权扰动的敏感性计算实验, 取扰动角大小落入区间 $(0^\circ, 5^\circ), \dots, (\lambda - 5^\circ, \lambda + 5^\circ), \dots, (175^\circ, 180^\circ)$ ($\lambda = 10^\circ, 20^\circ, \dots, 170^\circ$) 之一的 Adaline 的敏感性真实值和近似值作为观测数据, 直到每个区间都得到 20 个相应的观测数据. 再依据式 (16) 计算它们相应的绝对误差值, 最后取其绝对误差的平均值作为该观察点的结果数据. 实验结果如图 5 所示.

图 5 大小不同的扰动角下, Adaline 敏感性近似计算的绝对误差

Figure 5 The absolute error of the Adaline sensitivity under different perturbed angles

从图 5 中可以看出: 一方面, 随着 Adaline 输入维数的增大, 绝对误差迅速缩小. 其中, 除 5 维输入的绝对误差处于 $[0.02, 0.06]$ 之间外, 其他各维输入下的绝对误差都小于 0.01. 15 维和 20 维的绝对误差进一步收窄到千分位级; 另一方面, 对于输入维数确定的 Adaline, 在扰动区间 $[0, \pi]$ 的两端, 绝对误差来得小些. 这与本节有关 Adaline 敏感性近似计算误差的理论分析结论是完全一致的.

5 Madaline 敏感性

Madaline 敏感性计算是以 Adaline 敏感性为基础, 也是 Adaline 敏感性计算研究的主要目的之一. 与 Ad-aline 类似, Madaline 敏感性可定义为:

定义 2 对于输入 $X \in \{-1, 1\}^n$, 给定的权值 $W_i^l \in R^{n^{l-1}+1}$ 和权扰动值 $\Delta W_i^l \in R^{n^{l-1}+1}$ ($1 \leq i \leq n^l, 1 \leq l \leq L$), 以及输入扰动概率 P , Madaline 敏感性定义为对于所有可能的输入模式点因输入、权参数的扰动而引起网络输出发生变化的概率, 可记为

$$s_{\text{net}} = \frac{V_{\text{var}}}{V_n}, \quad (17)$$

其中, V_n 是所有输入模式点个数, V_{var} 是所有输出发生变化的模式点个数.

显然, 若要 Madaline 的输出不发生变化, 当且仅当其输出层各神经元 Adaline 的输出不发生变化. 由 A-daline 的定义可知, Madaline 输出层第 i 个神经元的输出不发生变化概率为 $(1 - s_i^L)$ ($1 \leq i \leq n^L$), 因此, 依据概率知识可以将 Madaline 的敏感性进一步表达如下:

$$s_{\text{net}} = 1 - \prod_{i=1}^{n^L} (1 - s_i^L), \quad (18)$$

其中, s_i^L ($1 \leq i \leq n^L$) 表示 Madaline 输出层第 i 个 Adaline 神经元的敏感性.

5.1 作为网络节点的 Adaline 敏感性近似计算

鉴于 Madaline 层间节点存在前后依赖关系, 第 3 节推导的有关 Adaline 敏感性近似计算表达式不能简单地用来求解作为 Madaline 节点的 Adaline 的敏感性, 本节将在前面 Adaline 敏感性计算的基础上, 推导适合作为 Madaline 节点的 Adaline 敏感性的更一般化计算表达式.

由 Madaline 结构可知, 当网络隐层中的某一神经元存在输出变化时, 即存在 $s_i^l > 0 (1 \leq i \leq n^l, 1 \leq l < L)$, 通过网络层间的传递, 其后继层各神经元的相应输入分量将有扰动发生, 并且这种扰动是以概率 s_i^l 形式给出的. 除此之外, 因环境干扰的随机性, Madaline 第一层各 Adaline 的输入扰动也可能是以概率形式给出的. 因此, 与单独的 Adaline 相比, 作为 Madaline 节点的 Adaline 的敏感性计算具有更为复杂的计算环境, 这种复杂性体现在输入参数扰动的随机性表示上.

为了反映 Adaline 的输入扰动, 在不影响权扰动表示的情况下, 给 Adaline 的每一输入分量 x_i 赋予一扰动概率 p_i , $p_i = 0$ 表示输入分量 x_i 不存在扰动, 而 $p_i = 1$ 表示输入分量 x_i 一定发生扰动. 显然, 该模型统一了 Madaline 中各层上 Adaline 输入扰动的表示, 并且涵盖第 3 节中讨论的情况, 更具一般性.

