

СИСТЕМА ВЫЯВЛЕНИЯ АНОМАЛЬНЫХ НАБЛЮДЕНИЙ В ДАННЫХ ОБ ОКАЗАНИИ МЕДИЦИНСКОЙ ПОМОЩИ НАСЕЛЕНИЮ

С. Л. Блюмин, Р. В. Щеглеватых, А. А. Найденов, А. С. Сысоев

*Кафедра прикладной математики, sysoev_as@stu.lipetsk.ru;
ФГБОУ ВО «Липецкий государственный технический университет»,
г. Липецк, Россия*

Ключевые слова: анализ конечных изменений; анализ чувствительности; аномальные наблюдения; классификация; математическое моделирование; поиск аномалий.

Аннотация: Дано описание математической модели нейросетевого классификатора данных о фактах оказания медицинской помощи в учреждениях Липецкой области в целях выявления нетипичных (аномальных) записей. Выявление аномалий относится к проблеме нахождения данных, не соответствующих некоторому ожидаемому поведению процесса или показателю, возникающему в системе. Из-за большого количества входов нейросетевой модели возрастает и время, которое требуется для обработки поступающей информации. Для оценки того, какие факторы должны быть переданы на вход нейросетевого классификатора, предложен подход к редукции нейросетевой модели, основанный на анализе чувствительности. Представлено описание комплекса программных средств для решения поставленной задачи.

Введение

Широкое внедрение информационных технологий во все сферы жизни человека приводит к упрощению рутинных операций, сокращению числа возникающих технических ошибок, облегчению доступа к необходимой информации, и, следовательно, повышению качества жизни. Бурно развивающиеся системы искусственного интеллекта способны анализировать информацию, поступающую в режиме реального времени, находить закономерности в имеющихся данных, а также выдавать рекомендации по управлению системой или процессом.

Известны применения интеллектуальных систем для анализа медицинской информации, особое место среди которых занимают экспертные медицинские системы, способные на основании анализа ретроспективной информации и выработки определенных онтологических правил построить предположение относительно диагностируемого у пациента заболевания и определить приемлемый вариант лечения [1 – 4]. Например, в исследовании [3] приведен подход к синтезу нечетких решающих правил для оценки и прогнозирования состояния пациента при различных типах заболеваний. Актуальными также являются задачи анализа рентгенологических снимков, снимков компьютерной и магнитно-резонансной томографий для анализа состояния больного и поиска аномалий в тканях [5 – 9]. Данные разработки особенно популярны, так как вплоть до настоящего времени

для гарантированной постановки диагноза необходимо было прибегать к не всегда возможному и порой опасному исследованию, такому как биопсия тканей пациента. Системы сегментации и распознавания изображений способны снять необходимость проведения таких процедур.

Актуальной является задача исследования статистической медицинской информации, в частности информации об оказании медицинской помощи населению [10, 11]. Несмотря на принятый в России «ГОСТ Р 52636–2006 Электронная история болезни. Общие положения», каждое медицинское лечебное учреждение разрабатывает системы, адаптированные для своих нужд. Однако хранимые в таких системах данные удовлетворяют положениям Международного стандарта электронных историй болезни ISO 13606 и могут быть использованы для анализа качества услуг, оказываемых лечебными учреждениями населению. Одной из возникающих в этой сфере проблем является поиск аномальных записей в статистической информационной системе электронных историй болезни [12]. Такие записи могут трактоваться как ошибочные или рассматриваться как фальсификация результатов в целях получения прибыли в рамках программ страховой медицины. Экспертами фонда обязательного медицинского страхования проводятся регулярные выборочные проверки предоставляемых данных, однако, проверка в автоматическом режиме с использованием методов искусственного интеллекта позволит ускорить данный процесс и сократить издержки на обеспечение экспертизы.

Цель исследования – представить разработанные модели, методы, а также программные реализации системы выявления аномальных наблюдений в данных об оказании медицинской помощи населению.

Задача нейросетевой классификации

Рассматриваемая проблема представляет собой задачу нахождения аномальных наблюдений в массивах данных. Решению схожих проблем посвящено достаточное количество работ. В зависимости от специфики организации данных и типа их аномальности могут быть применены различные подходы [13 – 16]. Рассматривается массив данных, состоящий из реализаций n объектов, каждый из которых описывается m -мерным вектором признаков $X = (X_1, \dots, X_m)$ и принадлежит к одному из двух классов: K_1 – нормальные наблюдения, K_2 – аномальные. Множество показателей X объектов классов K_1 и K_2 является обучающей выборкой.

Ставится задача для каждого нового наблюдения $x = (x_1, \dots, x_m)$ вектора признаков с учетом совершения некоторой ошибки второго рода (возможность того, что доля нормальных наблюдений будет отнесена к аномальным) определить класс объекта K_i , $i = 1, 2$, к которому его следует отнести, то есть нужно построить решающее правило (с учетом ошибки первого рода) $h(x): X \rightarrow \{d_1, d_2\}$, где $h(s) = d_1 \Leftrightarrow s \in K_1$.

