

**АССОЦИАТИВНАЯ ПАМЯТЬ НА ОСНОВЕ ЯДЕРНЫХ СЕТЕЙ****1. Введение**

В последние годы область ядерных методов (kernel machines) бурно развивается. Эти методы получили широкое применение в области распознавания образов, восстановления зависимостей, классификации и кластеризации. Общая идея ядерных методов состоит в использовании вспомогательного пространства, размерность которого существенно превышает размерность исходного. Соответствие между старым и новым пространствами устанавливается некоторым нелинейным преобразованием

$$\varphi: E_X \rightarrow E'_X.$$

Это преобразование подбирается таким, что классы векторов, линейно не отделимые в  $E_X$ , будут в  $E'_X$  линейно отделимы. Размерность пространства  $E'_X$  может быть очень большой и даже бесконечной, поэтому непосредственно оперировать векторами в нем затруднительно. Во многих случаях достаточно знать лишь скалярное произведение в пространстве  $E'_X$ , выраженное через элементы  $E_X$ . Эта функция называется ядром:

$$K(u, v) = (\varphi(u), \varphi(v)). \quad (1)$$

Наиболее распространенным из ядерных методов является метод опорных векторов (Support Vector Machine, SVM) [1]. Этот метод требует решения задачи квадратичной оптимизации с выпуклыми ограничениями типа неравенств. Размерность этой задачи равна размерности обучающего множества. При больших обучающих выборках (особенно, если данные имеют высокую размерность) точное решение этой задачи становится затруднительным, поэтому разработаны многочисленные приближенные методы [1].

Также к числу ядерных методов относится и метод опорных векторов с использованием наименьших квадратов (LS-SVM); в качестве критерия оптимальности он использует не величину маржи (поля, margin), а среднеквадратичную ошибку. Этот метод требует лишь квадратичной оптимизации без ограничений и в итоге сводится к решению систем линейных уравнений. Оптимальность этого метода в соответствии со среднеквадратическим критерием риска доказана в [2].

Таким образом, на сегодняшний день большинство ядерных методов представляют собой алгоритмы классификации, распознавания образов и восстановления зависимостей (регрессии), основанные на обучении с учителем.

В данной работе предпринята попытка расширить применимость ядерных методов и построить на их основе систему ассоциативной памяти. Для функционирования ассоциативной памяти типа Хопфилда ключевой является линейная независимость запоминаемых образов. В такой сети можно запомнить набор векторов, только если они линейно независимы. Более того, эта независимость должна быть "достаточно сильной", то есть каждый вектор должен лежать достаточно далеко от линейной оболочки всех остальных. Это означает, что количество запоминаемых образов не может превышать размерность запоминаемых векторов, а на практике должно составлять не более 25% (псевдоинверсный алгоритм обучения [Personnaz]) или 70% (разнасыщение [3]). По мере приближения количества запомненных образов к размерности пространства в синаптической матрице начинают преобладать диагональные элементы. Это приводит к резкому ослаблению аттракторных свойств и ухудшению качества функционирования такой системы как ассоциативной памяти.

Применение ядерного подхода позволяет перейти к пространству, в котором запоминаемое множество становится линейно независимым.

В качестве прототипа мы используем псевдоинверсную гетеро- и автоассоциативную память. Применение ядер в рамках данной парадигмы позволяет снять некоторые ограничения, связанные с линейностью исходной модели. В частности, удается ослабить ограничение на емкость памяти, которая в таких сетях не может превосходить 25 – 75% от размерности данных.

На базе этого подхода удалось также построить автоассоциативную память, способную к конвергенции в процессе экзамена.

## 2. Алгоритм

Рассмотрим гетероассоциативную память с псевдоинверсным алгоритмом обучения. Пусть дано пространство входов  $E_X$  и пространство выходов  $E_Y$  размерностей  $n$  и  $p$  соответственно. Нам необходимо запомнить  $m$  пар векторов:  $x_i \in E_X, y_i \in E_Y, i = 1 \dots m$ . Будем считать, что эти векторы образуют столбцы матриц  $X$  и  $Y$  соответственно.

