А.В. Савченко

# Статистическое распознавание образов на основе вероятностной нейронной сети с проверкой однородности[1]

**Аннотация.** Статистическое распознавание образов сведено к проверке гипотез об однородности выборок. Для ее оптимального решения в смысле минимума среднего байесовского риска предложена модификация вероятностной нейронной сети (PNN). Представлены результаты сравнительного анализа предложенной модификации с оригинальной PNN в задаче автоматической идентификации авторства литературного текста.

**Ключевые слова:** статистическое распознавание образов, вероятностная нейронная сеть, проверка гипотез об однородности выборок, принцип минимума информационного рассогласования Кульбака-Лейблера.

## Введение

Во многих областях прикладных исследований, таких как задачи искусственного интеллекта (распознавание изображений, речи), автоматизированная поддержка принятия решений, техническая и медицинская диагностика, экономический и лингвистический анализ и др., где вся доступная информация заключена в конечных выборках наблюдений [1], возникает необходимость по возможности точно ответить на вопрос: насколько существенно отличаются свойства анализируемых объектов между собой? Эта задача может быть сформулирована в терминах распознавания образов [1, 2] - по заданной *выборке* требуется определить, к какому классу она принадлежит. При этом вся доступная информация о классах содержится только в обучающих выборках, принадлежность которых к тому или иному классу заранее известна. В рамках универсального статистического подхода указанная задача обычно сводится к задаче статистической классификации [3, 4]. Ее оптимальное решение дает классический байесовский подход минимизации среднего риска [2, 3]. А наиболее распространенным на данный момент способом реализации этого подхода служит вероятностная нейронная сеть (probabilistic neural network, PNN) [5, 6], в которой для аппроксимации неизвестных плотностей вероятностей по обучающим выборкам используются ядерные функции Парзена [5]. Эта сеть первоначально была предназначена для классификации единичных объектов, однако PNN может применяться и для классификации выборки при предположении о независимости ее элементов (т.н. наивный байесовский классификатор [2, 3]). К сожалению, этот подход позволяет получить известное [1, 7] оптимальное решение задачи распознавания образов, основанное на ее редукции к проверке статистической однородности выборок, лишь в том случае, если объем обучающих выборок намного превосходит объем распознаваемой выборки. Поэтому в настоящей работе оптимальный алгоритм распознавания образов реализован на основе новой модификации [4] PNN, в которой указанное ограничение отсутствует. Показано, что в случае гистограммной оценки распределения (т.е. при стремлении параметра разброса гауссовского ядра Парзена к 0) критерий, лежащий в основе

[1] Исследование осуществлено в рамках Программы «Научный фонд НИУ ВШЭ» в 2013-2014 гг., проект № 12-01-0003.

предложенной PNN, эквивалентен известному выражению для проверки однородности двух выборок по принципу минимума информационного рассогласования Кульбака-Лейблера [8]. В рамках экспериментального исследования рассмотрено применение предложенной модификации в задаче идентификации автора литературного текста [9]. Полученные результаты и сделанные по ним выводы рассчитаны на широкий круг специалистов в области статистического распознавания образов.

## 1. Задача статистического распознавания образов

Наиболее общая формулировка большинства задач статистического распознавания образов состоит в следующем [4]. Требуется отнести предъявляемый объект анализа $\mathbf{X} = \{\mathbf{x}_j\}, j = \overline{1,n}$ – *выборку* объема *n* к одному из $R > 1$ классов, строго говоря, заранее точно не определенных. Здесь элемент выборки $\mathbf{x}_j = \{x_{j;1},...,x_{j;M}\}$ представляет собой некоторый вектор *признаков* фиксированной длины *M*. Каждый класс характеризуется тем, что принадлежащие ему объекты обладают некоей общностью или сходством. То общее, что объединяет объекты в класс, и называют *образом* [10]. В таком случае решение рассматриваемой задачи сводится к установлению отношения эквивалентности между соответствующим набором признаков объекта наблюдения $\mathbf{X}$ и одного из *R* образов в базе данных.

Проблема состоит в том [11], что каждому конкретному образу обычно присуща известная вариативность, т.е. изменчивость его признаков от одного образца наблюдения к другому, которая носит случайный характер. Обычно преодоление данной проблемы связывают со статистическим подходом [2, 3], когда в роли каждого образа выступает соответствующий закон распределения $\mathbf{P}_r$ объектов одного класса. Задача переходит в таком случае в задачу проверки статистических гипотез о неизвестном законе распределения. Здесь возникает новое препятствие - проблема встречных гипотез в отношении вида каждого из $R$ альтернативных распределений. Радикальное средство для ее преодоления [11] – это восстановление неизвестных законов $\mathbf{P}_r$, $r = \overline{1,R}$ в процессе предварительного обучения по выборкам $\mathbf{X}_r = \{\mathbf{x}_j^{(r)}\}, j = \overline{1,n_r}$ конечных объемов $n_r$, $r = \overline{1,R}$, принадлежность которых классу (образу) *r* заранее точно известна.

Предположим, что $\mathbf{X}$ представляет собой выборку независимых наблюдений, распределенных по закону $\mathbf{P} \subset \{\mathbf{P}_r\}$ из множества заданных альтернатив, какому именно – неизвестно. Тогда задача статистического распознавания объекта $\mathbf{X}$ может быть сведена к проверке *R* гипотез о его законе распределения:

$$W_r: \quad \mathbf{P} = \mathbf{P}_r, \quad r = \overline{1,R}. \tag{1}$$

Для ее решения обычно используется принцип минимума среднего байесовского риска [2, 3] - делается вывод о принадлежности входного объекта к классу $\nu$ с максимальной апостериорной вероятностью

$$\nu = \arg\max_{r \in \{1,...,R\}} P(\mathbf{X}|W_r) \cdot P(W_r) \tag{2}$$

Здесь $P(W_r)$ - априорная вероятность появления *r*-го класса, $P(\mathbf{X}|W_r)$ - правдоподобие, т.е. условная вероятность принадлежности объекта $\mathbf{X}$ классу *r*. В большинстве задач распознавания образов предполагается [12], что появление каждого класса равновероятно (полная априорная неопределённость)

$$P(W_r) = \frac{1}{R}.$$

А для оценки правдоподобия воспользуемся заданными обучающими выборками [13]

$$\hat{P}\left(\mathbf{X}|W_r\right)=\frac{1}{\left(n_r\right)^n}\prod_{j=1}^{n}\sum_{j_r=1}^{n_r}K\left(\mathbf{x}_j,\mathbf{x}_{j_r}^{(r)}\right) \tag{3}$$