Для того чтобы отнести новое наблюдение к одному из возможных классов, определим дискриминантную пороговую функцию $t(x): X \rightarrow R$. Тогда классификатор примет вид

$$h^\lambda(x) = \begin{cases} d_1, & t(x) \geq \lambda, \\ d_2, & t(x) < \lambda. \end{cases} \quad (1)$$

В заданном решающем правиле (1) в качестве функции $t(x)$ для решения поставленной задачи используется выходное значение нейросетевой модели.

В проведенном исследовании качество классификации оценивалось на основе точности и полноты, а также комбинаций указанных характеристик [17]. Точность классификации $P(h)$ определяется как доля объектов, которые были рас-

познаны как объекты класса аномальных наблюдений. Полнота классификации $R(h)$ показывает, какая доля объектов, реально относящаяся к классу аномальных наблюдений, была предсказана; их среднее гармоническое (F -мера) [18] определяется по формуле

$$F(h) = \frac{2P(h)R(h)}{P(h) + R(h)}. \quad (2)$$

Рассматриваемая задача относится к классу задач бинарной классификации, одним из способов оценки порога разделения для которых является использование кривых ошибок (или ROC-кривых) – графиков зависимости полноты классификации $R(h)$ от величины специфичности по всем возможным значениям порога разделения λ [19].

Для обнаружения аномалий могут использоваться алгоритмы, основанные на методах классификации, в которых в качестве инструментов применяются деревья решений, модели нечеткой логики, наивные байесовские модели, генетические алгоритмы, нейронные сети, опорные векторы и др. [20]. Для повышения точности обнаружения аномальных наблюдений предложены также комбинированные методы, объединяющие в себе преимущества нескольких алгоритмов. К таким комбинациям относятся каскадные методы классификации с обучением (комбинация наивных байесовских моделей и деревьев решений, деревьев решений и метода опорных векторов) и комбинации схем классификации с обучением и без него (например, модификации методов опорных векторов и метод кластеризации k -средних (*англ.* k -means)). Данные методы предполагают, что в результате анализа будет получена модель, описывающая профиль «нормального» наблюдения. Однако существует и принципиально иной подход, основанный на построении модели, выявляющей значения, отличающиеся от всех тех, которые характерны для показателей рассматриваемой системы или процесса. Структура, лежащая в основе работы такого метода, – изолирующий лес. Изоляция в этом смысле означает отделение одной группы наблюдений от другой. Чтобы применить эту идею к каждому наблюдению, необходимо рассчитать некоторую меру восприимчивости, определяющую порог разделения. Естественными структурами, разделяющими данные, являются случайно сгенерированные двоичные деревья, образцы которых рекурсивно разделяются [21, 22]. Метод имеет много преимуществ и хорошо справляется с обнаружением аномалий, в частности, он чувствителен к появлению контекстных аномалий, которые можно интерпретировать как технические ошибки фиксации данных или их намеренное искажение. Для обеспечения такого обнаружения необходимо построить массивы «нормальных» и аномальных реализаций, которые в дальнейшем будут анализироваться специалистами. Для повышения качества обнаружения аномалий [23] предлагается на первом этапе использовать изолирующий лес для фильтрации данных (формирования групп «нормальных» и аномальных наблюдений), затем анализировать все аномалии и выделять среди них контекстные (наблюдения, которые принципиально возможны, но не типичны по сравнению с соседними данными), а затем, применяя нейросетевой классификатор, строить модель, способную находить выбросы для новых исходных данных:

$$Y^{(n)} = \Psi^{(n)}\Psi^{(n-1)} \dots \Psi^{(1)}X, \quad (3)$$

где $Y^{(n)}$ – выход n -слойной нейронной сети (значение, характеризующее принадлежность наблюдения к множеству выбросов); X – вектор входных факторов, $\Psi^{(1)}, \Psi^{(2)}, \dots, \Psi^{(n)}$ – функции активации слоев нейронной сети.

Задача редукции нейросетевой модели

Следует отметить, что при большом количестве входов в нейронную сеть (3), такой подход является затратным с точки зрения использования вычислительных ресурсов.

Ранее синтезирован алгоритм анализа чувствительности модели по факторам, позволяющий провести редукцию нейросетевой модели с минимальным понижением ее точности [24]. В основу подхода положены методы анализа конечных изменений с применением теоремы Лагранжа о промежуточной точке. В качестве мер чувствительности выхода нейронной сети на изменения ее входов используются так называемые факторные нагрузки A_{x_i} , которые вычисляются из модели

$$\Delta Y^{(n)} = \sum_{i=1}^d \frac{\partial Y^{(n)}}{\partial x_i} \left(\dots, x_i^{(0)} + \alpha \Delta x_i, \dots \right) \Delta x_i = A_{x_1} \Delta x_1 + \dots + A_{x_d} \Delta x_d. \quad (5)$$

Полученные результаты нагрузок (5) усредняются, и на основе их анализа принимается решение о сокращении числа входов модели.

Процедура вычисления факторных нагрузок по модели (5) проводится столько раз, сколько существует конечных приращений в анализируемом наборе данных. Таким образом, возникает задача нахождения устойчивой к выбросам оценки среднего значения. В качестве такой оценки предлагается использовать взвешенное среднее Тьюки, алгоритм построения которого носит итерационный характер и включает в себя следующие шаги:

Шаг 1. Вычислить среднее значение выборки (в качестве начального приближения используется медиана).

