

## **АНАЛИЗ ИСПОЛЬЗОВАНИЯ НЕЙРОСЕТЕЙ ДЛЯ ДИАГНОСТИКИ РАКА ШЕЙКИ МАТКИ ПО МУЛЬТИСПЕКТРАЛЬНОМУ ИЗОБРАЖЕНИЮ**

**Е.Н. МАЛЫШЕВСКАЯ**

Рассмотрена проблема диагностики рака шейки матки с помощью нейронных сетей по мультиспектральному изображению. Для диагностики использованы нейронные сети с обратным распространением ошибки и сети с радиальным базисом. Приведены результаты экспериментальных исследований.

### **ВВЕДЕНИЕ**

В данной работе рассмотрена возможность диагностики рака шейки матки. Такая задача продиктована необходимостью ранней диагностики заболевания используя компьютерную систему, которая поможет врачу определить области с большим риском возникновения раковой трансформации ткани. Данная система базируется на утверждении, что оптические свойства здоровой клетки отличаются от свойств больной и это отличие более выражено, чем отличие клеток разных людей. В медицинском университете Аризоны (США) была внедрена оптическая система, в которой кроме обычного кольпоскопического тестирования проводилось снятие мультиспектральных электронных изображений тканей шейки матки. Были сделаны мультиспектральные снимки 108 пациенткам. Параллельно, те же самые пациентки прошли осмотр у врача, который состоял в том, что была сделана биопсия из определенных участков и отослана врачу-патологу на анализ. Участки, из которых была сделана биопсия, были обозначены на снимке и результаты биопсии были сопоставлены с указанными участками. Эта информация использовалась для создания алгоритма распознавания зон риска заболевания на основе сопоставления мультиспектрального снимка, результатов биопсии и вывода врача. На основе полученной информации создана система позволяющая проводить раннюю диагностику заболевания [1].

### **ПОСТАНОВКА ЗАДАЧИ**

По данным медицинской статистики рак шейки матки (РШМ) среди онкологических заболеваний, возникающих у женщин, занимает четвертое место (после рака желудка, кожи и молочных желез).

Первичный кольпоскопический осмотр пациентки врачом гинекологом определяет необходимость проведения биопсии и последующую консультацию онколога [3]. Рассмотренная в данной работе компьютерная система может быть использована врачом для преждевременной диагностики рака путем определения наличия определенных типов тканей в шейке матки без

проведения биопсии. Европейская Экспертная Группа, разрабатывающая Европейскую обобщенную обучающую программу, предложила следующую классификацию изменений эпителия шейки матки: SEA (squamous epithelial abnormalities) — доброкачественные изменения плоского эпителия позволяющие предположить CIN; Squamous cell changes — изменения плоского эпителия без четких признаков опухоли; CIN-I — дисплазия легкой степени; CIN-II — дисплазия умеренной степени; CIN-III — интраэпителиальная неоплазия тяжелой степени, данное понятие объединяет тяжелую дисплазию и внутриэпителиальный рак (CIS — carcinoma in situ) рак, подозрительный на инвазию, инвазивный плоскоклеточный рак.

**Цель работы** — создание компьютерной системы, которая сможет правильно классифицировать разные виды тканей по мультиспектральному изображению, применяя нейронные сети (НС). В данной работе рассмотрена, в первую очередь диагностика состояний дисплазии и неоплазии (CIN-I, II, III), а также наличие тканей с доброкачественными изменениями (MET). Такая задача диагностики является классической задачей для которой применяются НС. В работе применены нейронные сети с разными алгоритмами обучения.

## ОСНОВНЫЕ ИДЕИ МЕТОДА ОБРАТНОГО РАСПРОСТРАНЕНИЯ ОШИБКИ

Данный метод является методом обучения искусственных нейронных сетей, который впервые был описан Полем Дж. Вербосом в 1974 г., и далее развит Дэвидом И. Румельхартом и Рональдом Дж. Вильямсом в 1986 году.

Основная идея этого метода состоит в распространении сигналов ошибки от выходов сети к ее входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы.

Для возможности применения этого метода, передаточная функция нейронов должна быть дифференцируема.