Здесь $K\left(\mathbf{x}_j,\mathbf{x}_{j_r}^{(r)}\right)$ - так называемая ядерная (kernel)-функция [6]. Наиболее распространено гауссовское ядро Парзена

$$K\left(\mathbf{x}_j,\mathbf{x}_{j_r}^{(r)}\right)=\frac{1}{\left(2\pi\sigma^2\right)^{M/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{M}\left(x_{j;i}-x_{j_r;i}\right)^2\right), \tag{4}$$

где $\sigma=const>0$ - фиксированный параметр разброса. С учетом оценки (3) итоговое решение (2) может быть записано в виде

$$\nu=\arg\max_{r\in\{1,\ldots,R\}}\frac{1}{\left(n_r\right)^n}\prod_{j=1}^{n}\sum_{j_r=1}^{n_r}K\left(\mathbf{x}_j,\mathbf{x}_{j_r}^{(r)}\right) \tag{5}$$

Как известно, критерий (5) может быть реализован в виде вероятностной нейронной сети (PNN) [5] для задачи распознавания образов (Рис.1). В отличие от классической PNN [6], используемой для классификации, здесь к четырем слоям (входной слой, слой образов, слой суммирования и выходной слой) добавляется слой произведения, т.к. классифицируется не один объект $\mathbf{x}_j$, а некоторая выборка $\mathbf{X}$.

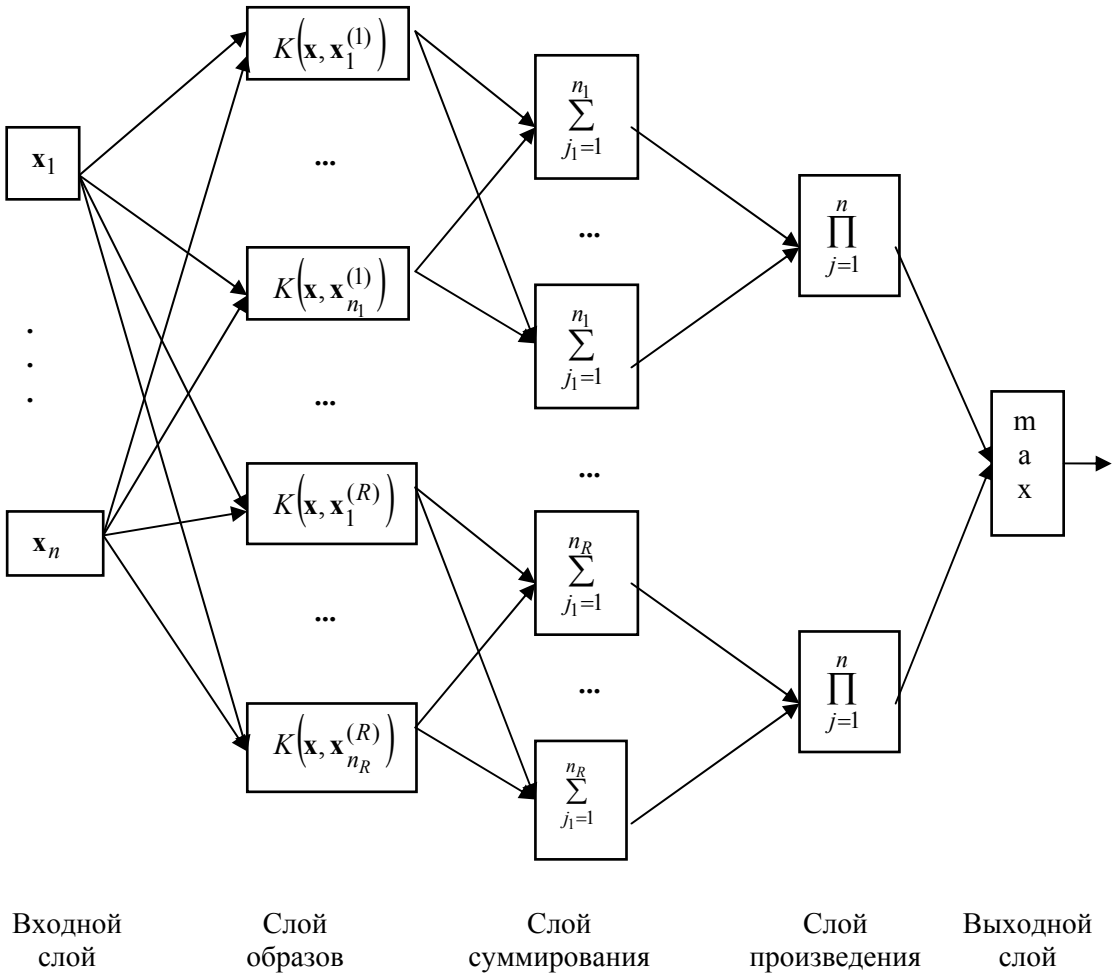



Рис.1. Вероятностная нейронная сеть в задаче распознавания образов (критерий (5)).

## 2.Синтез оптимального алгоритма

В поставленной задаче распознавания образов закон распределения $\mathbf{P}_r$ неизвестен, поэтому для решения требуется выполнить его оценку (3) по заданным обучающим выборкам. В этом состоит отмеченное в [1] существенное отличие от задачи статистической классификации (1), в которой распределения $\mathbf{P}_r$ предполагаются известными. Поэтому более правильным нам представляется использование подхода к распознаванию образов из [1], когда задача сводится к проверке $R$ гипотез о статистической однородности входной $\mathbf{X}$ и обучающей выборок $\mathbf{X}_r$.

$$W_r : \left\{\mathbf{x}_j\right\}, j=\overline{1,n} \text{ и } \left\{\mathbf{x}_j^{(r)}\right\}, j=\overline{1,n_r} \text{ имеют одинаковое распределение}$$

