

and IT Education" (Moscow, November 8, 2010) .- М.: MSU. N.V.Lomonosova, 2010, pp 85 - 95.

5. Nekhaev I.N, Vlasov A.A. Intelligent adaptive testing level of learning / IN Nekhaev, AA Vlasov // CAI-2010. Twelfth National Conference on Artificial Intelligence with international participation (September 20 - 24, 2010, Tver, Russia). Collection trudov.- FIZMATLIT M.: 2010, Vol.3, 257-263.

ИСПОЛЬЗОВАНИЕ МЕТОДА САМООРГАНИЗАЦИИ НЕЙРОННОЙ СЕТИ В ЗАДАЧЕ СИНТЕЗА МЕДИЦИНСКОГО КЛАССИФИКАТОРА

аспирант Нежинских С.С., д.т.н., профессор Миркин Е.Л.,
Международный Университет Кыргызстана, Бишкек, Кыргызская Республика,
E-mail: eugene_mirkin@mail.ru

USING METHOD OF SELF-ORGANIZING NEURAL NETWORK IN THE PROBLEM OF SYNTHESIS OF MEDICAL CLASSIFIER

Nezhinskikh, S.S.Mirkin E.L., International University of Kyrgyzstan, Bishkek, Kyrgyz Republic
E-mail: eugene_mirkin@mail.ru

Метод самоорганизации нейронной сети, предложенный в работе [1], использован здесь для решения задачи медицинской диагностики заболевания «диабет». Топология нейронной сети была создана автоматически, с учётом «сшивания» её элементов по валидационному признаку. Созданный медицинский классификатор показал хорошую разделяющую способность.

Введение. Повсеместная компьютеризация медицинских учреждений, а так же повышение производительности вычислительных систем, позволила использовать современные интеллектуальные методы принятия решений в области медицинской диагностики. Эффективность использования нейронных сетей в задачах диагностирования медицинских заболеваний заключается в том, что модель, построенная на нейронной сети, может работать с большим набором параметров, влияние которых на формулировку диагноза оценить достаточно сложно. Нейронные сети способны принимать решения, основываясь на выявляемых ими скрытых закономерностях в многомерных данных.

Главная проблема решения задач с использованием нейронных сетей заключается в формировании топологии сети, потенциально способной решить поставленную задачу. Процесс подбора приемлемой топологии нейронной сети является трудоёмким и зависит от личного опыта исследователя.

В статье для создания модели диагностирования заболевания сахарным диабетом, предлагается использовать метод самоорганизации нейронной сети [1].

Постановка задачи исследования. Ставится задача разработать модель классификатора, способного выявлять заболевание диабетом у пациентов. Приведём основные медицинские характеристики заболевания сахарным диабетом.

Сахарный диабет – это хроническое эндокринное заболевание, вызванное формированием инсулиновой недостаточности или инсулиновой резистентности в организме человека, в результате которой происходит развитие гипергликемии (продолжительного повышенного уровня глюкозы в крови) и нарушение метаболических процессов. Распространенность заболевания на сегодняшний момент в мире, согласно данным Всемирной Организации Здравоохранения, составляет около 347 млн. человек, и эта цифра растет каждый год [2]. Сахарный диабет сейчас находится на третьем месте рейтинга после онкологических и сердечно-сосудистых заболеваний.

Синтез нейросетевого классификатора.

Проблемой диагностики диабета занимались многие исследователи в области медицины и, к настоящему времени, был накоплен богатый статистический материал по этому заболеванию. Задачей работы нейронной сети будет определение исхода заболевания – отнесения пациента к одному из двух классов (1 – «здоров», 2 – «болен»). Для построения эффективного классификатора необходимо выявить, как параметры исходных данных влияют на принятие решения о том, к какому классу принадлежит пациент.