### Описание алгоритма

На рис. 1 изображена нейронная сеть. У сети есть входы  $x_1, \dots, x_n$ , выходы Outputs и внутренние узлы. Перенумеруем все узлы (включая входы и выходы) числами от 1 до  $N$ . Обозначим через  $w_{i,j}$  вес, стоящий на ребре, соединяющем  $i$ -й и  $j$ -й узлы, а через  $o_i$  — выход  $i$ -го узла. Для  $m$ -тестовых примеров с целевыми значениями выходов  $\{t_k^d\}$   $d = 1 \dots m$ ,  $k \in \text{Outputs}$  функция ошибки выглядит так:

$$E(\{w_{i,j}\}) = \frac{1}{2} \sum_{d=1}^m \sum_{k \in \text{Outputs}} (t_k^d - o_k(x_1^d, \dots, x_n^d))^2.$$

Для модификации весов реализуется стохастический градиентный спуск, то есть подправляются веса после каждого тестового примера. Необходимо двигаться в сторону, противоположную градиенту, то есть добавлять к каждому весу  $w_{i,j}$   $\Delta w_{i,j} = -\eta \frac{\partial E^d}{\partial w_{i,j}}$ , где

$$E^d(\{w_{i,j}\}) = \frac{1}{2} \sum_{k \in \text{Outputs}} (t_k^d - o_k^d)^2.$$

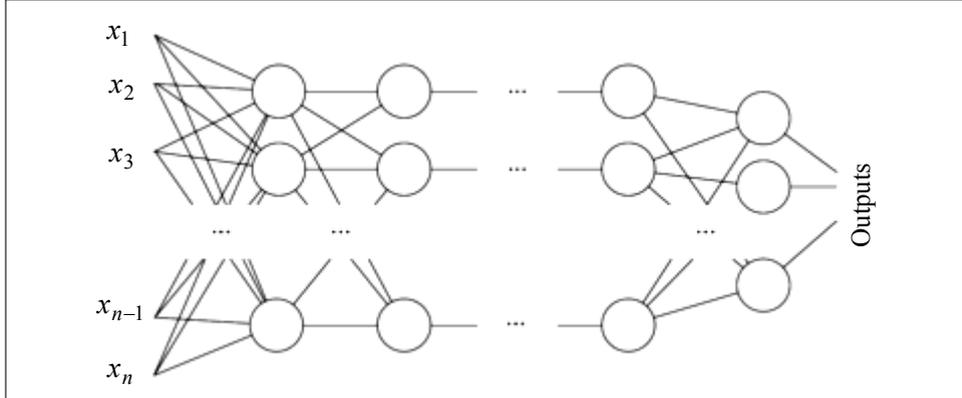


Рис. 1. Пример нейронной сети

Производная вычисляется следующим образом. Пусть сначала  $j \in \text{Output}$  (интересующий вес) входит в перцептрон последнего уровня. Необходимо отметить, что  $w_{i,j}$  влияет на выход перцептрона только как часть суммы

$$S_j = \sum w_{i,j} x_{i,j}, \text{ где сумма берется по входам } j\text{-го узла. Поэтому } \frac{\partial E^d}{\partial w_{i,j}} = \frac{\partial E^d}{\partial S_j} \frac{\partial S_j}{\partial w_{i,j}} = x_{i,j} \frac{\partial E^d}{\partial S_j}.$$

Аналогично,  $S_j$  влияет на общую ошибку только в рамках выхода  $j$ -го узла  $o_j$  (выход всей сети). Поэтому

$$\begin{aligned} \frac{\partial E^d}{\partial S_j} &= \frac{\partial E^d}{\partial o_j} \frac{\partial o_j}{\partial S_j} = \left( \frac{\partial}{\partial o_j} \frac{1}{2} \sum_{k \in \text{Outputs}} (t_k^d - o_k^d)^2 \right) \left( \frac{\partial \sigma(S_j)}{\partial S_j} \right) = \\ &= \left( \frac{1}{2} \frac{\partial}{\partial o_j} (t_j - o_j)^2 \right) (o_j(1 - o_j)) = -o_j(1 - o_j)(t_j - o_j). \end{aligned}$$

Если же  $j$ -й узел — не на последнем уровне, то у него есть выходы; обозначим их через  $\text{Outputs}(j)$ . В этом случае  $\frac{\partial E^d}{\partial S_j} = \sum_{k \in \text{Outputs}(j)} \frac{\partial E^d}{\partial S_k} \frac{\partial S_k}{\partial S_j}$  и

$$\frac{\partial S_k}{\partial S_j} = \frac{\partial S_k}{\partial o_j} \frac{\partial o_j}{\partial S_j} = w_{i,j} \frac{\partial o_j}{\partial S_j} = w_{i,j} o_j(1 - o_j).$$