*Теорема 1.* Оптимальное решение в смысле минимума среднего риска задачи статистического распознавания выборки независимых одинаково распределенных объектов (признаков) дает следующий критерий

$$\nu = \arg\max_{r\in\{1,\ldots,R\}} \frac{n^n\cdot(n_r)^{n_r}}{(n+n_r)^{n+n_r}} \prod_{j=1}^{n}\left(1+\frac{\sum\limits_{j_r=1}^{n_r} K\left(\mathbf{x}_j,\mathbf{x}_{j_r}^{(r)}\right)}{\sum\limits_{j_1=1}^{n} K\left(\mathbf{x}_j,\mathbf{x}_{j_1}\right)}\right)\cdot\prod_{j_r=1}^{n_r}\left(1+\frac{\sum\limits_{j_1=1}^{n} K\left(\mathbf{x}_{j_r}^{(r)},\mathbf{x}_{j_1}\right)}{\sum\limits_{j_{r;1}=1}^{n_r} K\left(\mathbf{x}_{j_r}^{(r)},\mathbf{x}_{j_{r;1}}^{(r)}\right)}\right) \quad (6)$$

*Доказательство*. Оптимальное решение принимается на основе принципа минимума среднего риска по выборке из объединенного выборочного пространства $\left\{\mathbf{X},\mathbf{X}_1,\ldots,\mathbf{X}_R\right\}$ в пользу класса $\nu$, такого, что

$$\nu = \arg\max_{r\in\{1,\ldots,R\}} \sup_{\mathbf{P}^*}\sup_{\mathbf{P}_j^*,\,j=\overline{1,R}} P\left(\left\{\mathbf{X},\mathbf{X}_1,\ldots,\mathbf{X}_R\right\}\middle|W_r\right)\cdot P\left(W_r\right) \quad (7)$$

Здесь $\mathbf{P}^*$ - распределение (неизвестное) выборки $\mathbf{X}$, $\mathbf{P}_j^*$ - неизвестное распределение $r$-го эталона. Учитывая независимость элементов выборок $\left\{\mathbf{X},\mathbf{X}_1,\ldots,\mathbf{X}_R\right\}$, функцию правдоподобия можно переписать в виде

$$\sup_{\mathbf{P}^*}\sup_{\mathbf{P}_j^*,\,j=\overline{1,R}} P\left(\left\{\mathbf{X},\mathbf{X}_1,\ldots,\mathbf{X}_R\right\}\middle|W_r\right)=\sup_{\mathbf{P}^*}P\left(\mathbf{X}\middle|W_r\right)\sup_{\mathbf{P}_r^*}P\left(\mathbf{X}_r\middle|W_r\right)\cdot\prod_{\substack{j=1\\ j\neq r}}^{R}\sup_{\mathbf{P}_j^*}P\left(\mathbf{X}_j\right),$$

или

$$\sup_{\mathbf{P}^*}\sup_{\mathbf{P}_j^*,\,j=\overline{1,R}} P\left(\left\{\mathbf{X},\mathbf{X}_1,\ldots,\mathbf{X}_R\right\}\middle|W_r\right)=\frac{\sup\limits_{\mathbf{P}^*}P\left(\mathbf{X}\middle|W_r\right)\sup\limits_{\mathbf{P}_r^*}P\left(\mathbf{X}_r\middle|W_r\right)}{\sup\limits_{\mathbf{P}_r^*}P\left(\mathbf{X}_r\right)}\cdot\prod_{j=1}^{R}\sup_{\mathbf{P}_j^*}P\left(\mathbf{X}_j\right).$$

Так как $\prod\limits_{j=1}^{R}\sup\limits_{\mathbf{P}_j^*}P\left(\mathbf{X}_j\right)=const(\mathbf{X})$ не зависит от классифицируемой выборки $\mathbf{X}$, (7) эквивалентно критерию

$$\nu = \arg\max_{r\in\{1,\ldots,R\}} \frac{\sup\limits_{\mathbf{P}^*} P(\mathbf{X}|W_r)\,\sup\limits_{\mathbf{P}_r^*} P(\mathbf{X}_r|W_r)}{\sup\limits_{\mathbf{P}_r^*} P(\mathbf{X}_r)} \cdot P(W_r)$$

Введя для симметрии в знаменатель множитель $\sup\limits_{\mathbf{P}^*} P(\mathbf{X})$, не зависящий от обучающей выборки $\mathbf{X}_r$, окончательно запишем

$$\nu = \arg\max_{r\in\{1,\ldots,R\}} \frac{\sup\limits_{\mathbf{P}^*} P(\mathbf{X}|W_r)\,\sup\limits_{\mathbf{P}_r^*} P(\mathbf{X}_r|W_r)}{\sup\limits_{\mathbf{P}^*} P(\mathbf{X})\cdot\sup\limits_{\mathbf{P}_r^*} P(\mathbf{X}_r)} \cdot P(W_r) \qquad (8)$$

Известно [1], что верхняя граница (sup) функции правдоподобия достигается при совпадение истинных распределений $\mathbf{P}^*$, $\mathbf{P}_r^*$ с их оптимальными несмещенными оценками. При этом для вычисления вероятностей при условии справедливости гипотезы $W_r$ используется совместная выборка $\{\mathbf{X}, \mathbf{X}_r\}$. Тогда для верхних границ вероятностей из (8) можно аналогично (3) записать следующие ядерные оценки

$$\sup_{\mathbf{P}^*} P(\mathbf{X}|W_r) = \frac{1}{(n+n_r)^n}\prod_{j=1}^{n}\left(\sum_{j_1=1}^{n} K\left(\mathbf{x}_j, \mathbf{x}_{j_1}\right) + \sum_{j_r=1}^{n_r} K\left(\mathbf{x}_j, \mathbf{x}_{j_r}^{(r)}\right)\right),$$

$$\sup_{\mathbf{P}_r^*} P(\mathbf{X}_r|W_r) = \frac{1}{(n+n_r)^{n_r}}\prod_{j_r=1}^{n_r}\left(\sum_{j_1=1}^{n} K\left(\mathbf{x}_{j_r}^{(r)}, \mathbf{x}_{j_1}\right) + \sum_{j_{r;1}=1}^{n_r} K\left(\mathbf{x}_{j_r}^{(r)}, \mathbf{x}_{j_{r;1}}^{(r)}\right)\right),$$

$$\sup_{\mathbf{P}_r^*} P(\mathbf{X}_r) = \frac{1}{(n_r)^{n_r}}\prod_{j_r=1}^{n_r}\sum_{j_{r;1}=1}^{n_r} K\left(\mathbf{x}_{j_r}^{(r)}, \mathbf{x}_{j_{r;1}}^{(r)}\right),$$

$$\sup_{\mathbf{P}^*} P(\mathbf{X}) = \frac{1}{n^n}\prod_{j=1}^{n}\sum_{j_1=1}^{n} K\left(\mathbf{x}_j, \mathbf{x}_{j_1}\right).$$

