

УДК 004.032.26

С.В. ПОПОВ, К.А. ШКУРО

Харьковский национальный университет радиоэлектроники, Украина

МЕТОД ПАРАМЕТРИЧЕСКОЙ ОПТИМИЗАЦИИ СЕТИ НА БАЗЕ ГИБРИДНЫХ НЕЙРОПОДОБНЫХ ЭЛЕМЕНТОВ, ОСНОВАННЫЙ НА МЕТОДЕ Ψ -ПРЕОБРАЗОВАНИЯ

В статье рассматриваются особенности обучения сети на базе гибридных нейроподобных элементов, вызывающие трудности в нахождении оптимального набора весов. Для поиска глобального экстремума критерия обучения предложен модифицированный метод, основанный на Ψ -преобразовании. Сформирована система ограничений, позволяющая выделить единственную среди множества эквивалентных областей, возникающих вследствие перестановочной и знаковой симметрии пространства параметров сети на базе гибридных нейроподобных элементов, что является необходимым для применения метода Ψ -преобразования.

Ключевые слова: параметрическая оптимизация, обучение, гибридный нейроподобный элемент, метод Ψ -преобразования, симметрия пространства параметров, глобальный экстремум.

Введение

В современных условиях, когда к качеству обработки информации представляются все более высокие требования, как никогда важна роль оптимизационных методов в настройке параметров моделей различного рода. Не являются исключением и искусственные нейронные сети (ИНС) и нейро-фаззи системы (НФС) различной структуры [1], чьи свойства с точки зрения качества получаемых решений в значительной мере определяются качеством процесса их структурной и параметрической настройки (обучения). В данной работе мы будем рассматривать сеть на базе гибридных нейроподобных элементов [2], частными случаями которой являются многие популярные архитектуры ИНС и НФС, в частности, многослойный персептрон, сеть с конечной импульсной характеристикой и другие.

Особое внимание следует уделить процедуре параметрической оптимизации сети, являющейся промежуточным этапом при проведении структурной оптимизации, например, с применением эволюционного подхода [3-4], поскольку качество работы метода структурной оптимизации коренным образом зависит от качества параметрической настройки. Низкокачественный алгоритм обучения может полностью свести на нет все усилия по поиску оптимальной архитектуры, что является недопустимым.

Основные сложности обучения сети на базе гибридных нейроподобных элементов возникают

вследствие нескольких причин:

- высокая размерность оптимизационной задачи;
- сложность рельефа критерия обучения (наличие множества локальных и глобальных экстремумов, плато и оврагов) [5];
- разрывы производных критерия обучения (в частности, при использовании треугольных функций принадлежности в нелинейных синапсах).

Наиболее эффективные методы обучения, используемые в теории искусственных нейронных сетей, используют производные критерия обучения для более эффективного поиска экстремума [6], что, соответственно, требует дифференцируемости самого критерия и функции, реализуемой сетью, во всей области поиска. Это условие выполняется для многих популярных архитектур ИНС. В случае же применения нелинейных синапсов на основе нечеткой системы условие дифференцируемости может нарушаться. В таком случае применение названных методов обучения может оказаться или невозможным или (при использовании дополнительных приемов, восстанавливающих значение производных в точках разрыва) недостаточно эффективным.

В силу перечисленных причин возникает задача создания новых методов обучения, которые были бы эффективны, несмотря на специфику сетей на базе гибридных нейроподобных элементов. В качестве основы для решения этой задачи используем известный метод Ψ -преобразования [7], обладающий рядом замечательных свойств, а именно: спо-

собностью находить глобальный экстремум функции и отсутствием требования дифференцируемости функции.