Здесь  $\frac{\partial E^d}{\partial S_k}$  — в точности аналогичная поправка, но вычисленная для узла следующего уровня (обозначим ее через  $\delta_k$ ). От  $\Delta k$  она отличается отсутствием множителя  $(-w_{i,j})$ . При помощи формул для вычисления поправки для узлов последнего уровня и выражения поправки для узла более

низкого уровня через поправки более высокого уровня можно составить алгоритм. Именно из-за этой особенности вычисления поправок алгоритм называется алгоритмом обратного распространения ошибки (backpropagation). Краткое резюме рассмотренного алгоритма:

- для узла последнего уровня

$$\delta_j = -o_j(1 - o_j)(t_j - o_j);$$

- для внутреннего узла сети

$$\delta_j = -o_j(1 - o_j) \sum_{k \in \text{Outputs}} \delta_k w_{j,k};$$

- для всех узлов

$$\Delta w_{i,j} = -\eta \delta_j x_{i,j}.$$

На вход алгоритма, кроме указанных параметров, необходимо также подавать в каком-либо формате структуру сети. На практике очень хорошие результаты показывают сети достаточно простой структуры, состоящие из двух уровней нейронов — скрытого уровня (hidden units) и нейронов-выходов (output units). Каждый вход сети соединен со всеми скрытыми нейронами, а результат работы каждого скрытого нейрона подается на вход каждому из нейронов-выходов. В таком случае достаточно подать на вход количество нейронов скрытого уровня [4].

### Недостатки алгоритма

Несмотря на многочисленные успешные применения обратного распространения, оно не является панацеей. Больше всего неприятностей приносит неопределенно долгий процесс обучения. В сложных задачах для обучения сети могут потребоваться дни или даже недели, она может и вообще не обучиться. Причиной может быть одна из описанных ниже.

*Паралич сети.* В процессе обучения сети значения весов могут в результате коррекции стать очень большими величинами. Это может привести к тому, что все или большинство нейронов будут функционировать при очень больших выходящих значениях, в области, где производная сжимающей функции очень мала. Так как посылаемая обратно в процессе обучения ошибка пропорциональна этой производной, то процесс обучения может практически остановиться. В теоретическом отношении эта проблема плохо изучена. Ее можно избежать уменьшив размер шага  $\eta$ , но это увеличивает время обучения. Различные эвристики используются для предохранения от паралича или для восстановления после него, но пока что они могут рассматриваться лишь как экспериментальные.

*Локальные минимумы.* Обратное распространение использует разновидность градиентного спуска, т.е. осуществляет спуск вниз по поверхности ошибки, непрерывно подстраивая веса в направлении к минимуму. Поверхность ошибки сложной сети сильно изрезана и состоит из холмов, долин, складок и оврагов в пространстве высокой размерности. Сеть может попасть в локальный минимум (неглубокую долину), когда рядом имеется гораздо более глубокий минимум. В точке локального минимума все направления ведут вверх, и сеть неспособна из него выбраться. Статистические методы

обучения могут помочь избежать этой ловушки, но они требуют большего затрата времени.

*Размер шага.* Внимательный разбор доказательства сходимости показывает, что коррекции весов предполагаются бесконечно малыми, но это неосуществимо на практике, так как ведет к бесконечному времени обучения. Размер шага должен браться конечным, и в этом вопросе приходится опираться только на опыт. Если размер шага очень мал, то сходимость слишком медленная, если же очень велик, то может возникнуть паралич или постоянная неустойчивость. Описан адаптивный алгоритм выбора шага, автоматически корректирующий размер шага в процессе обучения.

## СЕТИ РАДИАЛЬНОГО БАЗИСА

Самообучающиеся сети, как известно, широко используются для преобразования данных, например при распознавании образов в пространстве очень большой размерности. В этом случае для того, чтобы процедура обучения с учителем была эффективна, необходимо сначала сжать входную информацию тем или иным способом, либо выделить значимые признаки, понизив размерность, либо произвести квантование данных. Первый путь просто понижает число входов персептрона. Второй способ требует отдельного рассмотрения, поскольку лежит в основе очень популярной архитектуры — сетей радиального базиса (radial basis functions — RBF).