С учетом предположения о полной априорной неопределенности, (8) переходит в

$$\begin{aligned}\nu = \arg\max_{r\in\{1,\ldots,R\}} &\frac{1}{(n+n_r)^{n+n_r}}\prod_{j=1}^{n}\left(\sum_{j_1=1}^{n} K\left(\mathbf{x}_j, \mathbf{x}_{j_1}\right) + \sum_{j_r=1}^{n_r} K\left(\mathbf{x}_j, \mathbf{x}_{j_r}^{(r)}\right)\right)\times \\ &\times\prod_{j_r=1}^{n_r}\left(\sum_{j_1=1}^{n} K\left(\mathbf{x}_{j_r}^{(r)}, \mathbf{x}_{j_1}\right) + \sum_{j_{r;1}=1}^{n_r} K\left(\mathbf{x}_{j_r}^{(r)}, \mathbf{x}_{j_{r;1}}^{(r)}\right)\right)\times \\ &\times\left[\frac{1}{n^n\cdot(n_r)^{n_r}}\prod_{j=1}^{n}\sum_{j_1=1}^{n} K\left(\mathbf{x}_j, \mathbf{x}_{j_1}\right)\cdot\prod_{j_r=1}^{n_r}\sum_{j_{r;1}=1}^{n_r} K\left(\mathbf{x}_{j_r}^{(r)}, \mathbf{x}_{j_{r;1}}^{(r)}\right)\right]^{-1}\end{aligned} \qquad (9)$$

После несложных преобразований, окончательно получаем критерий (6). Теорема 1 доказана.

Параллельная (нейросетевая [5]) реализация выражения (6) и представляет собой предлагаемую модификацию PNN для задачи распознавания образов. Ее структурная схема [4] представлена на Рис.2.

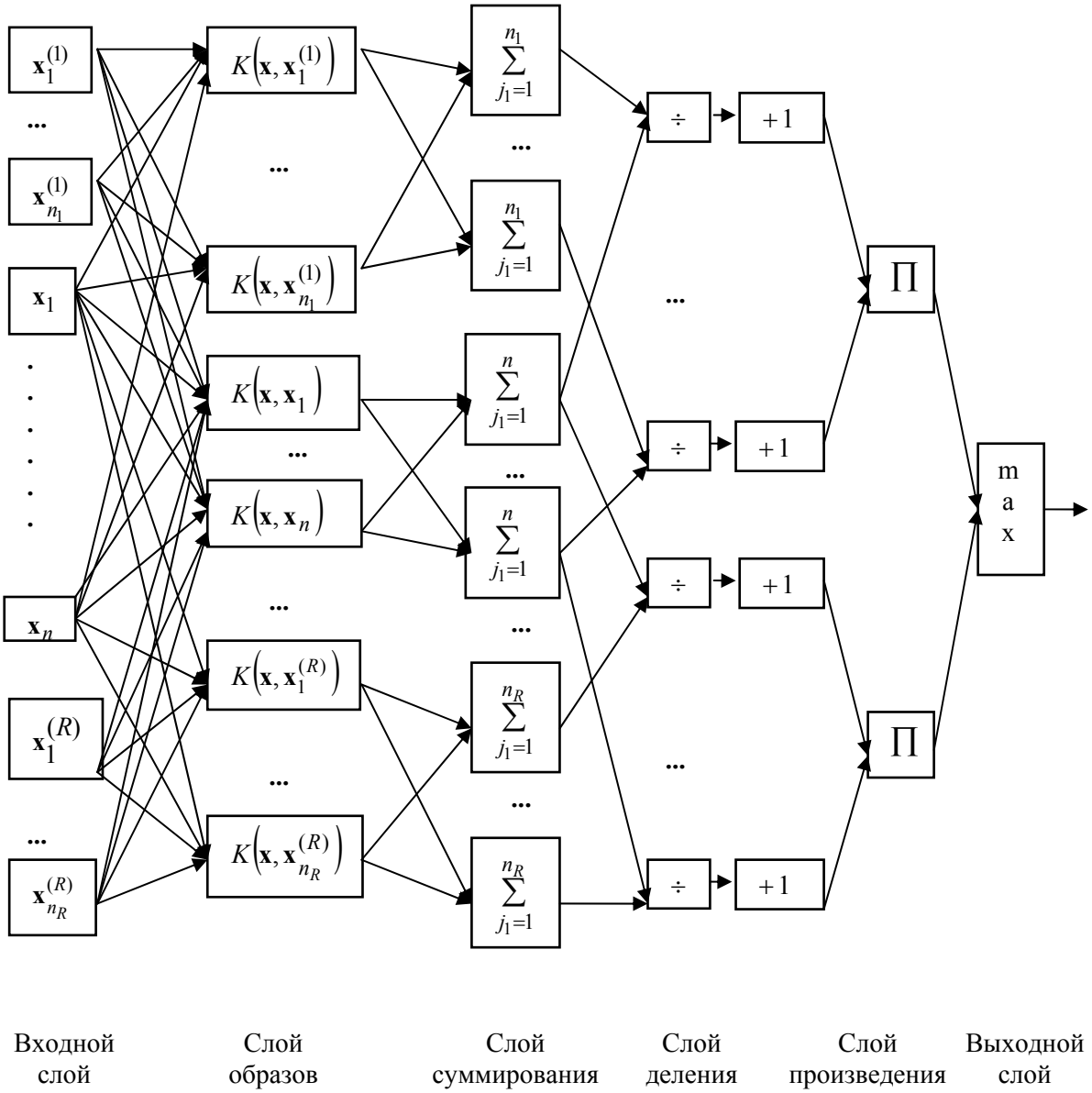



Рис.2. Предложенная модификация вероятностной нейронной сети в задаче статистического распознавания образов (критерий (6))

Здесь на вход сети поступает объединенная выборка $\{\mathbf{X}, \mathbf{X}_1,..., \mathbf{X}_R\}$ (а не только входная выборка $\mathbf{X}$, как на Рис.1), при этом не делается различие между входной и обучающими выборками. Во втором слое образов в дополнение к обучающим выборкам $\{\mathbf{X}_r\}$ присутствует ядерная функция от элементов входной выборки. Также по сравнению с рис. 1 в сеть (Рис. 2) добавлен новый слой деления, а в следующем за ним слое произведения в качестве множителей выступают не только элементы выборки признаков входного объекта, но и элементы *r*-й обучающей выборки.