Шаг 2. Определить расстояния от вычисленного среднего на шаге 1 до каждого элемента выборки. В соответствии с этими расстояниями элементам выборки присваиваются различные веса, с учетом которых среднее значение пересчитывается.

Пусть $A_{x_i} = \{A_{x_i}^1, \dots, A_{x_i}^n\}$ – выборка из рассчитанных факторных нагрузок для входа x_i ; $M_{A_{x_i}}$ – медиана выборки A_{x_i} ; S – медиана выборки $\{|A_{x_i}^1 - M_{A_{x_i}}|, \dots, |A_{x_i}^n - M_{A_{x_i}}|\}$ (абсолютное отклонение среднего). Для каждого элемента $A_{x_i}^k$ ($k = 1, \dots, n$) выборки A_{x_i} вычисляется отклонение от среднего

$$u_k = \frac{A_{x_i}^k - M_{A_{x_i}}}{cS + \xi},$$

где c – параметр, определяющий, насколько оценка чувствительна к выбросам; ξ – малая величина, основное назначение которой – исключить возможность деления на ноль.

Для нахождения веса каждого наблюдения выборки используется биквадратная функция вида

$$w(u) = \begin{cases} (1 - u^2)^2, & |u| \leq 1, \\ 0, & |u| > 1. \end{cases}$$

Оценка взвешенного Тьюки

$$T_{A_{x_i}} = \frac{\sum_{k=1}^n w(u_k) A_{x_i}}{\sum_{k=1}^n w(u_k)}.$$

Помимо точечной оценки среднего находят интервал для построения величины при помощи приближения распределением Стьюдента.

Симметричный доверительный интервал уровня $(1 - \alpha)$, %, задается формулой

$$T_{A_{x_i}} \pm t_{df}^{(1-\alpha/2)} \frac{S_{A_{x_i}}}{\sqrt{n}};$$

$$S_{A_{x_i}} = \sqrt{n} \frac{\sqrt{\sum_{|u_k| \leq 1} (A_{x_i}^k - T_{x_i})^2 (1 - u_k^2)^4}}{\left| \sum_{|u_k| \leq 1} (1 - u_k^2)(1 - 5u_k^2) \right|},$$

где $t_{df}^{(1-\alpha/2)}$ – $(1 - \alpha/2)$ – квантиль распределения Стьюдента с числом степеней свободы $df = \max(0, 7(n - 1), 1)$.

Схема описываемой системы представлена на рис. 1. Вначале данные, полученные из информационной системы, подвергаются преобразованию в необходимый формат. Затем полученные данные обрабатываются при помощи алгоритма изолирующего леса в целях вычисления показателя аномальности для каждой реализации. Набор входов модели дополняется полученными результатами. Среди существующих алгоритмов оценки аномальности наблюдения выбор конкретного алгоритма объясняется, во-первых, большей эффективностью изолирующего дерева, по сравнению с большинством подходов; во-вторых, отсутствием параметров настройки. Построенная модель с дополнительным входом в свою очередь подвергается анализу чувствительности по входам для последующей редукции. По выбранным наиболее значимым входам происходит структурная и параметрическая идентификация новой нейросетевой модели классификатора с сокращенным числом входных переменных. На последнем этапе проводится выявление аномальных записей в новых массивах данных об оказании медицинской помощи.

Схема информационных потоков между модулями системы представлена на рис. 2, интерфейс разработанного программного обеспечения – рис. 3.

В качестве входных факторов взяты отдельные статистические показатели из информационной системы персонифицированного учета медицинской помощи Липецкого областного фонда обязательного медицинского страхования (Липецкая область, Россия). Период фиксации: 01 – 30 апреля 2018 г. (570 111 случаев). На основе экспертного анализа выбраны показатели, однозначно характеризующие каждый случай оказания медицинской помощи, – 14 различных значений, таких как показатели, однозначно идентифицирующие пациента (например, номер полиса ОМС и возраст), оказание медицинской помощи (например, код, по которому можно определить, какая организация здравоохранения оказывает медицинскую помощь, специализация медицинского персонала, оказавшего анализируемый вид помощи), особенности конкретного случая (например, основной диагноз на первом, втором, третьем этапах госпитализации, дополнительный диагноз, длительность госпитализации) и т.п. Выбранные показатели были поданы на вход алгоритма изолирующего леса, полученные оценки аномальности для каждого наблюдения затем использовали в качестве входного фактора.