Тогда матрица связей такой сети удовлетворяет соотношению

$$BX = Y$$

и строится следующим образом:

$$B = YX^+. \quad (2)$$

То есть, эта матрица определяет проекционный оператор  $B: E_X \rightarrow E_Y$ , который для всех  $i$  удовлетворяет соотношению  $Bx_i = y_i$ . В формуле (2)  $X^+$  означает псевдообратную матрицу по Муру – Пенроузу. Она дается соотношением

$$X^+ = (X^T X)^{-1} X^T = S^{-1} X^T. \quad (3)$$

Здесь элементы матрицы  $S$  размером  $m \times m$  вычисляются как попарные скалярные произведения запоминаемых векторов:

$$s_{ij} = (x_i, x_j). \quad (4)$$

При экзамене на вход системы подается произвольный вектор  $x$ . По нему требуется определить реакцию сети  $y$ . Это можно сделать следующим образом:

$$\begin{aligned} y &= Bx = YS^{-1}z; \\ z &= X^T x; \\ z_i &= (x_i, x). \end{aligned} \quad (5)$$

Можно заметить, что для вычисления матрицы  $S$ , вектора реакции сети  $y$  и промежуточного вектора  $z$  необходимо знать лишь скалярные произведения запомненных входных векторов между собой и с вектором  $x$ .

Воспользуемся этой особенностью гетероассоциативной памяти для построения ядерного алгоритма. Заменяем пространство  $E_X$  на  $E'_X$  размерности  $n' \gg n$  ( $E'_X$  может быть также и бесконечномерным гильбертовым пространством). Векторы в  $E'_X$  выражаются через исходные посредством нелинейного преобразования  $\Phi: E_X \rightarrow E'_X$ . Будем называть  $E'_X$  пространством признаков (feature space).

Пусть теперь  $x_i' = \varphi(x_i)$ ,  $x_i \in E_X$  – входные векторы из обучающей последовательности. Обозначим  $K(u, v) = (\varphi(u), \varphi(v))$  – ядро.

Тогда, аналогично (3) – (5), получаем

$$\begin{aligned} s_{ij} &= K(x_i, x_j); \\ z_i &= K(x_i, x). \end{aligned} \tag{6}$$

Выражения (2 – 3), (5) можно вычислить явно, не переходя в большое пространство  $E'_X$ , а проводить процедуры формирования памяти и экзамена, используя лишь ядро.

Такой алгоритм является базовым для построения различных схем ассоциативной памяти. Недостаток его заключается в том, что такая сеть осуществляет линейное проектирование из пространства признаков в пространство выходов, она не способна к конвергенции в процессе экзамена.

Применимость ядерного алгоритма описывает

**Следствие 1 из теоремы Мерсера:** Если ядро  $K(u, v)$  удовлетворяет условию Мерсера [1], то

- 1) гильбертово пространство  $E'_X$  и отображение  $\varphi: E_X \rightarrow E'_X$  такое, что  $K(u, v) = (\varphi(u), \varphi(v))$  существуют;
- 2) для каждого набора  $x_i \in E_X$ ,  $y_i \in E_Y$ ,  $i = 1 \dots m$  матрица  $S$  неотрицательно определена;
- 3) если к тому же  $\dim E'_X > m$ , существует проекционный оператор  $B: E'_X \rightarrow E_Y$ , удовлетворяющий соотношению  $Bx'_i = y_i$ .