Заметим, что, если объем обучающих выборок велик, то в пределе при $n_r \to \infty$ выражение (6) переходит в (5). Действительно, в асимптотике обучающая выборка $\mathbf{X}_r$ полностью задает закон распределения $\mathbf{P}_r$, поэтому использование объединенной выборки $\{\mathbf{X}, \mathbf{X}_r\}$ не несет никакой дополнительной информации.

Предложенная модифицированная PNN (Рис. 2) [4] сохраняет все достоинства классической PNN (Рис. 1). При этом скорость сходимости к оптимальному байесовскому решению критерия (6) должна быть выше скорости сходимости критерия (5). Однако среднее время классификации в два раза превышает среднее время классификации PNN (Рис. 1). Поэтому далее приведем упрощенную версию критерия (6), которую можно использовать для распознавания дискретных объектов [11] при наличии ограничений на максимальное время классификации (распознавание в режиме реального времени) [14].

## 3. Распознавание дискретных случайных объектов

Предположим, что признаки $\mathbf{x}_j$ могут быть сгруппированы [1] в непересекающиеся области $\Delta_1,...,\Delta_N$, где $N$ - количество областей. Воспользуемся единичной ядерной функцией, являющейся пределом (4) при $\sigma \to 0$ в случае дискретных объектов

$$I\left(\mathbf{x}_j, \mathbf{x}_{j_r}^{(r)}\right) = \begin{cases} \dfrac{1}{\left|\Delta\left(\mathbf{x}_j\right)\right|}, & \Delta\left(\mathbf{x}_j\right) = \Delta\left(\mathbf{x}_{j_r}^{(r)}\right) \\ 0, & \Delta\left(\mathbf{x}_j\right) \neq \Delta\left(\mathbf{x}_{j_r}^{(r)}\right) \end{cases} \tag{10}$$

Здесь $\left|\Delta\right|$ - мощность интервала $\Delta$, а $\Delta\left(\mathbf{x}_j\right)$ - интервал, к которому принадлежит $\mathbf{x}_j$. В этом случае распределение $\mathbf{P}_r$ задается как вектор вероятностей $\mathbf{P}_r = \left\{p_{r,i} = P\left(\mathbf{x} \in \Delta_i\right)\right\}, i = \overline{1,N}$ принадлежности объекта к области $\Delta_i$. Далее для упрощения предположим, что все области $\Delta_1,...,\Delta_N$ имеют равную мощность $\left|\Delta_1\right| = ... = \left|\Delta_N\right|$.

Используя (3) и (10), вычислим оценку распределения обучающей выборки $\hat{\mathbf{P}}_r = \left\{\theta_{r,i}\right\}, i = \overline{1,N}$

$$\theta_{r,i} = \frac{1}{n_r} \sum_{j=1}^{n_r} I_{\Delta_i}\left(\mathbf{x}_j^{(r)}\right), \tag{11}$$

где

$$I_{\Delta_i}\left(\mathbf{x}_j^{(r)}\right) = \begin{cases} 1, & \Delta\left(\mathbf{x}_j^{(r)}\right) = \Delta_i \\ 0, & \Delta\left(\mathbf{x}_j^{(r)}\right) \neq \Delta_i \end{cases}$$

Аналогично оценим распределение входной выборки $\hat{\mathbf{P}} = \left\{w_i\right\}, i = \overline{1,N}$

$$w_i = \frac{1}{n} \sum_{j=1}^{n} I_{\Delta_i}\left(\mathbf{x}_j\right) \tag{12}$$

*Теорема 2*. Оптимальное решение (6) задачи распознавания выборки дискретных объектов для ядерной функции (10) эквивалентно известному критерию проверки статистической однородности двух выборок на основе принципа минимума информационного рассогласования Кульбака-Лейблера [8]

$$\nu = \arg\min_{r \in \{1,...,R\}} \left( n \sum_{i=1}^{N} w_i \log \frac{w_i \cdot \left(n + n_r\right)}{n \cdot w_i + n_r \cdot \theta_{r;i}} + n_r \sum_{i=1}^{N} \theta_{r,i} \log \frac{\theta_{r,i} \cdot \left(n + n_r\right)}{n \cdot w_i + n_r \cdot \theta_{r;i}} \right) \tag{13}$$

Здесь рассогласование Кульбака-Лейблера вычисляется между объединенной выборкой $\{\mathbf{X}, \mathbf{X}_r\}$ и наилучшей несмещенной оценкой $\hat{\mathbf{P}}_r^* = \{\hat{p}_{r,i}^*\}, i = \overline{1, N}$ распределения $\mathbf{P}_r^*$ при условии справедливости гипотезы $W_r$, где

$$\hat{p}_{r,i}^* = \frac{n}{n+n_r} \cdot w_i + \frac{n_r}{n+n_r} \cdot \theta_{r;i}$$

*Доказательство.* С учетом (11), (12) для выбранной ядерной функции (10) выражение (6) (или, что то же самое, (9)) будет эквивалентно критерию

$$\nu = \arg\max_{r\in\{1,\ldots,R\}} \frac{\prod_{i=1}^{N}\left(\frac{n\cdot w_i + n_r\cdot\theta_{r;i}}{n+n_r}\right)^{n\cdot w_i} \prod_{i=1}^{N}\left(\frac{n\cdot w_i + n_r\cdot\theta_{r;i}}{n+n_r}\right)^{n_r\cdot\theta_{r,i}}}{\prod_{i=1}^{N}(w_i)^{n\cdot w_i}\cdot\prod_{i=1}^{N}(\theta_{r,i})^{n_r\cdot\theta_{r,i}}},$$

или

$$\nu = \arg\min_{r\in\{1,\ldots,R\}} \prod_{i=1}^{N}\left(\frac{w_i\cdot(n+n_r)}{n\cdot w_i + n_r\cdot\theta_{r;i}}\right)^{n\cdot w_i} \cdot \prod_{i=1}^{N}\left(\frac{\theta_{r,i}\cdot(n+n_r)}{n\cdot w_i + n_r\cdot\theta_{r;i}}\right)^{n_r\cdot\theta_{r,i}}$$

Выполняя логарифмирование последнего выражения, окончательно получаем выражение (13). Теорема 2 доказана.