1. Постановка задачи исследования

Главная особенность метода Ψ -преобразования заключается в том, что объектом исследования и анализа является не сама оптимизируемая функция критерия обучения

$$E(w_1, w_2, \dots, w_{N_p}),$$

где E – критерий обучения;

w_i – настраиваемые параметры (синаптические веса) сети;

N_p – количество настраиваемых параметров сети,

а некоторая функция $\Psi(\zeta)$, образуемая в результате преобразования $E(w_1, w_2, \dots, w_{N_p})$. Функция $\Psi(\zeta)$, а также некоторые другие функции $\bar{w}_i(\zeta), (i=1, \dots, N_p)$, образующиеся в результате преобразования $E(w_1, w_2, \dots, w_{N_p})$ в $\Psi(\zeta)$, обладают рядом свойств, благодаря которым они могут быть использованы не только в случае задач, хорошо решаемых другими распространенными методами, но и в случае, когда их применение связано с большими трудностями. Это, в частности, многоэкстремальные задачи, а также задачи, в которых $E(w_1, w_2, \dots, w_{N_p})$ не всюду дифференцируема, то есть как раз тот класс задач, к которому относится обучение сетей на базе гибридных нейроподобных элементов.

Преобразование функции $E(w_1, w_2, \dots, w_{N_p})$ в $\Psi(\zeta)$ заключается в определении меры некоторого множества D^* , на котором $E(w_1, w_2, \dots, w_{N_p})$ не превосходит заданное значение ζ (подразумевается минимизация функции $E(w_1, w_2, \dots, w_{N_p})$). Величина ζ изменяется в пределах $E_{\min} \leq \zeta \leq E_{\max}$. Метод Ψ -преобразования позволяет описать закон изменения указанной меры, когда ζ изменяется от минимального до максимального значения функции $E(w_1, w_2, \dots, w_{N_p})$. В процессе построения преобразованной функции $\Psi(\zeta)$ образуются также функции $\bar{w}_i(\zeta), (i=1, \dots, N_p)$, благодаря которым пред-

ставляется возможным определять не только значение глобального экстремума, но и координаты указанного экстремума. Величины $\bar{w}_i(\zeta)$ представляют собой средние значения аргументов, на которых функция $E(w_1, w_2, \dots, w_{N_p}) \leq \zeta$. Таким образом, при изменении ζ в вышеуказанных пределах можно построить функции $\Psi(\zeta), \bar{w}_1(\zeta), \bar{w}_2(\zeta), \dots, \bar{w}_{N_p}(\zeta)$.

Основные свойства функции $\Psi(\zeta)$:

1. Преобразованная функция $\Psi(\zeta)$ независимо от прообраза $E(w)$ является функцией одной переменной.

2. Преобразованная функция $\Psi(\zeta)$ для всех $E(w) \in L_p(D)$ является монотонно неубывающей функцией.

3. Значение ζ^* , при котором $\Psi(\zeta) = 0$, равно величине глобального экстремума функции $E(w) \in L_p(D)$ при $m\{w : E(w) = E_{\min}\} = 0$, где $m\{\cdot\}$ – мера множества.

Для проверки правильности определения координат глобального экстремума может быть применена следующая процедура:

- определяется ноль функции $\bar{\Psi}(\zeta)$;
- полученное значение ζ^* подставляется в функцию $\bar{w}(\zeta)$ и определяются координаты w^* ;
- w^* подставляется в функцию $E(w)$ и полученное значение E^* сравнивается с ζ^* ;
- если E^* отличается от ζ^* более чем на заранее заданную величину допустимой погрешности δ , то есть $|E^* - \zeta^*| > \delta$, то координаты w^* определены неверно.

Такое расхождение чаще всего свидетельствует о близости по величине глобального экстремума и одного из локальных. Ошибка в определении w^* в таком случае зависит от меры множества D_δ , которое образуется в области расположения глобального экстремума, в результате проведения гиперповерхности, параллельной w_1, w_2, \dots, w_{N_p} на уровне наименьшего из локальных экстремумов.

Если же функция $E(w)$ имеет несколько равных глобальных экстремумов, то определение координат w^* одного из них описанным методом становится невозможным. Именно такой случай возникает, если гиперповерхность критерия обучения сети

$E(w)$ обладает каким-либо видом симметрии: перестановочной и/или знаковой [8]. Для исключения подобной ситуации на область поиска должны накладываться ограничения, выделяющие единственную эквивалентную область из всех возможных. Поскольку определение координат глобального экстремума методом Ψ -преобразования носит статистический характер и число проводимых испытаний ограничено, данному методу присуща некоторая погрешность. Для ее сокращения целесообразно в области глобального экстремума дополнительно проводить локальный поиск с целью уточнения его координат. Обе эти возможности реализованы в предлагаемом ниже модифицированном методе Ψ -преобразования.