依据全概率公式, 要计算一个带输入扰动概率的 n 维 Adaline 敏感性, 必需考虑从 1 到 n 个不同输入分量的各个扰动组合情况下的 Adaline 敏感性, 这样的组合共有 $2^n (= C_n^0 + C_n^1 + \cdots + C_n^n)$ 个. 因此, 最终作为网络神经元的 Adaline 敏感性计算式可表达为

$$s(\Delta W, P) = \sum_{i=1}^{2^n} p(D_i) s(\Delta W, \Delta X_{D_i}), \quad (19)$$

式中, $\Delta W = (\Delta w_0, \Delta w_1, \dots, \Delta w_n)^T$ 为权扰动向量, 各输入分量发生扰动的概率向量为 $P = (p_1, \dots, p_n)^T$, D_i 表示 2^n 个可能的输入扰动组合中第 i 个输入扰动组合, $p(D_i)$ 表示第 i 个输入扰动组合 D_i 发生的概率, ΔX_{D_i} 表示第 i 个输入扰动组合 D_i 情况下的 Adaline 输入扰动向量, $s(\Delta W, \Delta X_{D_i})$ 表示在权扰动 ΔW , 输入扰动 ΔX_{D_i} 情况下 Adaline 的敏感性. 设 D_i 所代表的输入扰动组合为 $k_1 k_2 \cdots k_n (k_j \in \{0, 1\}, 1 \leq j \leq n)$, 当 $k_j = 1$ 表示在该组合中 Adaline 第 j 个输入分量发生扰动, 当 $k_j = 0$ 表示未发生扰动. 显然,

$$p(D_i) = \prod_{\substack{m=1 \\ k_m=1}}^n p_m \prod_{\substack{l=1 \\ k_l=0}}^n (1 - p_l). \quad (20)$$

由第 3 节 Adaline 敏感性近似计算研究结论可知, 式 (19) 中 $s(\Delta W, \Delta X_{D_i})$ 可以通过式 (13) 等价变换为统一的权扰动, 然后由式 (9) 计算得到. 因此, Madaline 中的 Adaline 神经元的敏感性, 可通过分别计算出式 (19) 中的每一个输入扰动组合下的 Adaline 敏感性而得到.

然而, 由于式 (19) 要分别计算 2^n 种输入扰动组合情况下的 Adaline 敏感性, 而每种情况下的 Adaline 敏感性计算时间复杂度为 $O(n)$, 因而式 (19) 的时间复杂度为 $O(2^n \cdot n)$. 此外, 还要考虑 Madaline 网络中的 Adaline 个数. 如此高的指数级复杂度很难满足 Madaline 敏感性的应用需求. 为此, 如何解决 Madaline 网络中 Adaline 神经元敏感性计算的时间耗费问题将是有关 Madaline 敏感性计算课题中的关键问题.

面对作为网络节点的 Adaline 敏感性计算高时间耗费问题, 本文提出了如下解决思想: 根据 Adaline 敏感性的实际量值大小, 在计算 (19) 式中每一可能输入扰动组合下的 Adaline 敏感性时, 仅仅考虑计算那些对 Adaline 敏感性贡献较大的那部分组合, 而舍弃贡献小的部分.

为了实现上述思想, 可对任意组合下的 Adaline 敏感性量值预先设定一个取舍精度 (记为 α), 最终仅仅计算敏感性值大于或等于精度要求的组合部分, 而舍弃小于精度的组合部分. 那么这个取舍精度值 α 该如何设定呢? 鉴于通常情况下, 敏感性的精度取 0.0001 就足够了. 因而, 我们可以把任意输入扰动组合下的 Adaline 敏感性取舍精度 α 合理地设为 0.000 001, 即百万分位级. 当某一输入扰动组合下的 Adaline 敏感性值小于这个 α 时, 说明它对整个 Adaline 敏感性的贡献甚微, 大可忽略不计. 这样式 (19) 可改写为

$$s(\Delta W, P) = \sum_{i=1}^{2^n} p(D_i) s(\Delta W, \Delta X_{D_i}) \approx \sum_j p(D_j) s(\Delta W, \Delta X_{D_j}), \quad \text{当 } p(D_j) s(\Delta W, \Delta X_{D_j}) \geq \alpha. \quad (21)$$