Аналогично для ядерной функции (10) можно показать, что традиционное решение (5) эквивалентно критерию минимума информационного рассогласования Кульбака-Лейблера [8] между оценками распределения входного объекта $\hat{\mathbf{P}}$ и $r$-й обучающей выборки $\hat{\mathbf{P}}_r$

$$\nu = \arg\min_{r\in\{1,\ldots,R\}} \sum_{i=1}^{N} w_i \log\frac{w_i}{\theta_{r;i}} \tag{14}$$

Вычислительная сложность алгоритма распознавания образов на основе критериев (13), (14) за счет редукции обучающей выборки к ее гистограмме $\hat{\mathbf{P}}_r$ (11) в $\frac{n\cdot n_r}{N}$ раз ниже сложности алгоритмов (6) и (5) соответственно. Однако критерии (6), (5) являются намного более общими по сравнению с (13), (14), так как в них не накладывается ограничение (10) на используемую ядерную функцию.

## 4. Результаты экспериментальных исследований

Проиллюстрируем предложенную модификацию (6) вероятностной нейронной сети и ее упрощенную версию (13) на следующем примере распознавания образов из практики лингвистического анализа [9, 11]. Ставится задача автоматического распознавания автора фрагмента текста по заданным фрагментам произведений нескольких авторов (эталонам). В процессе проведенных исследований были рассмотрены восемь текстов: "Анна Каренина" и "Воскресение" Л.Н. Толстого, "Идиот" и "Преступление и наказание" Ф.М. Достоевского, "Мертвые души" и "Тарас Бульба" Н.В. Гоголя, "Тихий Дон" и "Поднятая целина" М.А. Шолохова. Тексты взяты из электронной библиотеки [15].

В эксперименте проводится сравнительный анализ вероятности ошибки распознавания традиционного подхода (5), (14), реализуемого в PNN (Рис. 1), и предложенной модификации (6), (13). Точность оценивалась следующим образом. Из каждого текста в обучающую выборку добавля-

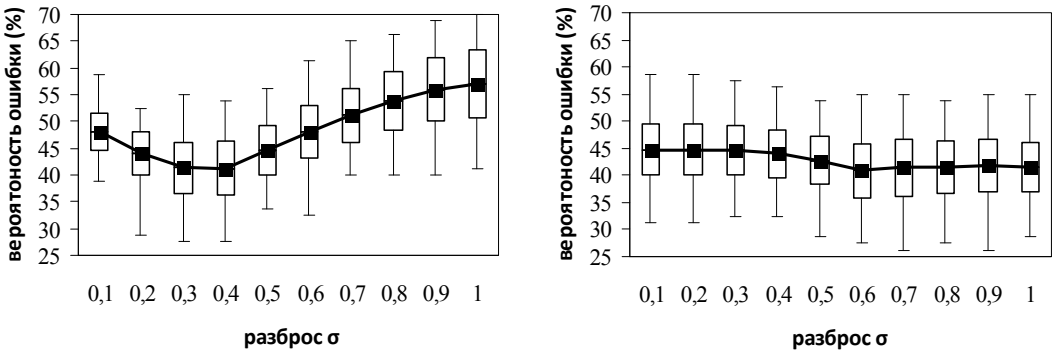
лись по одному выбранному наугад фрагменту фиксированного размера. Для каждого текста также наугад выбиралось 10 тестовых выборок, которые и подавались на вход алгоритмов распознавания (5) и (6). По 80 (8*10) экспериментам оценивалась средняя вероятность ошибки. Качество распознавания оценивалось как среднее арифметическое вероятности ошибки по 100 таким экспериментам (всего 8000 операций классификации).

В качестве показателя стилистических особенностей текстов, весьма условного, но, вместе с тем, информативного и наглядного, в первом эксперименте была выбрана частота знаков препинания для каждого отдельного предложения. Как известно [10], эти признаки показали достаточно высокое качество идентификации автора неизвестного текста для русского языка. Использовались следующие знаки препинания: ".", "?", "!", ",", ".", "-", ":", ";", "(". Таким образом, вектор признаков **x** содержит *M=9* элементов. Объем выборки *n* определяется числом предложений в выбранном фрагменте.

В первом случае и обучающие, и тестовые выборки признаков определялись по фрагментам текстов из 25000 символов. Результаты распознавания - диаграмма "ящик с усами" для зависимости вероятности ошибки классификации от параметра разброса $\sigma$ для критериев (5) и (6) показаны на Рис.3.

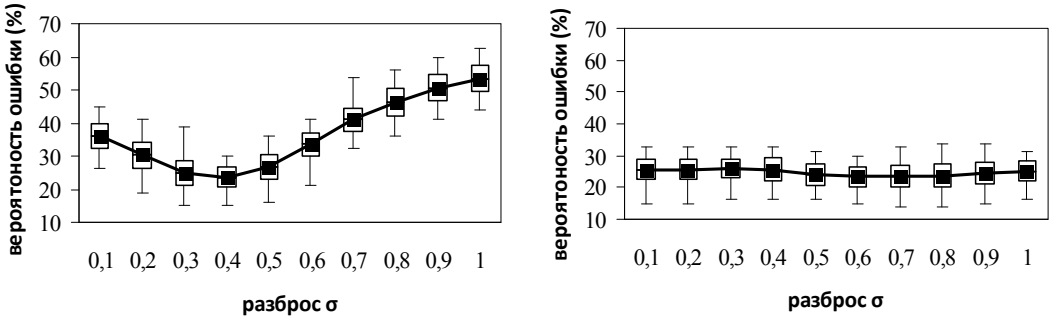
Как видно из этого рисунка, минимальная вероятность ошибки 40,8% для предложенного выражения (6) несколько ниже аналогичного показателя - 41,3% - для традиционного подхода (5). При этом основной показатель качества предложенной модификации - это устойчивость к изменению параметра разброса $\sigma$ - даже в худшем случае вероятность ошибки не превышает 45%. В отличие от этого, PNN значительно больше зависит от выбора оптимального значения разброса - в худшем случае вероятность ошибки приближается к 60%.

Во втором случае первый эксперимент был повторен, но для построения обучающих выборок использовались фрагменты текста объемом 100000 символов. Тестовая выборка формировалась из фрагмента объемом 25000 символов. На Рис.4 представлены результаты этого эксперимента.