2. Модифицированный метод Ψ -преобразования для обучения сетей на базе гибридных нейроподобных элементов

Сеть на базе гибридных нейроподобных элементов реализует функцию

$$(\hat{y}_1(k), \hat{y}_2(k), \dots, \hat{y}_m(k)) = \Phi(w_1, w_2, \dots, w_{N_p}, x_1(k), x_2(k), \dots, x_n(k)) \quad (1)$$

или в векторной форме

$$\hat{y}(k) = \Phi(w, x(k)). \quad (2)$$

Ошибка обучения на шаге k определяется выражением

$$e(k) = y(k) - \hat{y}(k), \quad (3)$$

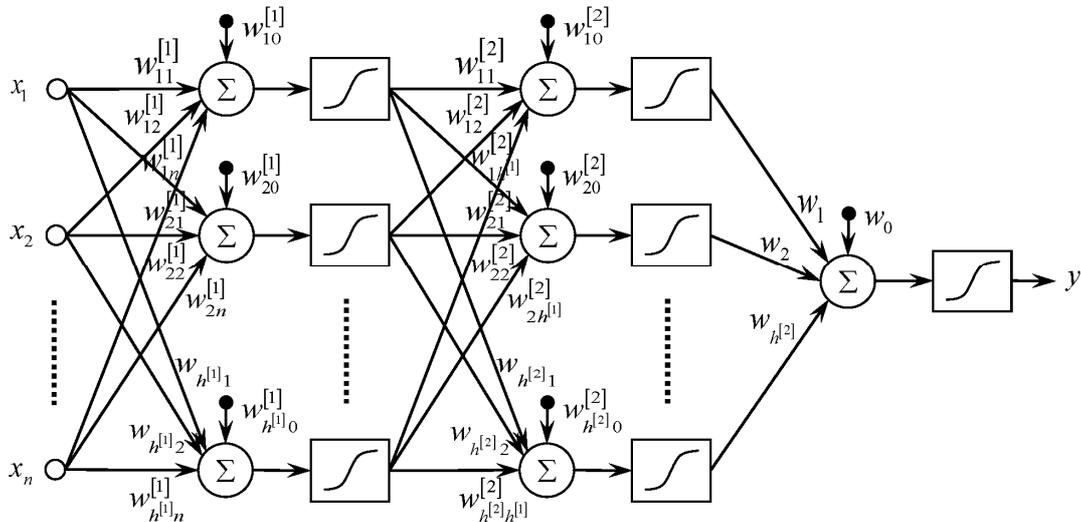


Рис. 1. Многослойный персептрон с двумя скрытыми слоями

где $y(k) = (y_1(k), y_2(k), \dots, y_m(k))$ – обучающий сигнал.

Стандартный квадратичный критерий обучения при этом может быть записан в следующем виде:

$$E = \sum_{k=1}^N \|e(k)\|^2, \quad (4)$$

где N – длина обучающей выборки.

При заданной обучающей выборке, то есть фиксированных значениях $x(k), y(k), (k = 1, \dots, N)$ критерий обучения E становится функцией только настраиваемых весов w . Обратим внимание, что для метода Ψ -преобразования конкретный вид критерия обучения не имеет значения, поэтому квадратичный критерий приведен здесь только в качестве примера.

Поскольку, как отмечалось выше, функция сети $\Phi(w, x(k))$ обладает свойством симметрии пространства параметров, то и критерий обучения $E(w)$ обладает тем же свойством. А как известно, метод Ψ -преобразования неэффективен в ситуациях, когда оптимизируемая функция имеет несколько локальных экстремумов, близких по значению к глобальному экстремуму. В данном случае ситуация еще сложнее, так как функция имеет множество равных глобальных экстремумов.