为了减少不需要的敏感性计算, 可以在计算前, 事先对各输入分量的扰动概率按降序排列, 然后从高到低逐个相乘 (不妨把乘积记为 β), 即 $\beta = p_{t_1}, \beta = \beta \cdot p_{t_2}, \dots$, 其中, $t_j (1 \leq t_j \leq n)$ 表示 Adaline 第 t_j 个输入分量, $j (1 \leq j \leq n)$ 表示 p_{t_j} 在降序队列中的序号, 当 $\beta \leq \alpha$ 时即停止累乘, 记下最后的累乘概率个数, 它正好等于最后的 j , 不妨设最后的 $j = H$, 则 H 成了所有需要考虑计算的输入扰动组合中扰动元素 ($p_i > 0, 1 \leq i \leq n$) 个数的最小上限, 因为当所需要考虑计算的组合中扰动元数个数大于 H 时, 有 $p(D_i) s(\Delta W, \Delta X_{D_i}) \leq p(D_i) < \beta = p_{t_1} p_{t_2} \cdots p_{t_H} \leq \alpha$ 成立. 所以, 在式 (21) 的计算中, 我们只要考虑最多有 H 个输入分量扰动的组合情况下 Adaline 敏感性的计算, 这样, 需考虑的输入扰动组合数可由原来的 $2^n (= \sum_{i=1}^n C_n^i)$ 个减少到最多 $\sum_{i=1}^H C_n^i$ 个. 进一步可有如下的分析, 在实际应用中, Adaline 的敏感性的值一般会在 $[0, 0.5000]$ 范围内, 因此最坏的一种情况是每一个输入分量发生扰动的概率都是 0.5 的情况. 由于 $0.5^{20} \approx 0.000\ 001$, 所以, 在 Adaline 敏感性的实际计算中, 值得考虑的输入扰动组合其扰动元素个数最多不超过 20, 即值得考虑的输入扰动组合数最多为 $\sum_{i=1}^{20} C_n^i$. 这样, 在计算 Madaline 的 n 维神经元 Adaline 的敏感性时, 式 (21) 最坏的时间复杂度可降为 $O(c * n)$ (c 为常数).

5.2 Madaline 敏感性近似计算算法

通过式 (21), 按从前往后逐层依次计算各神经元 Adaline 的敏感性, 最终得到网络输出层各 Adaline 神经元的敏感性. 再由式 (18) 可最后计算出 Madaline 敏感性. 具体计算算法如下:

输入: Madaline 的结构 $n^0 - n^1 - \cdots - n^L$, 权 $W_i^l (1 \leq l \leq L, 1 \leq i \leq n^l)$, 权扰动 $\Delta W_i^l (1 \leq l \leq L, 1 \leq i \leq n^l)$, 输入扰动概率 $P = (p_1^0, p_2^0, \dots, p_{n^0}^0)$, 组合计算的取舍精度 α .

- (1) 初始化: $P \Rightarrow S$;
- (2) 对给定 Madaline 所有层, 逐层循环计算:
 - ① 把 S 中元素按降序排序;
 - ② 将 S 排序后的元素从高到底逐次相乘, 直到乘积 $\leq \alpha$;
 - ③ 将连乘中元素的个数 $\Rightarrow H$;
 - ④ 对当前层中所有神经元 Adaline, 逐个循环计算:
 - a) 输入扰动组合数从 1 到 $\sum_{i=1}^H C_n^i$ (其中, i 为输入扰动分量个数), 逐个组合循环计算: 依据式 (13) 和式 (9), 计算当前输入扰动组合下的 Adaline 敏感性.
 - b) 依据式 (21) 计算当前 Adaline 的敏感性.
 - ⑤ 将得到的当前层所有神经元的敏感性赋给 S , 即 $(s_1^i, \dots, s_{n^i}^i) \Rightarrow S$.
- (3) 依据 Madaline 输出层敏感性, 由式 (18) 最终计算出的 Madaline 敏感性近似值 s_{net} .