В статье для обучения классификатора были использованы данные, предоставленные «Национальным институтом диабета, желудочно-кишечных и почечных заболеваний» (г. Финикс, штат Аризона, США), находящиеся в открытом доступе в «Репозитории машинного обучения», интернет-ресурса «Центр для машинного обучения и интеллектуальных систем» (Университет Калифорнии, г. Ирвайн, штат Калифорния, США) [3,4,5].

В таблице 1 представлена структура базы пациентов, вышеуказанного медицинского учреждения. Количество записей соответствует 768 пациентам. Все пациенты - женщины не моложе 21 года.

Таблица 1. Структура базы данных Национального института диабета, желудочно-кишечных и почечных заболеваний.

№	Параметр	Ед. измерения	Запись 1	Запись 2	...	Запись 768
1	Число случаев беременности	Число	6	1	...	1
2	Концентрация глюкозы	мг/дл	148	85		93
3	Диастолическое артериальное давление	мм рт. ст.	72	66		70
4	Толщина кожной складки трехглавой мышцы	Мм	35	29		31
5	2-х часовой сывороточный инсулин	ед/мл	0	0		0
6	Индекс массы тела	кг/м ²	33.6	26.6	...	30.4
7	Числовой параметр наследственности диабета	Число	0.627	0.351		0.315
8	Возраст	Лет	50	31		23
9	Наличие заболевания	Число	1	0		0

Приведём описание полей, характеризующих состояние пациента на момент диагностирования:

1. Число случаев беременности.
2. Концентрация глюкозы в плазме крови (определяется по результатам перорального глюкозотолерантного теста) – для большей точности постановки диагноза часто используют тест на толерантность к глюкозе. Он также носит название «построение сахарной кривой» или «исследование под нагрузкой». С помощью него можно не только подтвердить наличие у пациента заболевания сахарным диабетом, но и выявить скрытую форму диабета, который не удастся распознать с помощью других анализов.
3. Диастолическое артериальное давление – показывает давление в артериях в момент расслабления сердечной мышцы. Это минимальное давление в артериях, оно отражает сопротивление периферических сосудов.
4. Толщина кожной складки трехглавой мышцы (кожно-жировая складка над трицепсом) – влияет на оценку энергетического баланса. Толщина кожной складки над трехглавой мышцей плеча зависит от телосложения и роста. При ожирении этот показатель не дает дополнительных сведений по сравнению с индексом массы тела, но при сильном истощении он особенно информативен.
5. 2-х часовой сывороточный инсулин.
6. Индекс массы тела – величина, позволяющая оценить степень соответствия массы человека и его роста и, тем самым, косвенно оценить, является ли масса недостаточной, нормальной или избыточной. Рассчитывается по формуле $I = \frac{m}{h^2}$, где m — масса тела в килограммах, h — рост в метрах.
7. Числовой параметр наследственности диабета.
8. Возраст.
9. Наличие заболевания: 1 – «болен», 0 – «здоров».

Создание модели нейронной сети.

Для создания нейросетевого классификатора будем использовать метод формирования топологии самоорганизующейся нейронной сети [1]. При этом, определение дискретных моментов времени эволюции сети будет улучшено с помощью ранней остановки обучения нейронной сети, что позволяет устранить эффект переобучения. Для этого разделим исходные данные на три группы: обучающую, валидационную и тестирующую. Первая группа данных будет использована для настройки весов и смещений нейронной сети, вторая – для расчёта интегрированной ошибки обучения сети. Наконец, третья группа, необходима для контроля качества прогноза обученной модели и не используется при обучении нейронной сети.

Предлагается, в соответствии с методом [1], улучшенный алгоритм самоорганизации нейронной сети:

- Первый шаг. Построение топологии начинается с одного нейрона, для которого задаются начальные веса w_{1x}^1 , где верхний индекс – номер слоя; нижний индекс – номер нейрона соединённый с x – индексом веса, и смещение b_1^1 . Так же, на данном шаге необходимо задать максимальную сложность топологии сети.
- Второй шаг. В дополнение к контролю интегрированной ошибки E_n (где n – номер эпохи), процесс обучения будет завершён, если интегрированная ошибка сети при обучении на валидационной группе

данных E_n^{val} будет увеличиваться по отношению к E_n в течении 6 эпох. Дальнейшая последовательность действий соответствует используемому методу.