Для наглядности рассмотрим ограничения для выделения одной из эквивалентных областей в пространстве параметров на простом частном случае сети на базе гибридных нейроподобных элементов – многослойном персептроне с двумя скрытыми слоями (рис. 1).

Перестановочная симметрия проявляет себя здесь следующим образом: нейроны внутри каждого

скрытого слоя обладают одинаковыми наборами входных и выходных связей, а, следовательно – могут быть переставлены в произвольном порядке.

Отсюда мы имеем в общей сложности $(h^{[1]})! \times (h^{[2]})!$ эквивалентных областей перестановочной симметрии в пространстве параметров, а ограничения, выделяющие одну эквивалентную область в каждом скрытом слое, имеют вид:

$$w_{10}^{[s]} < w_{20}^{[s]} < \dots < w_{h^{[s]}0}^{[s]}, \quad (5)$$

где s – номер скрытого слоя.

Кроме того, если в нейронах скрытых слоев в качестве активационной применяется нечетная функция, например, часто используемая в теории искусственных нейронных сетей функция гиперболического тангенса $\psi(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$, то дополни-

тельно появляется так называемая знаковая симметрия. Она выражается в том, что значение выражения (1) не изменится, если одновременно изменить на противоположные знаки синаптических весов на входе и выходе любого нейрона любого скрытого слоя. Это свойство приводит к дополнительному появлению $h^{[1]2} \times h^{[2]2}$ эквивалентных областей внутри каждой эквивалентной области перестановочной симметрии.

Для выделения единственной эквивалентной области знаковой симметрии – наложение ограничений на знаки параметров смещения нейронов скрытых слоев сети:

$$w_{j0}^{[s]} \geq 0, \quad (6)$$

где j – номер нейрона внутри скрытого слоя, $j = 1, \dots, h^{[s]}$.

Таким образом, в многослойном персептроне с несколькими (в частном случае – с двумя) скрытыми слоями выделение одной эквивалентной области перестановочной и знаковой симметрий можно достичь совместным применением ограничений (5) и (6), причем оба вида ограничений касаются одних и тех же параметров сети – смещений нейронов скрытых слоев.

Теперь расширим полученные результаты на общий случай архитектуры сети на базе гибридных нейроподобных элементов (рис. 2).

Так как эта архитектура не имеет слоев нейронов (хотя они и могут образовываться в частных случаях), то ограничения вида (5) и (6) не могут в

ней применяться непосредственно. Более того, поскольку гибридные нейроподобные элементы могут содержать синапсы разных типов, и архитектура сети может быть неполносвязной и нерегулярной, нахождение эквивалентных областей и выделение единственной из них может оказаться сложной задачей. Для ее решения введем понятия скрытого элемента и группы эквивалентных элементов сети.

Скрытым элементом сети будем называть гибридный нейроподобный элемент, не имеющий связей ни с одним из выходов сети. Так, в многослойном персептроне все нейроны скрытых слоев сети являются скрытыми элементами. Каждый скрытый элемент создает знаковую симметрию в пространстве параметров при условии, что его активационная функция и функции всех синапсов, соединенных с его выходом, являются нечетными функциями. При выполнении этих условий выделение одной эквивалентной области достигается наложением ограничения вида (6) на параметр смещения этого скрытого элемента.

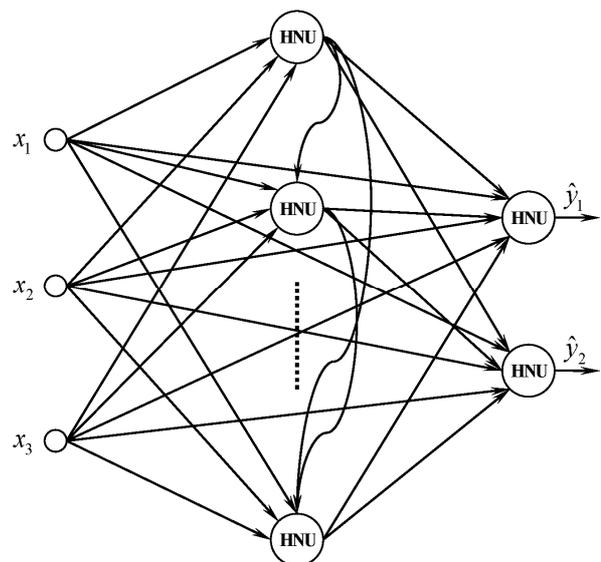


Рис. 2. Полносвязная архитектура сети на базе гибридных нейроподобных элементов.