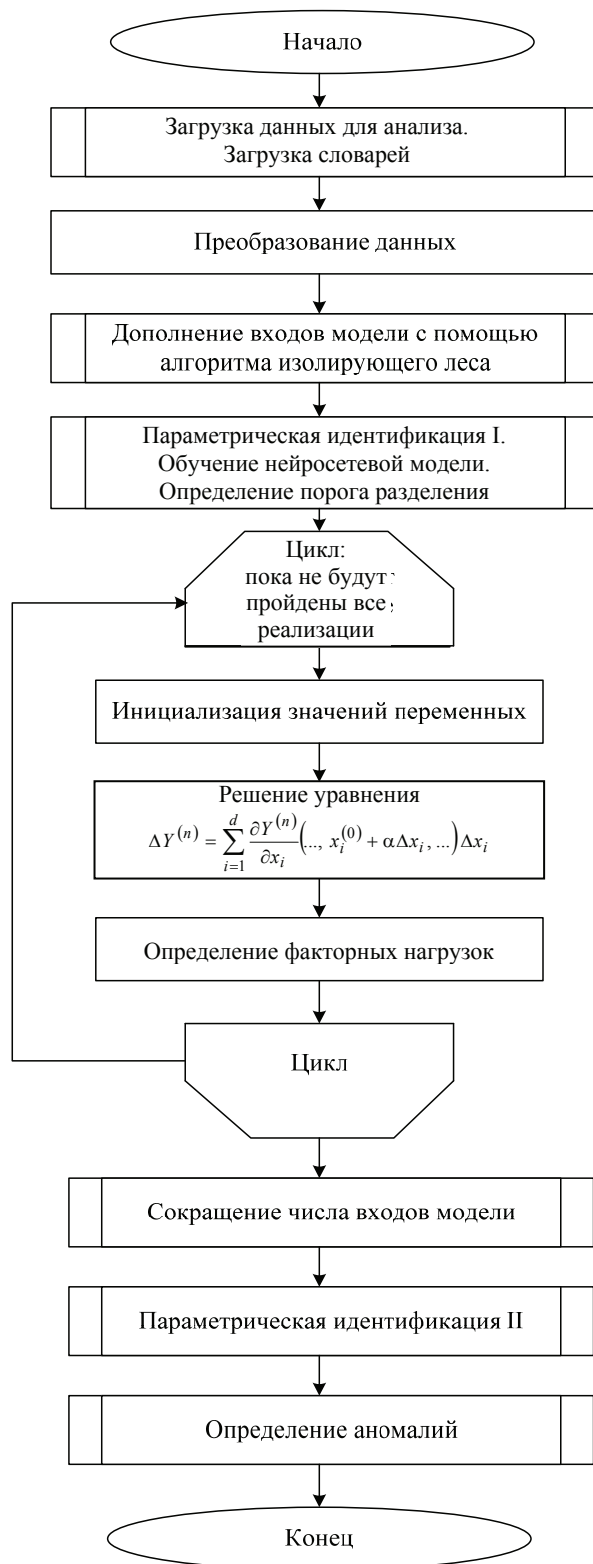


Рис. 1. Схема реализации модели поиска аномальных наблюдений

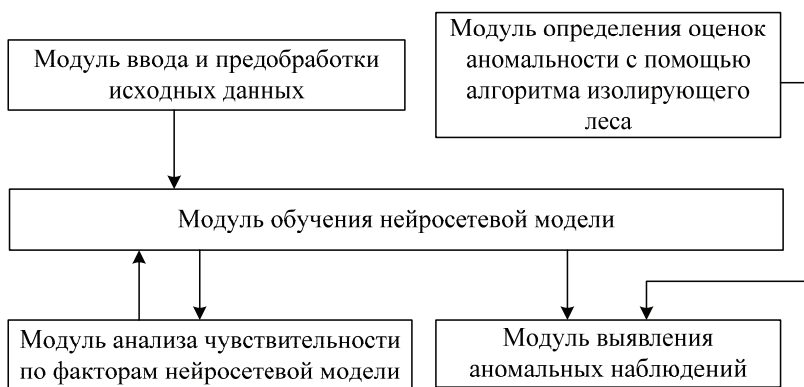


Рис. 2. Схема информационных потоков между модулями системы

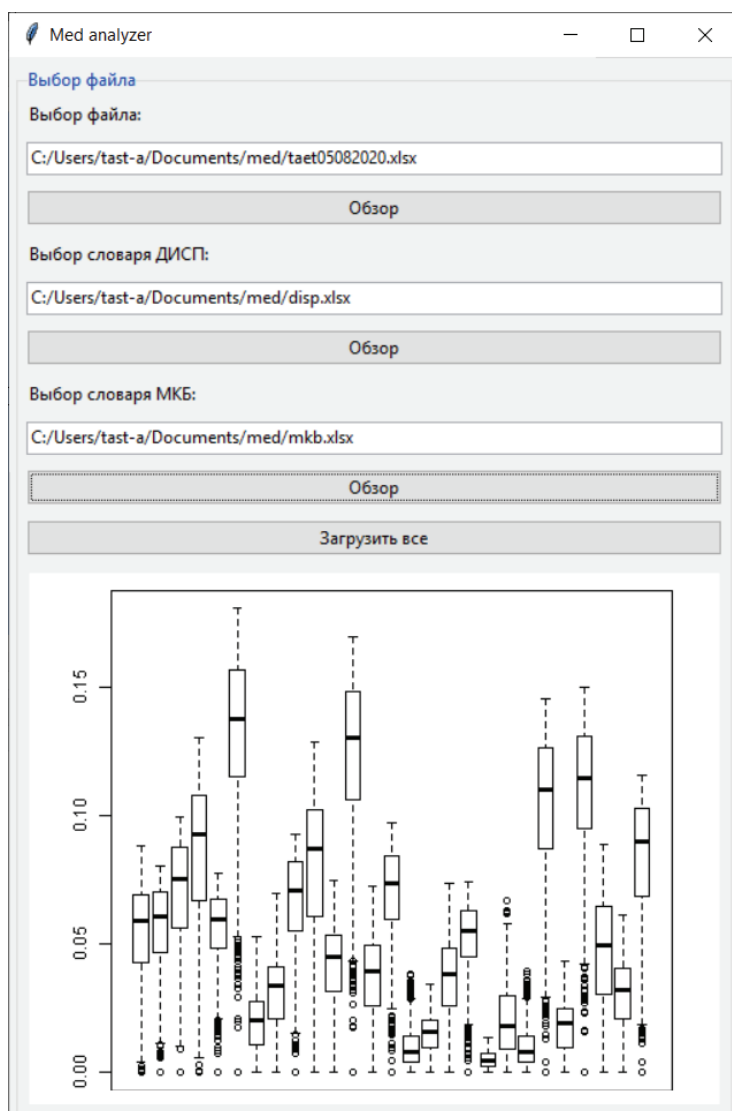


Рис. 3. Интерфейс модуля загрузки и преобработки данных

Вычислительный эксперимент

Была использована нейросетевая модель (3), имеющая структуру

$$y_k = \psi_1 \left(b_0 + \sum_{j=1}^3 w_j \psi_2 \left(b_1 + \sum_{m=1}^{15} w_{jm} x_m \right) \right), \quad (6)$$

где y_k – модельные значения; X – вектор факторов системы; w_j и w_{jm} – весовые коэффициенты; b_0 и b_1 – свободные коэффициенты нейронов выходного и скрытого слоев соответственно; $\psi_1(\text{net}) = \psi_2(\text{net}) = 1/(1+\exp(-\text{net}))$ – логистические функции активации. Точность предсказания моделью (6) аномалий в новом наборе данных составила 92,37 %.

На рисунке 4 представлена гистограмма с результатами анализа чувствительности построенной модели, демонстрирующими значимость входов модели. Отметим, что проведенные ранее исследования показывают непротиворечивость полученных результатов. Но в отличие