- На каждом последующем шаге проверяется количество нейронов в каждом слое. Учитывая заданную максимальную сложность топологии нейронной сети, параллельно в проверяемый слой добавляется один нейрон. Если количество нейронов в проверяемом слое равно двум, то добавляется новый слой с одним нейроном и параметрами $b_1^{l+1} = 0$, $w_1^{l+1} = [1 \ 0]$. За один цикл разрешено добавить лишь один нейрон и/или один слой.

При этом запоминается конфигурация нейронной сети Net_i (где i – номер эволюции сети), при которой интегрированная ошибка E_n была улучшена. Если E_n не изменялась в течении 6 эволюций сети, то текущая топология сети считается избыточной: конфигурация сети возвращается к последней успешной $Net_{current} = Net_i$, а количество нейронов в текущем слое в последующих шагах не изменяется.

Компьютерное моделирование.

Исходные данные были разделены следующим образом: обучающая выборка составила 80% от исходного количества, валидационная – 15%, тестирующая – 5%.

Множители максимальной сложности сети заданы в виде вектора, состоящего из двух чисел [2 0.5], что означает: в первом слое максимальное количество нейронов равно количеству входного множества x умноженного на первое число заданного вектора ($N_{max}^1 = 8 * 2 = 16$); во втором, и в каждом последующем слое, максимальное количество нейронов равно произведению максимального количества нейронов в предыдущем слое на второе число заданного вектора: $N_{max}^2 = N_{max}^1 * 0.5 = 8$. Начальная топология нейронной сети состоит из одного нейрона с функцией активации f – гиперболический тангенс. Начальные значения для данного нейрона заданы в виде: $b_1^1 = [-0.32]$, $w_1^1 = [-0.41 \ 0.49 \ -0.98 \ -0.90 \ 0.33 \ 0.21 \ 0.05 \ 0.46]$. При достижении порога $\bar{E} = 0.025$, либо достижении максимальной сложности топологии, сеть будет считаться обученной.

На шестой эпохе, первый этап обучения сети методом «обратного распространения ошибки» [7] был прерван, так как интегрированная ошибка сети на валидационной группе увеличивалась с 6 по 12-ую эпоху. Сеть считалась необученной, так как интегрированная ошибка нейронной сети E уменьшилась с величины $E_0 = 0.87$ до $E_6 = 0.15$, не достигнув заданного порога $\bar{E}$.

В соответствии с предложенной процедурой, на втором этапе в нейронную сеть были добавлены два нейрона – один параллельно в первый слой, второй в новый слой.

После добавления третьего нейрона, интегрированная ошибка сети не изменилась (в соответствии с правилами валидации) и составила $E_{20} = 0.136$, что аналогично предыдущей эволюции топологии сети.

На четвертом этапе эволюции сети, число нейронов в первом слое достигло пяти, а значение интегрированной ошибки E_{24} значения 0.13. В течении последующих шести эпох добавление нейронов в первый слой не привело к изменению этого значения. В соответствии с предложенным алгоритмом, нейронная сеть была возвращена к текущему состоянию, а максимальное количество нейронов для первого слоя установлено в значение 5.

После добавления третьего нейрона во второй слой, интегрированная ошибка сети достигла значения $E_{26} = 0.122$. Дальнейшее добавление нейронов во второй слой не привело к улучшению данного показателя, и сеть была сконфигурирована в соответствии с текущим состоянием. Хотя E_{26} не достигла целевого порогового уровня $\bar{E}$, дальнейшее обучение нейронной сети было завершено, так как максимальная сложность топологии сети была достигнута.