图 6 Madaline 敏感性近似计算实验结果

Figure 6 The experimental results on the Madaline sensitivity

(a) Madaline with 10-20-1; (b) Madaline with 15-20-1; (3) Madaline with 20-20-1

输出: Madaline 敏感性近似值 s_{net} .

6 Madaline 敏感性近似计算验证实验

为了验证本文第 5 节有关 Madaline 敏感性近似计算方法, 设计如下一组实验: 对结构为 10-20-1, 15-20-1, 20-20-1 的 Madalines 各执行 20 次随机敏感性计算实验, 其中每次实验中各网络节点在 $[-1, 1]$ 区间内取得随机初始权值, 在 $[-0.5, 0.5]$ 区间内取得权随机扰动值.

每次实验选取 Madaline 真实敏感性 (标记为 Actual)、近似敏感性以及相应的近似计算绝对误差作为最后的实验观察值. 为了与已有的研究结论进行比较, 分别用 Wang^[14]、Stevenson^[4] 以及本文的方法计算其中的近似敏感性, 并分别标记为 Wang's, Stevenson's, Ours.

图 6 展示了该组实验的实验结果. 从图 6 可以看出, 与 Wang、Stevenson 提出的方法相比, 本文提出的 Madaline 敏感性近似计算方法得到的结果更加接近真实敏感性, 其近似计算绝对误差也远小于 Wang 和 Stevenson 方法所得值, 并且随着输入维数、隐层神经元个数的增加, 这种优势越来越明显.

7 总结

本文全面总结了已有的 Madaline 敏感性的研究现状, 指出了存在的问题和导致的原因, 提出一种用离散随机技术来定量计算和定性分析 Madaline 敏感性的新思路, 并给出了具体的计算公式和计算算法. 所提的思路充分地适应了 Madaline 的离散特性, 方便了敏感性近似计算公式的推导和误差

分析, 进一步完善了 Madaline 敏感性计算理论; 所给的算法在尽可能减少精度损失的情况下, 大大降低了计算复杂度, 并且较目前已有的研究结果限制条件少, 更具有一般性. 从多个角度精心设计的大量实验充分验证了所给公式和算法的准确性和有效性, 以及敏感性近似计算误差的理论分析所得出的结论.

虽然对 Madaline 敏感性的研究已经取得很大的进展, 并有了成功的应用, 但是我们认为在理论上还有值得深入探索的地方. 首先, 如本文在引言中所指出的, 采用外接超球面模型来近似表示 Adaline 的输入空间不够确切, 如何建立更为确切的数学模型来进一步完善敏感性计算理论是值得深入研究的. 其次, 现有方法计算的敏感性是统计意义下的均值, 而非确定的值, 这可能导致以此为尺度来评判两个或多个 Madaline 网络的敏感性大小时, 出现与实际情况的某种偏离. 究其原因, 敏感性均值是利用了超球的对称性来考虑有相同大小扰动弧面的所有情况, 这虽然简化了计算, 但是忽略了形成扰动弧面的两超平面在空间中的具体位置, 如何充分利用两超平面的位置信息来进一步精确计算 Adaline 敏感性也是值得深入研究的.

接下来, 我们一方面将从理论上完善 Madaline 敏感性的研究, 另一方面将以 Madaline 敏感性作为度量工具, 应用它来解决 Madaline 所面临的一些基本且实际的问题, 如高效学习算法的设计、合理网络结构的构建、高性能网络的实现等. 我们相信, 本文的 Madaline 敏感性研究结果将能够为这些课题的研究提供理论依据和技术支持.