от известных подходов (например, использование коэффициентов чувствительности Соболя), предложенный подход не основан на аппроксимации статистических параметров исследуемой структуры и, в отличие от стратегии Гарсона (подход, используемый для оценки чувствительности входов нейронных сетей), оперирует как параметрами, так и факторами исследуемой модели.

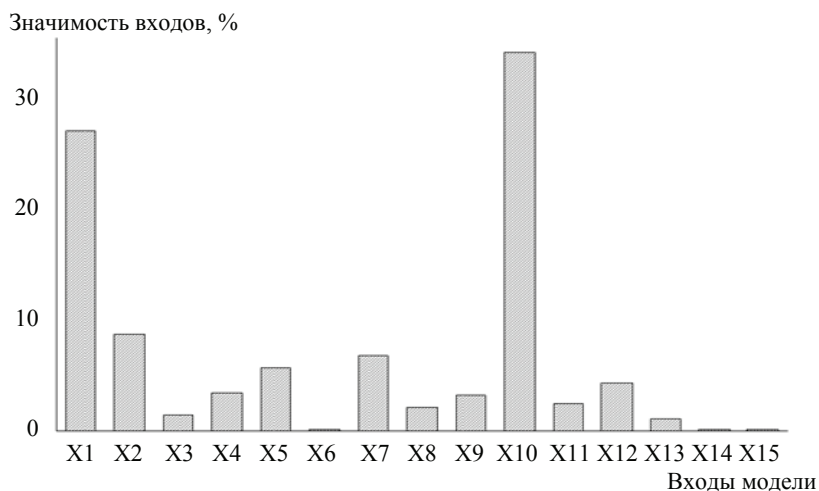


Рис. 4. Результаты анализа чувствительности модели (6)

Заключение

Разработанный программный комплекс реализует предложенные модели и подходы, направленные на решение задачи определения аномальных реализаций в массивах данных об оказании медицинской помощи населению. Актуальность этих проблем подтверждается постоянным увеличением расходов на проведение экспертных проверок в указанной области. Предлагаемая система способна проводить автоматическую проверку поступающей новой информации, а также имеет возможность адаптации к новым условиям функционирования. Открытым остается вопрос определения порогового значения важности входа рассматриваемой модели, однако в каждом конкретном случае данный вопрос решается путем экспертной оценки.