**Доказательство:**

- 1) напрямую следует из теоремы Мерсера;
- 2) истинно потому, что матрица  $S$  составлена из попарных скалярных произведений набора векторов  $x'_i \in E'_X$  (матрица Грамма);
- 3) такой оператор всегда можно построить на линейной оболочке векторов  $\langle x'_i \rangle_{i=1}^m \in E'_X$  –  $m$ -мерном подпространстве. Затем его можно непрерывно продолжить на все  $E'_X$ . □

Условие Мерсера формулируется следующим образом: Пусть функция  $K(\mathbf{u}, \mathbf{v}): Q \times Q \rightarrow \Re$  непрерывна и симметрична, а  $Q$  – некоторый компакт в  $E_X$ . Для того, чтобы существовало пространство  $E'_X$ , в котором  $K(.,.)$  определяет скалярное произведение, необходимо и достаточно, чтобы для любого  $g \in L_2(Q)$  выполнялось неравенство

$$\iint_{u, v \in Q} K(u, v) g(u) g(v) \geq 0.$$

К сожалению, мы не можем гарантировать обратимость матрицы  $S$ . Обратимость достигается тогда и только тогда, когда все  $\langle x'_i \rangle_{i=1}^m \in E'_X$  линейно независимы. Это условие может не соблюдаться при некоторых типах ядра на конкретных наборах векторов. На практике для устранения этой проблемы используют регуляризацию по Тихонову: вместо  $S$  применяется матрица

$$S_\mu = S + \mu I$$

при малом  $\mu > 0$ . Эта матрица всегда обратима, так как  $S$  неотрицательно определена.

Другим подходом к решению этой проблемы является инкрементальное построение матрицы  $S$ . На каждом шаге размерность ее повышают на единицу, добавляя очередной запоминаемый вектор. Если получившаяся матрица оказалась необратимой, вектор не запоминается. Для обращения новой  $S$  применяется техника работы с блочными матрицами [4, 5].

Попробуем оценить емкость памяти: для запоминания  $m$  образов в такой сети необходимо хранить матрицу  $S$  размером  $m \times m$ . То есть, можно сказать, что ядерная ассоциативная память обладает способностью запоминать количество образов, равное числу нейронов. Эта оценка является максимальной и в некоторых случаях может не достигаться на практике. Так, например, при выборе линейного ядра, описываемый механизм отождествляется с обычной гетероассоциативной памятью.

### 3. Модификации ядерного алгоритма

#### 3.1. Автоассоциативная память

Описанный алгоритм можно применять и в качестве автоассоциативной памяти. При этом  $x_i \in E_X, y_i \in E_Y, x_i = y_i, i = 1 \dots m$ . Также необходимо, чтобы эти векторы были биполярными. Для этого потребуем совпадения пространств  $E_X$  и  $E_Y$ , а также самих входов и выходов. Матрица  $S$  по-прежнему формируется по формуле (6). Экзамен производится по итеративной процедуре: на вход подается вектор  $x_i$ , по нему по формулам (5 – 6) вычисляется постсинаптический потенциал  $y_i$ . Затем применяется функция активации, и вычисляется состояние системы на следующем шаге:

$$x_{i+1} = f(y_i). \quad (7)$$

Процедура производится до конвергенции к устойчивому состоянию – аттрактору. Аттракторы таких систем описывает теорема 1.

**Теорема 1.** Если для системы автоассоциативной памяти (4 – 8) выполняются условия следствия 1, а также матрица  $S$  обратима, аттракторами процедуры конвергенции такой сети являются лишь неподвижные точки и двойные циклы.

Доказательство: аналогично соответствующей теореме для сетей типа Хопфилда, используя функцию энергии

$$E_t = -\frac{1}{2} K(x_t, y_t). \quad (8)$$

По следствию из теоремы Мерсера, существует самосопряженный проекционный оператор  $C: E'_x \rightarrow E'_x$ , такой, что  $Cx'_i = x'_i$ . Используя свойства скалярного произведения в пространстве  $E'_x$ , можно написать следующую цепочку равенств:

$$\begin{aligned} E_t - E_{t+1} &= -\frac{1}{2} (x'_t, Cx'_{t-1}) + \frac{1}{2} (x'_{t+1}, Cx'_t) = -\frac{1}{2} (x'_{t-1}, Cx'_t) + \frac{1}{2} (x'_{t+1}, Cx'_t) = \\ &= \frac{1}{2} K(y_t, x_{t+1}) - \frac{1}{2} K(y_t, x_{t-1}). \end{aligned} \quad (9)$$

Для (большинства) ядер эта величина всегда неотрицательна, поскольку  $x_{t-1}$  является ближайшим биполярным вектором к  $y_t$ . Это гарантирует неотрицательность (9) для всех ядер типа RBF, а также ядер вида  $K(x, y) = R(\langle x, y \rangle)$  (здесь  $\langle x, y \rangle$  – скалярное произведение).