а) Критерий (5) б) Критерий (6)

Рис. 3. Зависимость вероятности ошибки от разброса σ; частота знаков пунктуации; обучающая выборка - 25000 символов

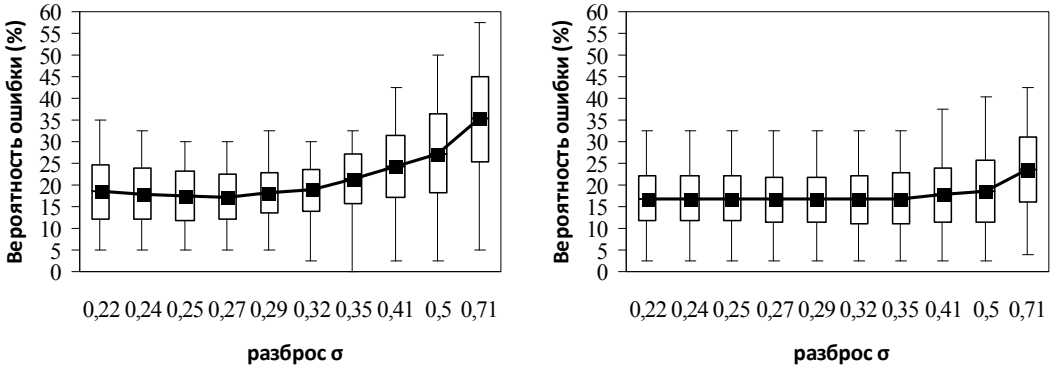



а) Критерий (5) б) Критерий (6

Рис. 4. Зависимость вероятности ошибки от разброса σ; частота знаков пунктуации; обучающая выборка - 100000 символов

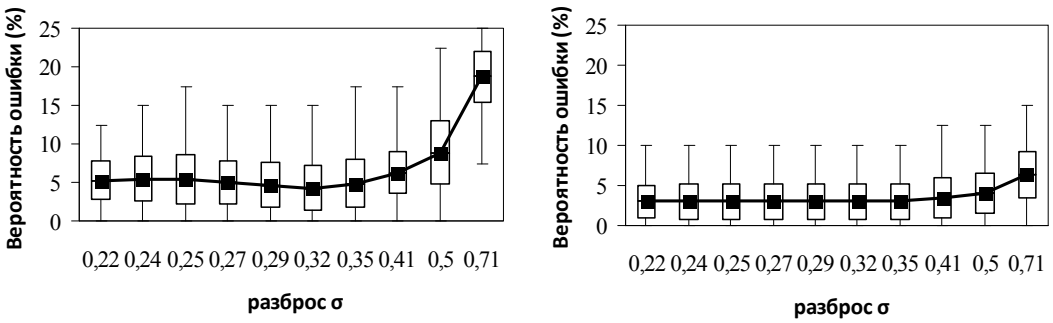
В этом случае точность классификации значительно превышает точность предыдущего эксперимента. Минимальная вероятность ошибки 23% для предложенной модификации (6) несколько ниже аналогичного показателя (23,6%) традиционного критерия (5). И снова предложенная сеть, в отличие от PNN, оказывается устойчивой к изменению разброса $\sigma$.

Во втором эксперименте применялись наиболее традиционные для задачи распознавания авторства текста признаки - частота символьных триграмм [9] (последовательностей из трех символов, широко использующийся на практике частный случай N-грамм [16]). Для вычисления ядерной функции (4) вместо расстояния Евклида использовалось количество одинаковых символов в сравниваемых триграммах. Результаты идентификации в виде зависимости вероятности ошибки от разброса $\sigma$ для критериев (5) и (6) показаны на Рис.5 (25000 символов в обучающей выборке) и Рис. 6 (100000 символов в обучающей выборке).




а) Критерий (5) б) Критерий (6)

Рис. 5. Зависимость вероятности ошибки от разброса $\sigma$; частота триграмм; обучающая выборка - 25000 символов




а) Критерий (5) б) Критерий (6)

Рис. 6. Зависимость вероятности ошибки от разброса σ; частота триграмм; обучающая выборка - 100000 символов

Как видно, в этом случае качество классификации значительно выше, чем для первого эксперимента. Однако в целом выводы остаются теми же: предложенный подход позволяет получить лучшую точность и более устойчив к изменению параметра ядерной функции.

В заключительном эксперименте проведено сравнение упрощенной версии предложенной модификации (13) с выражением (14). В качестве признака использовалась частота появления *N=500* наиболее распространенных слов [17], которые выбирались из числа слов, входящих в выбранные эталонные фрагменты. При этом выбирались только слова в нижнем регистре (не имена собственные) длиной не менее трех символов. В качестве предварительной обработки проводился стемминг (stemming), т.е. для каждого слова путем отбрасывания окончания выде-

лялась его основа [16]. Использовался известный алгоритм [17] - из слов удалялись следующие окончания: -ому, -ему, -ами, -ыми, -ями, -ими, -ашь, -ешь, -ишь, -ого, -его, -ый, -ой, -ий, -ые, -ие, -ое, -ее, -ия, -ая, -ию, -ии, -ую, -ах, -ых, -их, -ть, -ов, -ев, -им, -ом, -ым, -ам, -ям, -ем, -ут, -ют, -ит, -ет, -ат, -ят, -а, -я, -о, -е, -ы, -и, -ь, -ю, -у. Здесь уже точность классификации 86,7% для предложенного критерия (13) существенно превосходит точность 72,2% критерия (14). Все результаты (усредненная вероятность ошибки совместно с диапазоном ее изменения) вместе с наилучшими результатами из предыдущих экспериментов (признак частоты символов пунктуации в предложении) сведены в таблицу.

Оценки вероятности ошибки для идентификации автора фрагмента текста

| **Признаки** | **Частота символов пунктуации** | | **Частота триграмм** | | **Частота слов** | |
|---|---|---|---|---|---|---|
| **Количество символов** | Критерий (5) | Критерий (6) | Критерий (5) | Критерий (6) | Критерий (14) | Критерий (13) |
| 25000 - обучение, 25000 - тестирование | 41,3%±7,5% | 40,8%±7,4% | 17,4%±7,7% | 15,9%±7,3% | 27,8%±8,2% | 13,3%±6,5% |
| 100000 - обучение, 25000 - тестирование | 23,6%±4,5% | 23,0%±4,4% | 4,2%±4,3% | 3,0%±3,4% | 20,6%±5,4% | 3,8%±2,9% |

По результатам проведенного эксперимента можно сделать следующие выводы. Во-первых, точность классификации для традиционной вероятностной нейронной сети (5) несколько ниже аналогичного показателя для предложенной модификации (6). Особенно это заметно при использовании их реализаций (13), (14) на основе информационного рассогласования Кульбака-Лейблера при распознавании дискретных объектов. И, во-вторых, задача выбора оптимального значения параметра разброса $\sigma$ гауссовской ядерной функции Парзена (4) для критерия (6) не столь актуальна, как для традиционного подхода (5).