Список литературы

1. Abu-Nasser, B. Medical Expert Systems Survey / B. Abu-Nasser // International Journal of Engineering and Information Systems (IJEAIS). – 2017. – Vol. 1, Issue 7. – P. 218 – 224.
2. Singla, J. Medical Expert Systems for Diagnosis of Various Diseases / J. Singla, D. Grover, A. Bhandari // International Journal of Computer Applications. – 2014. – Vol. 93, Issue 7. – P. 36 – 43.
3. Корневский, Н. А. Использование нечеткой логики принятия решений для медицинских экспертных систем / Н. А. Корневский // Медицинская техника. – 2015. – Т. 49, № 1. – С. 33 – 35.
4. An Expert System for Diabetes Prediction Using Auto Tuned Multi-Layer Perceptron / M. Jahangir, H. Afzal, M. Ahmed [et al.] // Intelligent systems conference (IntelliSys), 7–8 September 2017, London. – IEEE, 2018. – P. 722 – 728.
5. Body-Wide Hierarchical Fuzzy Modeling, Recognition, and Delineation of Anatomy in Medical Images / J. K. Udupa, D. Odhner, L. Zhao [et al.] // Medical Image Analysis. – 2014. – Vol. 18, Issue 5. – P. 752 – 771.
6. Xue, W. Text Detection and Recognition for Images of Medical Laboratory Reports with a Deep Learning Approach / W. Xue, Q. Li, Q. Xue // IEEE Access. – 2019. – Vol. 8. – P. 407 – 416.
7. Active Deep Neural Network Features Selection for Segmentation and Recognition of Brain Tumors Using MRI Images / M. I. Sharif, J. P. Li, M. A. Khan, M. A. Saleem // Pattern Recognition Letters. – 2020. – Vol. 129. – P. 181 – 189.
8. Доронищева, А. В. Методы распознавания медицинских изображений для задач компьютерной автоматизированной диагностики. – Текст : электронный / А. В. Доронищева, С. З. Савин // Современные проблемы науки и образования. – 2014. – № 4. – URL : https://www.elibrary.ru/download/elibrary_22285960_75182931.pdf (дата обращения: 24.08.2021).
9. Parks, C. L. Automated Facial Recognition of Computed Tomography-Derived Facial Images: Patient Privacy Implications / C. L. Parks, K. L. Monson // Journal of Digital Imaging. – 2017. – Vol. 30, Issue 2. – P. 204 – 214.
10. Марухина, О. В. Применение методов Data Mining для выявления скрытых закономерностей в задачах анализа медицинских данных / О. В. Марухина, Е. Е. Мокина, Е. В. Берестнева // Фундаментальные исследования. – 2015. – № 4. – С. 107 – 113.
11. Минжасова, А. И. Статистический анализ медицинских данных / А. И. Минжасова // Прикладная математика и фундаментальная информатика. – 2015. – № 2. – С. 193 – 198.
12. Щеглевых, Р. В. Исследование нейросетевой модели обнаружения аномальных наблюдений в массивах данных / Р. В. Щеглевых, А. С. Сыроев // Прикладная математика и вопросы управления. – 2021. – № 1. – С. 23 – 40. doi: 10.15593/2499-9873/2021.1.02
13. Agrawal, Sh. Survey on Anomaly Detection using Data Mining Techniques / Sh. Agrawal, J. Agrawal // Procedia Computer Science. – 2015. – Vol. 60. – P. 708 – 713.
14. Amor, N. B. Naive Bayes vs Decision Trees in Intrusion Detection Systems / N. B. Amor, S. Benferhat, Z. Elouedi // Proceedings of the ACM Symposium on Applied Computing, 14 – 17 March 2004, Nicosia, Cyprus. – IEEE, 2004. – P. 420 – 424.
15. Kaur, N. Survey Paper on Data Mining Techniques of Intrusion Detection / N. Kaur // International Journal of Science, Engineering and Technology Research. – 2013. – Vol. 2, Issue 4. – P. 799 – 804.
16. Survey of Fraud Detection Techniques / Y. Kou, C. T. Lu, S. Sirwongwattana, Y. P. Huang // Proceedings of the IEEE International Conference Networking, Sensing and Control, 21 – 23 March 2004, Taipei, Taiwan. – IEEE, 2004. – Vol. 2. – P. 749 – 754.

17. Bramer, M. Estimating the Predictive Accuracy of a Classifier / M. Bramer // Principles of Data Mining. – London : Springer-Verlag London, 2020. – P. 79 – 92.
 18. Sawade, C. Active Estimation of F-Measures / C. Sawade, N. Landwehr, T. Scheffer // Advances in Neural Information Processing Systems : 24th Annual Conference on Neural Information Processing Systems, Proceedings of a meeting held 6 – 9 December 2010, Vancouver, British Columbia, Canada. – Curran Associates, Inc., 2010. – Vol. 2. – P. 2083 – 2091.
 19. Obuchowski, N. A. Nonparametric Analysis of Clustered ROC Curve Data / N. A. Obuchowski // Biometrics. – 1997. – Vol. 53, No. 2. – P. 567 – 578.
 20. Parmar, J. D. Anomaly Detection in Data Mining: A Review / J. D. Parmar, J. T. Patel // International Journal of Advanced Research in Computer Science and Software Engineering. – 2017. – Vol. 7, Issue 4. – P. 32 – 40.
 21. Liu, F. T. Isolation Forest / F. T. Liu, K. M. Ting, Z. H. Zhou // 2008 Eight IEEE International Conference on Data Mining (ICDM), 15 – 19 December 2008, Pisa, Italy. – Los Alamitos, 2008. – Art. 10472172. – P. 413 – 422. doi: 10.1109/ICDM.2008.17
 22. Liu, F. T. Isolation-Based Anomaly Detection / F. T. Liu, K. M. Ting, Z. H. Zhou // ACM Transactions on Knowledge Discovery from Data. – 2012. – Vol. 6, Issue 1. – P. 1 – 39. doi: 10.1145/2133360.2133363
 23. Щеглеватых, Р. В. Математическая модель обнаружения аномальных наблюдений с использованием анализа чувствительности нейронной сети / Р. В. Щеглеватых, А. С. Сысоев // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8, № 1 (28). – С. 14–15. doi: 0.26102/2310-6018/2020.28
 24. Sensitivity Analysis of Neural Network Models: Applying Methods of Analysis of Finite Fluctuations / A. Sysoev, A. Ciurlia, R. Sheglevatykh, S. Blyumin // Periodica Polytechnica Electrical Engineering and Computer Science. – 2019. – Vol. 63, Issue 4. – P. 306 – 311. doi: 10.3311/PPee.14654
-