Группа эквивалентных элементов сети – два или более скрытых элемента, имеющих одинаковый тип и одинаковые (по типам и структурным параметрам синапсов, а также направлению) наборы входных и выходных связей. В многослойном персептроне группу эквивалентных элементов составляет каждый скрытый слой сети. Каждая группа эквивалентных элементов создает перестановочную симметрию в пространстве параметров сети, поскольку произвольное изменение порядка нумерации скрытых элементов в группе не оказывает влияния на функцию, реализуемую сетью. При этом выделение одной эквивалентной области достигается наложением ограничения вида (5) на параметры

смещения скрытых элементов каждой группы.

Таким образом, перед применением метода Ψ -преобразования для параметрической оптимизации сети на базе гибридных нейроподобных элементов необходимо проанализировать ее архитектуру и определить все скрытые элементы сети, создающие знаковую симметрию, а также все группы эквивалентных элементов сети, создающие перестановочную симметрию. В соответствии с этими данными создаются системы ограничений вида (5) для всех групп эквивалентных элементов сети и вида (6) для всех скрытых элементов сети.

Каждое ограничение вида (6) сокращает гиперобъем области поиска в пространстве параметров вдвое, а каждое ограничение вида (5) дает сокращение еще в $h!$ раз, где h – количество скрытых элементов в группе. В итоге, суммарное сокращение гиперобъема области поиска может достигать для сложных архитектур нескольких десятков порядков.

3. Процесс обучения сети на базе гибридных нейроподобных элементов

Перед нами стоит задача минимизации критерия обучения $E(w)$ (4) по синаптическим весам w_i , ($i = 1, \dots, N_p$). Первым шагом для этого является определение границ множества D , на котором определена функция $E(w)$, то есть необходимо установить границы интервалов $[w_1^{\min}, w_1^{\max}]$, ..., $[w_{N_p}^{\min}, w_{N_p}^{\max}]$ для генерации случайных чисел. Затем генерируется S наборов, равномерно распределенных в вышеуказанных интервалах случайных чисел, образующие вектора синаптических весов сети w^v , ($v = 1, \dots, S$). Определяются ограничения вида (5) и (6), которые выделяют единственную эквивалентную область в пространстве параметров сети. Каждый вектор w^v проверяется на принадлежность к этой области и в случае нахождения за ее пределами w^v заменяется на его проекцию внутрь этой области \bar{w}^v , которая определяется путем изменения знака и перестановки элементов внутри вектора w^v в соответствии с указанными ограничениями.

После этого в каждой точке w^v вычисляется значение функции $E(w^v)$ и определяются нижняя E_{\min} и верхняя E_{\max} границы изменения функции $E(w)$ на D . Интервал $[E_{\min}, E_{\max}]$ разбивается на

T равных частей и вычисляются значения $\zeta_\tau = E_{\min} + \tau \frac{E_{\max} - E_{\min}}{T}$, ($\tau = 1, \dots, T$). Для каждого τ вычисляется количество случаев s_τ , где $E(w) \leq \zeta_\tau$, и определяется статистическая вероятность $P_\zeta^{\text{ст}} = \frac{s}{S}$, которая является оценкой $\bar{\Psi}(\zeta_\tau)$. По полученным точкам тем или иным способом строится аппроксимирующая кривая $\bar{\Psi}(\zeta)$ и путем ее экстраполяции находится значение $\hat{\zeta}^*$, при котором $\bar{\Psi}(\zeta) = 0$, являющееся оценкой глобального минимума функции $E(w)$.