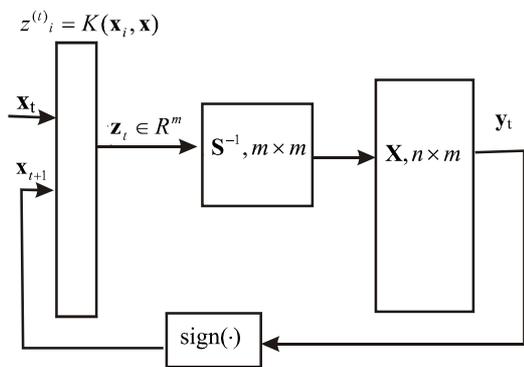


Рис. 1. Схема алгоритма автоассоциативной памяти

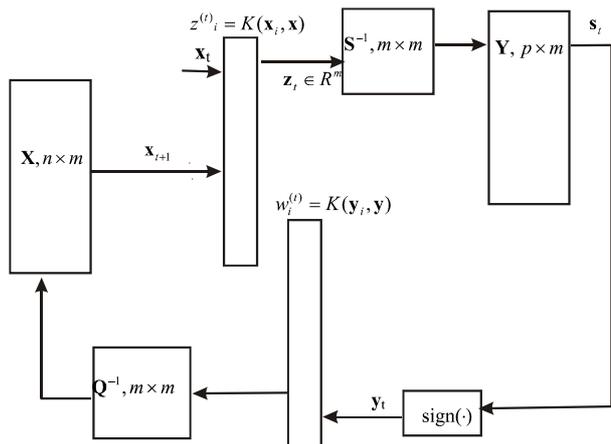


Рис. 2. Схема алгоритма гетероассоциативной памяти

предлагаемых алгоритмов были проведены эксперименты на некоторых тестовых задачах.

#### 4.1. Задача "Двойная спираль"

Это задача классификации двумерных векторов на два класса. Из-за сложной формы разделяющих кривых она трудна для многих типов нейронных сетей [7].

Данные для этого примера показаны на рис. 3. Выборка из 384 точек, принадлежащих к двум классам (поровну распределенных между ними), разбивалась случайным образом на примерно равные обучающую и тестовую части. При экспериментах с ядерными системами было использовано ядро типа радиальных базисных функций (radial basis function, RBF) в виде

$$K(u, v) = \exp\left(-\frac{1}{2\sigma^2} \|u - v\|^2\right) \quad (10)$$

с  $\sigma = 0.12$ . Эксперимент был проведен также с некоторыми известными нейропарадигмами: каскадной feed-forward, и каскадной ассоциативной памятью [8]. Результаты тестирования приведены в табл. 1.

Схема алгоритма конвергенции (экзамена) для ядерной автоассоциативной памяти приведена на рис. 1.

Подобную схему можно построить и для гетероассоциативной памяти. При этом входы либо выходы системы будут биполярными, а для замыкания процедуры экзамена строится обратное преобразование из  $E_Y$  в  $E_X$  (рис. 2).

Возможно также использование ядерных методов в модульной ассоциативной памяти. Это можно сделать двумя способами. В первом случае строится двухслойная сеть, первый слой которой представляет собой ядерную гетероассоциативную память с  $m \gg p$ , на втором слое функционирует модульная автоассоциативная память с древесной структурой, как в [6]. Второй способ предполагает организацию системы модулей, каждый из которых – ядерная автоассоциативная память с организацией процедуры экзамена по формулам (7 – 8).