A System for Detecting Anomalous Observations in Data on Healthcare Services

S. L. Blyumin, R. V. Scheglevatykh, A. A. Naydenov, A. S. Sysoev

*Department of Applied Mathematics, sysoev_as@stu.lipetsk.ru;
Lipetsk State Technical University, Lipetsk, Russia*

Keywords: analysis of final changes; sensitivity analysis; abnormal observations; classification; math modeling; search for anomalies.

Abstract: A description of the mathematical model of a neural network classifier of data on healthcare in the institutions of the Lipetsk region is given in order to identify atypical (abnormal) records. Anomaly detection refers to the problem of finding data that is inconsistent with some expected process behavior or metric occurring in the system. Due to the large number of inputs to the neural network model, the time it takes to process the incoming information also increases. To assess what factors should be transmitted to the input of the neural network classifier, an approach to the reduction of the neural network model based on sensitivity analysis is proposed. The description of a set of software tools for solving the problem is presented.

References

1. Abu-Nasser B. Medical Expert Systems Survey, *International Journal of Engineering and Information Systems (IJEAIS)*, 2017, vol. 1, issue 7, pp. 218-224.

2. Singla J., Grover D., Bhandari A. Medical Expert Systems for Diagnosis of Various Diseases, *International Journal of Computer Applications*, 2014, vol. 93, issue 7, pp. 36-43.
3. Korenevskiy N.A. [The use of fuzzy decision-making logic for medical expert systems], *Meditsinskaya tekhnika* [Medical technology], 2015, vol. 49, no. 1, pp. 33-35. (In Russ., abstract in Eng.)
4. Jahangir M., Afzal H., Ahmed M. [et al.] Intelligent systems conference (IntelliSys), 7-8 September, 2017, London, IEEE, 2018, pp. 722-728.
5. Udupa J.K., Odhner D., Zhao L. [et al.] Body-Wide Hierarchical Fuzzy Modeling, Recognition, and Delineation of Anatomy in Medical Images, *Medical Image Analysis*, 2014, vol. 18, issue 5, pp. 752-771.
6. Xue W., Li Q., Xue Q. Text Detection and Recognition for Images of Medical Laboratory Reports with a Deep Learning Approach, *IEEE Access*, 2019, vol. 8, pp. 407-416.
7. Sharif M.I., Li J.P., Khan M.A., Saleem M.A. Active Deep Neural Network Features Selection for Segmentation and Recognition of Brain Tumors Using MRI Images, *Pattern Recognition Letters*, 2020, vol. 129, pp. 181-189.
8. https://www.elibrary.ru/download/elibrary_22285960_75182931.pdf (accessed 24 August 2021).
9. Parks C.L., Monson K.L. Automated Facial Recognition of Computed Tomography-Derived Facial Images: Patient Privacy Implications, *Journal of Digital Imaging*, 2017, vol. 30, issue 2, pp. 204-214.
10. Marukhina O.V., Mokina Ye.Ye., Berestneva Ye.V. [Application of Data Mining Methods to Reveal Hidden Patterns in the Problems of Medical Data Analysis], *Fundamental'nyye issledovaniya* [Fundamental Research], 2015, no. 4, pp. 107-113. (In Russ., abstract in Eng.)
11. Minzhasova A.I. [Statistical analysis of medical data], *Prikladnaya matematika i fundamental'naya informatika* [Applied mathematics and fundamental informatics], 2015, no. 2, pp. 193-198. (In Russ.)
12. Shcheglevatykh R.V., Sysoyev A.S. [Research of a neural network model for detecting anomalous observations in data arrays], *Prikladnaya matematika i voprosy upravleniya* [Applied mathematics and control issues], 2021, no. 1, pp. 23-40, doi: 10.15593/2499-9873/2021.1.02 (In Russ., abstract in Eng.)
13. Agrawal Sh., Agrawal J. Survey on Anomaly Detection using Data Mining Techniques, *Procedia Computer Science*, 2015, vol. 60, pp. 708-713.
14. Amor N.B., Benferhat S., Elouedi Z. Naive Bayes vs Decision Trees in Intrusion Detection Systems, Proceedings of the ACM Symposium on Applied Computing, 14 - 17 March, 2004, Nicosia, Cyprus, IEEE, 2004, pp. 420-424.
15. Kaur N. Survey Paper on Data Mining Techniques of Intrusion Detection, *International Journal of Science, Engineering and Technology Research*, 2013, vol. 2, issue 4, pp. 799-804.
16. Kou Y., Lu C.T., Sirwongwattana S., Huang Y.P. Survey of Fraud Detection Techniques, Proceedings of the IEEE International Conference Networking, Sensing and Control, 21 - 23 March, 2004, Taipei, Taiwan, IEEE, 2004, vol. 2, pp. 749-754
17. Bramer M. *Principles of Data Mining*, London: Springer-Verlag London, 2020, pp. 79-92.
18. Sawade C., Landwehr N., Scheffer T. Advances in Neural Information Processing Systems, 24th Annual Conference on Neural Information Processing Systems, Proceedings of a meeting held 6 - 9 December, 2010, Vancouver, British Columbia, Canada, Curran Associates, Inc., 2010, vol. 2, pp. 2083-2091.
19. Obuchowski N.A. Nonparametric Analysis of Clustered ROC Curve Data, *Biometrics*, 1997, vol. 53, no. 2, pp. 567-578.
20. Parmar J.D., Patel J.T. Anomaly Detection in Data Mining: A Review, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2017, vol. 7, issue 4, pp. 32-40.