Поскольку в задаче обучения сети на базе гибридных нейроподобных элементов важно не само значение экстремума, а его координаты, определим их на основе полученной оценки $\hat{\zeta}^*$. Для этого на основе данных, полученных в ходе статистических испытаний для построения функции $\bar{\Psi}(\zeta)$, строится функция $\bar{w}(\zeta)$, где $\bar{w}_i(\zeta)$ – среднее значение тех реализаций w_i , где $E(w_1, w_2, \dots, w_{N_p}) \leq \zeta$. Функции $\bar{w}_i(\zeta)$ также аппроксимируются и по ним путем подстановки $\hat{\zeta}^*$ вычисляется оценка координат глобального экстремума \hat{w}^* .

Для проверки достоверности полученной оценки значение \hat{w}^* подставляется в функцию $E(w)$ и значение $E(\hat{w}^*)$ сравнивается с оценкой $\hat{\zeta}^*$. В случае расхождения полученных оценок более, чем на заданную погрешность δ , необходимо уточнение функций $\bar{\Psi}(\zeta)$ и $\bar{w}(\zeta)$, для чего проводится еще S статистических испытаний на множестве D . После чего все оценки пересчитываются с учетом общего числа статистических испытаний $2S$.

Если же $|E(\hat{w}^*) - \hat{\zeta}^*| \leq \delta$, то оценка координат глобального экстремума выполнена верно, однако это еще не гарантирует равенство оценки \hat{w}^* и истинных координат w^* , поскольку все проделанные операции носят статистический характер и получаемые оценки сходятся по вероятности к истинным значениям только при $S \rightarrow \infty$, что требует слишком больших вычислительных затрат. Поэтому для уточнения координат целесообразно, используя в качестве стартовой точки полученную оценку, применить некоторый метод локального поиска. При этом полагается, что оценка \hat{w}^* находится в обла-

ти притяжения глобального экстремума и эта область выпукла, следовательно, где бы ни располагалась оценка \hat{w}^* относительно истинного значения w^* , это значение возможно отыскать методом локального поиска.

Поскольку функция $E(w)$ может быть недифференцируемой, то на выбор метода локального поиска накладывається соответствующее ограничение. Одним из методов, удовлетворяющих этому ограничению, а также имеющих возможность использовать некоторую информацию уже накопленную на этапе выполнения Ψ -преобразования, является комплексный метод Бокса [9]. В качестве исходных вершин комплекса могут быть использованы N_p точек w^v с наименьшими значениями функции $E(w^v)$ и оценка координат глобального экстремума \hat{w}^* .

Заключение

Высокая размерность оптимизационной задачи, сложность рельефа критерия обучения и разрывы его производных вызывают трудности в нахождении оптимального набора при обучении сети на базе гибридных нейроподобных элементов, что требует применения специализированных методов обучения. В качестве основы для построения такого метода был избран метод Ψ -преобразования, обладающий способностью находить глобальный экстремум функции и отсутствием требования дифференцируемости функции. Однако невозможность применения метода в случае наличия нескольких равных глобальных экстремумов и невысокая точность определения координат глобального экстремума вследствие статистического характера метода требует его модификации. Для этого были введены ограничения в виде неравенств на настраиваемые параметры сети, позволяющие выделить единственную эквивалентную область в пространстве параметров и существенно сократить гиперобъем области поиска и повысить качество поиска оптимальной архитектуры сети на базе гибридных нейроподобных элементов. Для уточнения координат глобаль-

ного экстремума предлагается использовать локальный поиск в области его притяжения.