#### 4. Экспериментальные результаты

Для проверки практической применимости

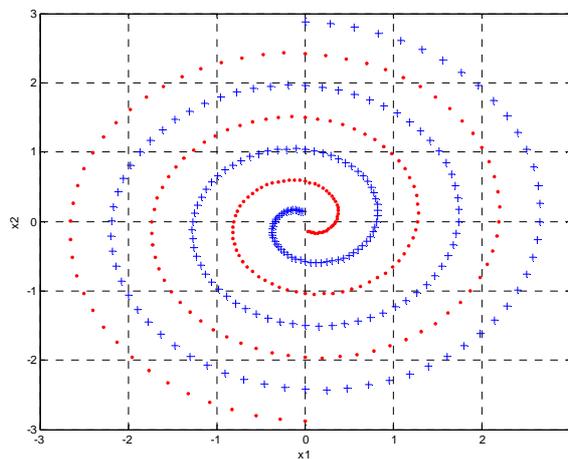


Рис. 3. Задача "двойная спираль"

Таблица 1. Качество распознавания в задаче "двойная спираль"

Сеть	Число слоев	Число нейронов в сети	Время обучения, с	Качество распознавания при независимом тесте, %
Cascade Backpropagation	20	231	35	94,3
Cascade associative	55	1596	1,1	92,7
Kernel associative	1	192	0,46	94,2

#### 4.2. Запоминание биполярных векторов в автоассоциативной памяти

В данном эксперименте изучалась сеть автоассоциативной памяти по алгоритму (4 – 8). В качестве данных использовались биполярные векторы различной размерности (от 16 до 256) и в различном количестве (от 32 до 1024), но всегда  $m > n$ . Ядро выбиралось в виде (11) с различным  $\sigma$ . Сеть сохраняет способность к конвергенции с аттракторным радиусом больше 1 вплоть до  $m = 20n$ .

Для изучения аттракторных свойств была выбрана сеть, запоминающая 512 образов размерностью 32. Множество (биполярных) векторов было сгенерировано таким образом, что расстояние по Хеммингу между любой их парой было не менее 10. Использовалось ядро (\*1) с  $\sigma = 1$ . При этом средний аттракторный радиус составил 6,85, его среднеквадратическое отклонение 0,93. Среднее число итераций конвергенции при этом составляло 2,3. Для  $\sigma = 2$  средний аттракторный радиус равнялся 4,33, его СКО 0,56. Система делала в среднем 2,97 итераций конвергенции. Для 256 запоминаемых образов и  $\sigma = 2$  аттракторный радиус составил в среднем 6,25, а число итераций возросло до 3,55.

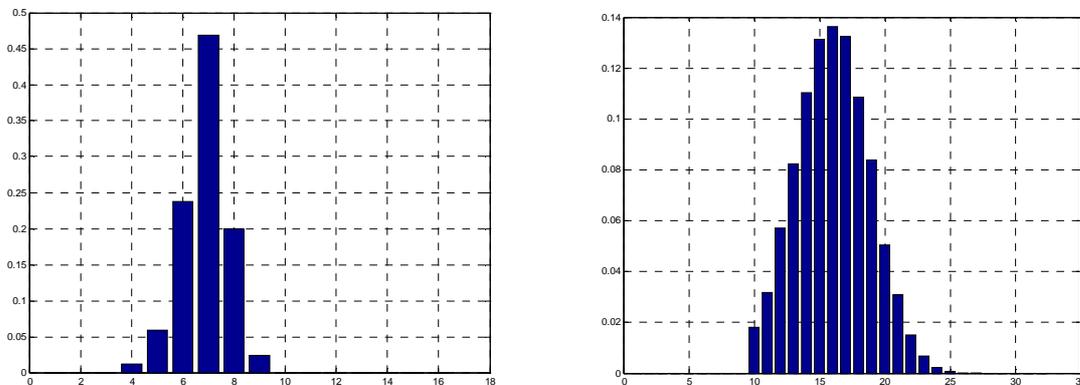


Рис. 4. Распределение аттракторного радиуса (слева) и расстояния между векторами (справа) для ядерной АП

Распределение векторов по расстоянию Хемминга показано на рис. 4 (справа). Там же приведена и гистограмма аттракторного радиуса (слева). Гистограмма АР строилась на основе экзамена всех запомненных векторов. При этом входной вектор отличался от эталона на фиксированную величину шума.