### Аппроксиматоры с локальным базисом

Сети радиального базиса напоминают персептрон с одним скрытым слоем, осуществляя нелинейное отображение  $\mathcal{R}^d \Rightarrow \mathcal{R}^m$   $y = \sum_i h_i \phi(w_i, x)$ , являю-

щееся линейной комбинацией базисных функций. Но в отличие от персептронов, где эти функции зависят от проекций на набор гиперплоскостей  $\sigma(wx)$ , в сетях радиального базиса используются функции (чаще всего — гауссовы), зависящие от расстояний до опорных центров:  $y = \sum_i h_i \phi_i(|w_i - x|)$ ,

$\phi_i(z) = e^{-z^2 / \sigma_i^2}$ . Все наборы базисных функций обеспечивают возможность аппроксимации любой непрерывной функции с произвольной точностью. Основное различие между ними в способе кодирования информации на скрытом слое. Если персептроны используют глобальные переменные (наборы бесконечных гиперплоскостей), то сети радиального базиса опираются на компактные шары, окружающие набор опорных центров (рис. 2).

В первом случае в аппроксимации в окрестности любой точки участвуют все нейроны скрытого слоя, во втором — лишь ближайšie. Как следствие такой неэффективности, в последнем случае количество опорных функций, необходимых для аппроксимации с заданной точностью, возрастает экспоненциально с размерностью пространства. Это основной недостаток сетей радиального базиса. Основное же их преимущество над персептронами — в простоте обучения [5].

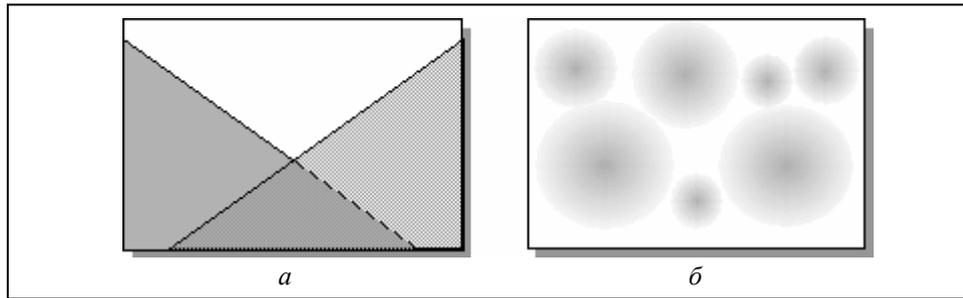


Рис. 2. Глобальная (персептроны) и локальная (сети радиального базиса) методы аппроксимации

### Гибридное обучение

Относительная автономность базисных функций позволяет разделить обучение на два этапа. На первом этапе обучается первый (соревновательный) слой сети, осуществляя квантование данных. На втором этапе происходит быстрое обучение второго слоя матричными методами, т.к. нахождение коэффициентов второго слоя представляет собой линейную задачу.

Подобная возможность раздельного обучения слоев является основным достоинством сетей радиального базиса. В целом же, области применимости персептронов и сетей радиального базиса коррелируют с найденными выше областями эффективности квантования и понижения размерности.

### ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ

Большее количество показателей имеют очень высокую корреляцию друг с другом, что может негативно влиять на работу нейронной сети для классификации тканей на больную/здоровую. Для этого во избежание потенциальных проблем в двух случаях показатели меняются. В одном случае, для обучения нейронной сети используется только подмножество показателей, между которыми коэффициент корреляции не превышает 0,95 (определен экспериментально). В другом случае, используется метод основных компонент, показатели которого трансформируются в набор независимых некоррелируемых показателей. Эти методы должны улучшить работу нейронной сети и увеличить точность классификации.

Метод главных компонент (МГК) (в англоязычной литературе — principal component analysis — PCA) — один из наиболее распространенных методов факторного анализа. Суть метода состоит в декомпозиции матрицы данных  $X$  и представлении ее в виде произведения двух матриц:  $T$  (матрица счетов, scores) и  $P$  (матрица нагрузок, loadings):  $X = TPT$ , что, по сути, является переходом к новым переменным.

Одно из важнейших свойств такого преобразования — ортогональность новых переменных:  $PPT = I$ ,  $TTT = I$ , где  $I$  — единичная матрица значительно упрощает дальнейшие операции с матрицами и позволяет ранжировать новые переменные в соответствии с их значимостью. Как правило, используются только значимые переменные («главные компоненты»), а незначимые переменные (соответствующие случайным погрешностям) ис-

ключаются из дальнейшего рассмотрения, что позволяет сжать данные и значительно снизить погрешность расчетов [6].

## РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Имеются диагнозы двух врачей, представляющие собой процентное содержание клеток каждого из шести возможных типов тканей (Squamous, Columnar, MET, CIN1, CIN2, CIN3). В таблице приведены результаты работы нейронных сетей. При проведении экспериментов использовался метод Cross validation (скользящий контроль). Процедура скользящего контроля заключается в следующем. Фиксируется некоторое множество разбиений выборки на две части: обучающую и проверочную. Для каждого разбиения выполняется настройка алгоритма по обучающей подвыборке и вычисляется частота его ошибок на проверочной подвыборке. Оценка скользящего контроля определяется как средняя по всем разбиениям частота ошибок на проверочной. Фактически скользящий контроль непосредственно измеряет обобщающую способность метода обучения на заданой конечной выборке.

**Т а б л и ц а .** Сравнительный анализ работы нейронных сетей

Средне-квадратичное отклонение (СКО)	Алгоритм обратного распространения ошибки. Метод сопряженных градиентов	Алгоритм обратного распространения ошибки. Метод квази Ньютона	Алгоритм обратного распространения ошибки. Метод масштабируемых сопряженных градиентов	Сеть радиального базиса
СКО (>0,95) Squamous	0,0607	0,0593	0,0499	0,0325
СКО (PCA) Squamous	0,0502	0,0621	0,0507	0,0326
СКО (>0,95) Columnar	0,0459	0,0457	0,0449	0,0295
СКО (PCA) Columnar	0,0471	0,0480	0,0408	0,0297
СКО (>0,95) MET	0,0561	0,0530	0,0492	0,0244
СКО (PCA) MET	0,0580	0,0577	0,0582	0,0247
СКО (>0,95) CIN1	0,0462	0,0852	0,0408	0,0220
СКО (PCA) CIN1	0,0817	0,0958	0,0458	0,0224
СКО (>0,95) CIN2	0,0427	0,0497	0,0345	0,0158
СКО (PCA) CIN2	0,0392	0,0500	0,0352	0,0158
СКО (>0,95) CIN3	0,0734	0,0789	0,0739	0,0305
СКО (PCA) CIN3	0,0656	0,0755	0,0877	0,0306

В таблице представлено СКО для 4 НС, с двумя методами предварительной обработки данных (СКО(>0,95)) для подмножества показателей, между которыми коэффициент корреляции не превышает 0,95 и СКО (РСА) (данные с использованием метода основных компонент) для 6 типов тканей. Из таблицы видно, что сеть RBF дает наименьшую ошибку, при этом оба метода предварительной обработки данных равноценны.

## ВЫВОДЫ

1. Результаты проведенных экспериментов показали целесообразность использования нейронных сетей для решения задач классификации типов тканей на медицинских изображениях.

2. Исходя из проведенных экспериментов видно, что сеть RBF всегда дает наилучшие результаты.

3. Использование различных способов предварительной обработки данных значительно улучшает работу нейронных сетей. После проведения ряда экспериментов оба способа оказались достаточно эффективными.

4. Разработанный подход диагностики рака шейки матки с использованием нейронных сетей может быть рекомендован для дальнейшей клинической апробации.

## ЛИТЕРАТУРА

1. *Schoonmaker J. et al.* Automatic Segmentation of Uterine Cervix for in vivo Localization and Identification of Cervical Intraepithelial Neoplasia. — <http://spie.org/Documents/ConferencesExhibitions/mi07-final.pdf>.
2. *Воробьева Л.И.* Цитологический скрининг рака шейки матки // Здоровье Украины. — 2008. — № 2/1. — С. 18.
3. *Koss L.G.* The complex test for cervical cancer detection // The Journal of the American Medical Association. — 1989. — **261**, № 5. — С. 737–743.
4. *Зайченко Ю.П.* Основы проектирования интеллектуальных систем. Навчальний посібник. — Київ: Видавничий Дім «Слово», 2004. — 352 с.
5. *Domagoj Kovacevic, Sven Loncaric.* Radial Basis Function-based Image Segmentation using a Receptive Field // Computer-Based Medical Systems. — 1997. Proceedings., Tenth IEEE Symposium. — P. 126–130.
6. *Richard G. Brereton* Introduction to multivariate calibration in analytical chemistry // Analyst. — 2000. — № 125. — P. 2125–2154.

Поступила 02.07.2008