21. Liu F.T., Ting K.M., Zhou Z.H. 2008 Eight IEEE International Conference on Data Mining (ICDM), 15 - 19 December, 2008, Pisa, Italy, Los Alamitos, 2008, art. 10472172, pp. 413-422, doi: 10.1109/ICDM.2008.17

22. Liu F.T., Ting K.M., Zhou Z.H. Isolation-Based Anomaly Detection, *ACM Transactions on Knowledge Discovery from Data*, 2012, vol. 6, issue 1, pp. 1-39, doi: 10.1145/2133360.2133363

23. Sheghevatykh R.V., Sysoyev A.S. [A mathematical model for detecting anomalous observations using sensitivity analysis of a neural network], *Modelirovaniye, optimizatsiya i informatsionnyye tekhnologii* [Modeling, optimization and information technologies], 2020, vol. 8, no. 1 (28), pp. 14-15, doi: 0.26102/2310-6018/2020.28 (In Russ., abstract in Eng.)

24. Sysoev A., Ciurlia A., Sheghevatykh R., Blyumin S. Sensitivity Analysis of Neural Network Models: Applying Methods of Analysis of Finite Fluctuations, *Periodica Polytechnica Electrical Engineering and Computer Science*, 2019, vol. 63, issue 4, pp. 306-311, doi: 10.3311/PPee.14654

Erkennungssystem für anomale Beobachtungen in Daten zur medizinischen Versorgung der Bevölkerung

Zusammenfassung: Es ist die Beschreibung des mathematischen Modells des neuronalen Netzklassifizierers von Daten über die medizinische Versorgung in den Einrichtungen der Region Lipetsk gegeben, um atypische (anormale) Aufzeichnungen zu identifizieren. Die Erkennung von Anomalien bezieht sich auf das Problem, Daten zu finden, die nicht mit einem erwarteten Prozessverhalten oder einer im System auftretenden Metrik übereinstimmen. Aufgrund der großen Anzahl von Eingaben in das neuronale Netzmodell erhöht sich auch die Zeit, die zur Verarbeitung der eingehenden Informationen benötigt wird. Um zu beurteilen, welche Faktoren auf die Eingabe des neuronalen Netzklassifikators übertragen werden sollten, wird ein Ansatz zur Reduktion des neuronalen Netzmodells basierend auf einer Sensitivitätsanalyse vorgeschlagen. Die Beschreibung einer Reihe von Softwarewerkzeugen zur Lösung des Problems ist vorgestellt.

Système de la détection des observations anormales dans les données sur les soins médicaux à la population

Résumé: Est donnée la description du modèle mathématique de la classe neuroréseau des données sur les faits de la prestation des soins médicaux dans les établissements de la région de Lipetsk afin d'identifier les enregistrements atypiques (anormaux). Le phénomène d'anomalie se réfère au problème de la détection des données qui ne correspondent pas à un comportement attendu d'un processus ou à un indicateur qui se produit dans le système. En raison du grand nombre d'entrées du modèle de réseau neuronal, le temps nécessaire pour traiter les informations entrantes augmente également. Pour évaluer les facteurs à transmettre à l'entrée du classificateur de neurones, est proposée une approche pour la réduction du modèle neuroréseau basée sur une analyse de sensibilité. Est présentée une description de l'ensemble des outils logiciels pour résoudre le problème posé.

Авторы: *Блюмин Семен Львович* – доктор физико-математических наук, профессор кафедры прикладной математики; *Щеглеватых Роман Вячеславович* – соискатель кафедры прикладной математики; *Найденев Антон Александрович* – магистрант; *Сысоев Антон Сергеевич* – кандидат технических наук, доцент кафедры прикладной математики, ФГБОУ ВО «Липецкий государственный технический университет», г. Липецк, Россия.

Рецензент: *Фролов Сергей Владимирович* – доктор технических наук, профессор, заведующий кафедрой «Биомедицинская техника», ФГБОУ ВО «ТГТУ», г. Тамбов, Россия.