#### 5. Обсуждение экспериментальных результатов

Проведенная серия экспериментов показывает, что предложенный алгоритм успешно функционирует в качестве как авто-, так и гетероассоциативной памяти. Пример с двойной спиралью показывает возможность его работы в качестве классификатора: достигнутое качество распознавания сравнимо со средними показателями классификаторов, решающих эту задачу.

Особенно интересна ядерная автоассоциативная память с процедурой конвергенции в процессе экзамена. В процессе экспериментов были исследованы некоторые свойства таких сетей с ядром типа RBF. При этом параметр  $\sigma$  ядра играет роль показателя степени взаимодействия между образами в памяти. При небольших  $\sigma$  при экзамене определяющим является образ, ближайший к входному. В этом случае число итераций конвергенции невелико, аттракторный радиус максимален для памяти с большим числом образов ( $m/n > 10$ ). Матрица  $S$  близка к единичной, и в целом такая система похожа на сеть Хемминга: так же, как и сеть Хемминга, она осуществляет отбор среди эталонов по максимуму соответствия входному образу.

При  $\sigma = 1,5 \dots 2$  аттракторный радиус незначительно падает, число итераций растет, что говорит об усложнении поведения ассоциативной памяти. Матрица  $S$  уже существенно отличается от единицы. При меньшем заполнении памяти такая конфигурация характеризуется сохранением аттракторного радиуса при дальнейшем увеличении числа итераций.

Поведение системы сильно зависит от характера ядра. Для ядер типа RBF  $K(u, v) = R\left(\frac{\|u - v\|^2}{\rho}\right)$ ,

где  $\rho$  – некоторая характерная длина (для гауссова ядра ее роль играет  $\sigma$ ). Предельным случаем такого ядра является дельта-функция  $K(u, v) = \delta(u - v)$ . Если использовать ядро в виде дельта-функции, матрица  $S$  становится единичной, и система перестает реагировать на входы, которые сколько-нибудь отличаются от эталонов. В этом случае вектор  $z$  из (6) будет равен нулю. В общем случае поведение сети зависит от соотношения между параметром ядра  $\rho$  и средним расстоянием между запоминаемыми входными векторами  $r$ . Если  $r \gg \rho$ , при формировании вектора  $z$  доминирует ближайший сосед входа  $x$ , матрица  $S$  все еще близка к единичной.

Наоборот, при  $r \ll \rho$  элементы матрицы  $S$  можно разложить в ряд, где  $r_{ij} = \|x_i - x_j\|$  (будем считать ядро аналитическим):

$$s_{ij} = 1 - a_2 \left(\frac{r_{ij}}{\rho}\right)^2 + O\left(\left(\frac{r_{ij}}{\rho}\right)^4\right). \quad (11)$$

При малых  $r$  старшими членами ряда (2) можно пренебречь. В таком случае при  $m > n$  матрица  $S$  будет плохо обусловленной и качество ассоциативной памяти неудовлетворительно. Зависимость числа обусловленности матрицы  $S$  от  $\sigma$  в случае гауссова ядра показана на рис. 5.

Если  $\sigma$  велико, в спектре матрицы  $S$  доминирует одно большое собственное значение. Это говорит о том, что в пространстве  $E^1$  преобразованные векторы вытянуты преимущественно вдоль одного направления. Это затрудняет работу обратного преобразования и в конечном счете делает невозможным функционирование ассоциативной памяти с такими параметрами.

Наиболее интересные результаты достигаются, когда  $r$  и  $\rho$  одного порядка. В этом случае матрица  $S$  имеет нетривиальный вид, процедура экзамена занимает несколько итераций, и ассоциативная память демонстрирует хорошие аттракторные свойства.