参考文献

- 1 Hoff M E Jr. Learning phenomena in networks of adaptive switching circuits. PhD dissertation. Stanford: Dept Electrical Eng, Stanford Univ, 1962. 1-91
- 2 Glanz F H. Statistical extrapolation in certain adaptive pattern recognition systems. PhD dissertation. Stanford: Dept Electrical Eng, Stanford Univ, 1965. 1-142
- 3 Winter R G. Madaline rule II: a new method for training networks of Adalines. PhD dissertation. Stanford: Dept Electrical Eng, Stanford Univ, 1989. 1-91
- 4 Stevenson M, Winter R, Widrow B. Sensitivity of feedforward neural networks to weight errors. IEEE Trans Neural Netw, 1990, 1: 71-80
- 5 Choi J Y, Choi C H. Sensitivity analysis of multilayer perceptron with differentiable activation functions. IEEE Trans Neural Netw, 1992, 3: 101-107
- 6 Piché S W. The selection of weight accuracies for Madalines. IEEE Trans Neural Netw, 1995, 6: 432-445
- 7 Alippi C, Piuri V, Sami M G. Sensitivity to errors in artificial neural networks: a behavioral approach. IEEE Trans Circuits Syst I-Regul Pap, 1995, 42: 358-361
- 8 Oh S H, Lee Y. Sensitivity analysis of a single hidden-layer neural networks with threshold function. IEEE Trans Neural Netw, 1995, 6: 1005-1007
- 9 Cheng A Y, Yeung D S. Sensitivity analysis of neocognitron. IEEE Trans Syst Man Cybern C, 1999, 29: 238-249
- 10 Zeng X, Yeung D S. Sensitivity analysis of multilayer perceptron to input and weight perturbations. IEEE Trans Neural Netw, 2001, 12: 1358-1366
- 11 Yeung D S, Sun X. Using function approximation to analyze the sensitivity of the MLP with antisymmetric squashing activation function. IEEE Trans Neural Netw, 2002, 13: 34-44
- 12 Zeng X, Yeung D S. A quantified sensitivity measure for multilayer perceptron to input perturbation. Neural Comput, 2003, 15: 183-212
- 13 Zeng X, Wang Y, Zhang K. Computation of Adalines' sensitivity to weight perturbation. IEEE Trans Neural Netw, 2006, 17: 515-519
- 14 Wang Y, Zeng X, Yeung D S, et al. Computation of Madalines' sensitivity to input and weight perturbations. Neural Comput, 2006, 18: 2854-2877

- 15 Zurada J M, Malinowski A, Usui S. Perturbation method for deleting redundant inputs of perceptron networks. *Neurocomput*, 1997, 14: 177–193
- 16 Bernier J L, Ortega J, Rojas I, et al. Improving the tolerance of multilayer perceptrons by minimizing the statistical sensitivity to weight deviations. *Neurocomput*, 2000, 31: 87–103
- 17 Bernier J L, Ortega J, ROS E, et al. A quantitative study of fault tolerance, noise immunity and generalization ability of MLPs. *Neural Comput*, 2000, 12: 2941–2964
- 18 Engelbrecht A P. A new pruning heuristic based on variance analysis of sensitivity information. *IEEE Trans Neural Netw*, 2001, 12: 1386–1399
- 19 Zeng X, Yeung D S. Hidden neuron pruning of multilayer perceptrons using a quantified sensitivity measure. *Neurocomput*, 2006, 69: 825–837
- 20 Yeung D, Ng W, Wang D, et al. Localized generalization error model and its application to architecture selection for radial basis function neural network. *IEEE Trans Neural Netw*, 2007, 18: 1294–1305
- 21 Zeng X, Shao J, Wang Y, et al. A sensitivity-based approach for pruning architecture of Madalines. *Neural Comput Appl*, 2009, 18: 957–965

Approximate computation of Madaline sensitivity based on discrete stochastic technique

ZHONG ShuiMing*, ZENG XiaoQin, LIU HuiYi & XU Yan

Institute of Pattern Recognition and Intelligent System, College of Computer and Information Engineering, Hohai University, Nanjing 210098, China

*E-mail: zhongyi@hhu.edu.cn

Abstract The computation of the sensitivity of a Madaline's output to its parameter perturbation is systematically discussed. Firstly, according to the discrete feature of Adalines, a method based on discrete stochastic technique is proposed, which derives some analytical formulas for the computation of Adalines' sensitivity. The method can theoretically solve some problems that are unsolvable by the existing methods based on continuous stochastic techniques, release some unpractical constraints, and make it available to theoretically analyze the approximation error of Adalines' sensitivity. Secondly, on the basis of the sensitivity of Adalines and the structural characteristics of Madalines, a new selection strategy depending on a type of dedication degree for computing Madalines' sensitivity is proposed, which is superior to current popular way of simply averaging in both precision and complexity. The proposed formulas and algorithm have the advantages of simplicity, low computational complexity, small approximation error, and high generality, as have been verified by a great amount of experimental simulations.

Keywords Adaline, Madaline, feedforward networks, sensitivity, parameter perturbation, stochastic discrete technique