На рисунке 2 приведён график изменения интегрированной ошибки обучения сети в процессе её эволюции.

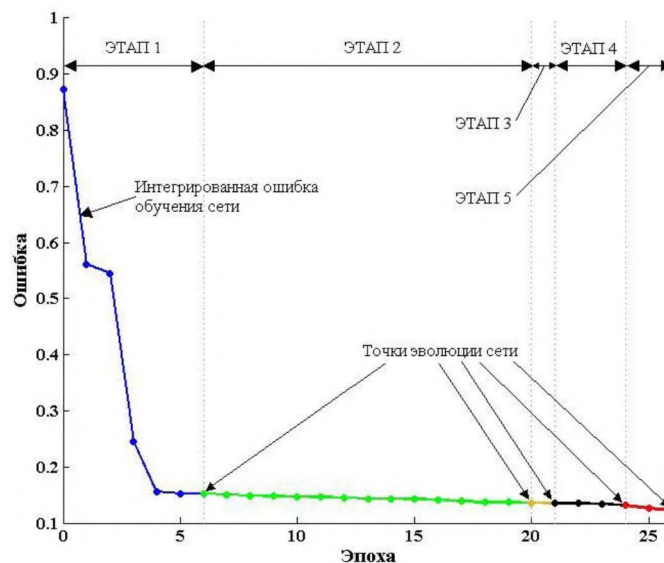


Рис. 2. Динамика изменения интегрированной ошибки обучения сети

Эффективность работы обученной нейронной сети составила 77.6%, что является приемлемым результатом. Этот факт позволяет использовать классификатор в медицинских целях.

Вывод. В работе использован метод самоорганизации нейронной сети [1] для создания медицинского классификатора. При этом методика самоорганизации топологии нейронной сети была модифицирована автоматическим «сшиванием» компонентов сети по валидационному признаку. Ранняя остановка обучения нейронной сети, привела к эффективному определению дискретных моментов времени эволюции сети, тем самым позволила устранить эффект переобучения.

Список литературы

1. Миркин Е. Л., Нежинских С. С. Метод формирования топологии самоорганизующейся нейронной сети в процессе её обучения // «Проблемы автоматизации и управления», Бишкек. 2014. – №2. С.28 – 36.
2. Danaei G, Finucane MM, Lu Y, Singh GM, Cowan MJ, Paciorek CJ et al. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. Lancet, 2011, 378(9785):31–40.
3. Center for Machine Learning and Intelligent Systems; Machine Learning Repository; University of California, Irvine. [<http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>].
4. The Human Genome Project and Diabetes: Genetics of Type II Diabetes. New Mexico State University. 1997. 1 June 2006. [<http://darwin.nmsu.edu/~molbio/diabetes/disease.html>].
5. National Health and Nutrition Examination Survey (NHANES I, II) [<http://www.cdc.gov/nchs/nhanes.htm>].
6. Аксенов С.В., Новосельцев В.Б. Организация и использование Нейронных сетей (методы и технологии). НТЛ, Томск, 2006.
7. Саймон Хайкин. Нейронные сети. Полный курс. 2-ое изд. 2006. – М.: Вильямс, 2006. – 1104 с.

УДК 004.584

РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ ЛОГИСТИЧЕСКОЙ СИСТЕМЫ НА ОСНОВЕ RFID-ТЕХНОЛОГИЙ

Злобина М.И. Алтайский государственный университет, Барнаул, Россия,
E-mail: marizlo.ftf@mail.ru

DEVELOPMENT OF AUTOMATED LOGISTICS SYSTEM BASED ON RFID-TECHNOLOGY

Zlobina M.I. Altai State University, Barnaul, Russia, E-mail: marizlo.ftf@mail.ru

В статье проведено исследование систем сбора и обработки данных. Спроектирована автоматизированная логистическая система инвентаризационного учета. Приведен один из вариантов её реализации.