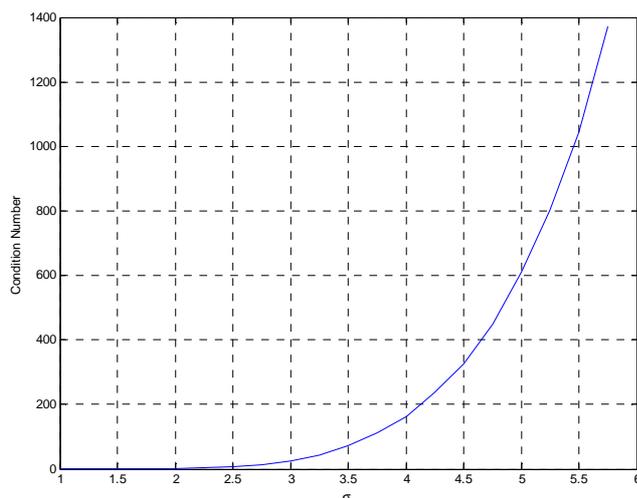


Рис. 5. Число обусловленности матрицы  $S$  в зависимости от параметра  $\sigma$

## 6. Заключение

В данной работе предложена модель ассоциативной памяти на основе ядерного подхода. Описаны несколько алгоритмов гетеро- и автоассоциативной памяти, дающие возможность конвергенции в процессе экзамена.

В зависимости от выбора ядра такие системы ведут себя по-разному. Так, например, выбор квадратичного ядра (в виде скалярного произведения) приводит к реализации сети Хопфилда с проекционным алгоритмом обучения. Такая схема может оказаться предпочтительнее традиционной, так как требует не более  $n(n+m)$  чисел для запоминания  $m$   $n$ -мерных векторов по сравнению с  $n^2/2$  для хранения проекционной матрицы.

Тем не менее предлагаемый метод требует хранения симметрической матрицы размером  $m \times m$  ( $m$  – число запоминаемых образов). Следует также отметить, что эффективность работы всех алгоритмов, основанных на применении ядерного подхода, сильно зависит от выбора функции ядра. Подбор ядра проводится по эвристическим соображениям для каждой конкретной задачи.

В случае неквадратичного ядра появляется возможность запомнить количество векторов, существенно превышающее их размерность. Однако, если ядро становится слишком "заостренным" (сильная нелинейность), свойства такой системы приближаются к сети Хемминга.

Наиболее интересны промежуточные случаи. Здесь число итераций конвергенции может быть довольно большим, сильно проявляется взаимодействие между образами. Можно предположить, что такая система способна восстанавливать образ даже при утрате одного из эталонов. Вопросы устойчивости ядерной ассоциативной памяти к такого рода искажениям, а также свойства ее аттракторных областей, составляют предмет дальнейшего исследования.

## СПИСОК ЛИТЕРАТУРЫ

1. Vapnik V. Statistical Learning Theory. – John Wiley & Sons, NY, 1998. – P. 736.
2. Smale S. On the Mathematical Foundations of Learning // Bull. Am. Math. Soc. – 2001. – Vol. 39, N 1. – P. 1 – 49.
3. Резник А.М., Городничий Д.О., Сычев А.С. Регулирование обратной связи в нейронных сетях с проекционным алгоритмом обучения // Кибернетика и системный анализ. – 1996. – № 6. – С. 153 – 163.
4. Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание: Пер. с англ. – М.: Наука, 1977. – 237 с.
5. Пайпс Л. А. Матрицы в технике: Пер. с англ. // Современная математика для инженеров / Под ред. Э.Ф. Беккенбаха. – М.: Изд. ин. лит., 1959.
6. Reznik A.M. and Dekhtyarenko O.K., Modular Neural Associative Memory Capable of Storage of Large Amounts of Data // Proc. of the International Joint Conference on Neural Networks. – Portland, Oregon. – 2003. – July 20 – 24. – P. 3031 – 3035.
7. Reed R.D, Marks R.J. II Neural Smothing: supervised learning in feedforward artificial neural networks // MIT Press. – 1998.
8. Многофункциональный нейрокомпьютер Neuroland / Резник А.М., Калина Е.А., Сычев А.С., Садовая Е.Г, Дехтяренко А.К., Галинская А.А. // Математические машины и системы. – 2003. – № 1. – С. 36 – 45.