Литература

1. Haykin, S. *Neural Networks. A Comprehensive Foundation [Text]* / Haykin S. – Upper Saddle River: Prentice Hall, 1999. – 842 p.
2. Попов, С.В. Гибридный нейроподобный элемент – новый тип строительного блока искусственных нейронных сетей [Текст] / С.В. Попов, К.А. Шкуро // *Научный вестник Донбасской государственной машиностроительной академии.* – 2011. – №2 (8Е). – С. 87-92.
3. Попов, С.В. Эволюционная нейро-фаззи сеть на базе гибридных нейроподобных элементов [Текст] / С.В. Попов, К.А. Шкуро // *17 міжнародна конференція з автоматичного управління «Автоматика-2010». Тези доповідей. Том 2.* – Харків: ХНУРЕ, 2010. – С. 193-194.
4. Popov, S. *Evolutionary Optimized Network of Hybrid Neuron-Like Units [Text]* / S. Popov, K. Shkuro // *Neural Networks and Artificial Intelligence (ICNNAI-2012) : proceedings of the 7th International Conference (Minsk, Belarus, 10-12 October, 2012).* – Minsk : BSUIR, 2012. – P. 32-35.
5. Fukumizu, K. *Local minima and plateaus in hierarchical structures of multilayer perceptrons [Text]* / K. Fukumizu, S. Amari // *Neural Networks.* – 2000. – Vol 13. – P. 317-327.
6. Shepherd, A.J. *Second-Order Methods for Neural Networks (Fast and Reliable Training Methods for Multi-Layer Perceptrons) [Text]* / A.J. Shepherd. – London: Springer, 1997. – 145 p.
7. Чичинадзе, В.К. *Решение невыпуклых нелинейных задач оптимизации [Текст]* / В.К. Чичинадзе. – М.: Наука, 1983. – 256 с.
8. Попов, С.В. *О симметрии пространства параметров многослойных вычислительных структур [Текст]* / С.В. Попов // *Тези доповідей учасників міжнародної науково-практичної конференції «Інтелектуальні системи прийняття рішень та інформаційні технології».* – Чернівці, 19-21 травня, 2004. – С. 59-60.
9. Химмельблау Д. *Прикладное нелинейное программирование [Текст]* / Д. Химмельблау. – М.: Мир, 1975. – 536 с.

Поступила в редакцію 12.05.2013, рассмотрена на редколлегии 29.05.2013

Рецензент: д-р техн. наук, проф., проф. каф. искусственного интеллекта Е.В. Бодянский, Харьковский национальный университет радиоэлектроники, Харьков.

МЕТОД ПАРАМЕТРИЧНОЇ ОПТИМІЗАЦІЇ МЕРЕЖІ НА БАЗІ ГІБРИДНИХ НЕЙРОПОДІБНИХ ЕЛЕМЕНТІВ, ЗАСНОВАНИЙ НА МЕТОДІ Ψ -ПЕРЕТВОРЕННЯ

С.В. Попов, К.О. Шкуро

У статті розглянуто особливості навчання мережі на базі гібридних нейроподібних елементів, що викликають труднощі в знаходженні оптимального набору ваг. Для пошуку глобального екстремуму критерію

навчання запропоновано модифікований метод, заснований на Ψ -перетворенні. Сформовано систему обмежень, що дозволяє виділити єдину серед множини еквівалентних областей, які виникають внаслідок перестановної і знакової симетрії простору параметрів мережі на базі гібридних нейроподібних елементів, що є необхідним для застосування методу Ψ -перетворення.

Ключові слова: параметрична оптимізація, навчання, гібридний нейроподібний елемент, метод Ψ -перетворення, симетрія простору параметрів, глобальний екстремум.

A Ψ -TRANSFORM BASED OPTIMIZATION METHOD FOR NETWORK OF HYBRID NEURON-LIKE UNITS

S.V. Popov, K.O. Shkuro

Special features of learning of the network of hybrid neuron-like units are considered that cause difficulty in finding an optimal set of weights. To find the global optimum of the learning criterion, a modified method based on Ψ -transform is proposed. A system of restrictions necessary for the application of the Ψ -transform method is proposed that selects one of equivalent areas in the parameters space of the network of hybrid neuron-like units, arising from permutation and sign symmetry.

Key words: parametric optimization, learning, hybrid neuron-like unit, Ψ -transform method, parameter space symmetry, global optimum.

Попов Сергей Витальевич – д-р техн. наук, ст. науч. сотр., главный научный сотрудник проблемной научно-исследовательской лаборатории автоматизированных систем управления, Харьковский национальный университет радиоэлектроники, Харьков, Украина.

Шкуро Кристина Александровна – аспирант каф. искусственного интеллекта, Харьковский национальный университет радиоэлектроники, Харьков, Украина, e-mail: shkurokristina@gmail.com.