## Заключение

Предложенная в настоящей работе модификация (Рис. 2) вероятностной нейронной сети [4] является обобщением классической PNN на задачу статистического распознавания образов. Выражение (6) обладает всеми преимуществами PNN [5] над другими нейросетевыми классификаторами. Прежде всего, это высокая скорость обучения по сравнению с традиционным алгоритмом обратного распространения ошибки. А дополнительное обучение может быть выполнено в режиме реального времени. Кроме того, структура сети (Рис. 2) не содержит рекурсивных связей, поэтому она допускает полную реализацию в параллельном варианте. Наконец, основное достоинство предложенного классификатора заключается в том, что при увеличении числа обучающих выборок решение (6) сходится к оптимальному байесовскому решению [1]. При этом скорость сходимости существенно превосходит скорость сходимости классической PNN в особенно актуальном для распознавания образов случае, когда объем обучающей выборки для каждого класса близок к объему входной выборки $n_r \approx n$.

Проведенный эксперимент показал, что, хотя наилучшая точность классификации у оригинальной PNN (Рис. 1) и модификации (Рис. 2) в случае гауссовской ядерной функции (4) практически совпадают, предложенный подход оказался значительно более устойчивым к изменению параметра разброса $\sigma$ функции активации (5). Последнее обстоятельство позволяет использовать предложенный критерий (6) и его упрощенную версию (13) при варьирующихся во времени условиях наблюдения.

Отметим еще одно потенциальное преимущество [7] использования предложенной модификации (6) в задаче распознавания образов по сравнению с оригинальной версией критерия (5) - синтезированное выражение (6), в отличие от (5), является симметричным. Действительно, в задаче проверки однородности выборки считаются равнозначными. А в статистической классификации

наблюдается асимметрия между распределением, заданным обучающей выборкой, и распределением входного объекта. Это обстоятельство может служить дополнительным обоснованием критерия (6), т.к. симметрия является желательным свойством [2] рассогласования между объектами для многих алгоритмов распознавания образов (таких, как кластеризация).

К сожалению, синтезированная сеть (Рис. 2) обладает также и известными недостатками сетей PNN [5] - использование достаточно большого объема памяти для хранения всех обучающих выборок и относительно низкая скорость классификации, обусловленная полным перебором всего множества альтернатив [12, 14]. Более того, вычислительная сложность критерия (6) в два раза превосходит сложность традиционной PNN (5). Однако это обстоятельство не должно служить препятствием к практической реализации (6) для распознавания образов, т.к. зачастую в этой задаче объем обучающих выборок (количество эталонов) относительно невелик

Для снижения вычислительной сложности и требуемого объема оперативной памяти в работе предложена упрощенная версия (13) алгоритма (6) для достаточно распространенного предельного случая ( $\sigma \to 0$ ) гауссовской ядерной функции. Повышение вычислительной эффективности связано с редукцией размерности пространства признаков - переходу от выборок к оценкам их распределений (гистограммам). Показано, что в случае распознавания дискретных случайных объектов критерий (6) эквивалентен критерию проверки однородности входной и обучающей выборок, основанному на принципе минимума информационного рассогласования Кульбака-Лейблера [8]. Более того, для этого случая предложенный подход (13) позволил существенно повысить точность распознавания по сравнению с реализацией (14) классической сети PNN.

Таким образом, благодаря проведенному исследованию предложена новая модификация вероятностной нейронной сети (Рис. 2) для задачи распознавания образов, обладающая высокими эксплуатационными свойствами. Синтезированный критерий (6) может быть с незначительными исправлениями использован в таких актуальных задачах, как распознавание изображений [12, 14], речи [7] и др.

## Литература


1. Боровков А.А. Математическая статистика: дополнительные главы, М.: Наука, 1984. 144с.
2. Koutroumbas K., Theodoridis S. Pattern recognition, Boston: Academic Press, 2008. - 840 с.
3. Vapnik V.N. Statistical learning theory. – New York: Wiley; 1998. – 732 p.
4. Savchenko A.V., Probabilistic neural network with homogeneity testing in recognition of discrete patterns set // Neural Networks. – 2013. – Vol. 46.– pp. 227-241.
5. Specht D.F. Probabilistic neural networks // Neural Networks. – 1990. – Vol. 3. – pp.109–118.
6. Rutkowski L. Adaptive probabilistic neural networks for pattern classification in time-varying environment //IEEE Transactions on Neural Networks.- 2004.- Vol.15 (4).- pp. 811-827
7. Савченко В.В. Обнаружение и исправление ошибок в задачах автоматического распознавания речи на основе принципа минимума информационного рассогласования // Известия вузов России. Радиоэлектроника. 2012. Вып.4. С.10-18.
8. Kullback S. Information theory and statistics. New York: Dover Pub, 1997. – 408 p.
9. Шевелев О.Г. Методы автоматической классификации текстов на естественном языке. – Томск: ТМЛ-Пресс, 2007. – 144с.
10. Цыпкин Я.З. Адаптация и обучение в автоматических системах, М.: Наука, 1968. - 400 с.
11. Савченко В.В., Савченко А.В. Принцип минимального информационного рассогласования в задаче распознавания дискретных объектов // Известия вузов России. Радиоэлектроника. – 2005. – Вып.3. – С.10-18.
12. Савченко А.В. Трехпороговая система автоматического распознавания изображений// Искусственный интеллект и принятие решений. - 2011. – №4. - С.102-109.
13. Айзерман М.А., Браверман Э.М., Розоноэр Л.И. Метод потенциальных функций в теории обучения машин. М.: Наука, 1970. -384 с.
14. Savchenko A.V., Directed enumeration method in image recognition // Pattern Recognition. – 2012. – 45(8). – pp. 2952-2961.
15. Библиотека Максима Мошкова [Электронный ресурс]. http://www.lib.ru. – Режим доступа: свободный.
16. Маннинг К., Рагхаван П., Шютце Х. Введение в информационный поиск, М.: Вильямс, 2011 - 528 с.
17. Николенко С.И., Тулупьев А.Л. Самообучающиеся системы, М.: МЦНМО, 2009. - 288 с.


**Савченко Андрей Владимирович.** Доцент Национального исследовательского университета Высшая школа экономики. Окончил Нижегородский государственный технический университет в 2008 году. Кандидат технических наук. Автор 40 печатных работ. Область научных интересов: распознавание образов, искусственный интеллект, новые информационные технологии. E-mail: avsavchenko@hse.ru