

5. X. Xie and G. A. Beni, "Validity Measure for Fuzzy Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 13, pp. 841-847, 1991.

## ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ АНАЛИЗА МЕДИЦИНСКИХ ДАННЫХ В ОНЛАЙН-РЕЖИМЕ

Бражникова Е.Н., Перова И.Г.

*Харьковский национальный университет радиоэлектроники, 61166, г. Харьков, пр. Науки, 14, email:  
yelyzaveta.brazhnykova@nure.ua, rikywenok@gmail.com*

На сегодняшний день методы Medical Data Mining [1-3] являются широко распространенными в области медицины. Среди задач, возникающих в рамках Medical Data Mining особое место занимает проблема диагностики состояния организма по множеству показателей.

Именно поэтому работа посвящена разработке и исследованию адаптивных гибридных нейро-фаззи систем для информационной технологии медицинского диагностирования в режиме онлайн. Информационная технология состоит из нескольких этапов. На первом этапе признаки одного пациента  $(x_1, x_2, \dots, x_n)^T$  и диагноз  $d$  поступают на вход системы препроцессинга данных, после которого мы получаем предобработанные данные (нормированные, центрированы, кодированные), которые могут быть использованы в качестве входных для информационной технологии. Таким образом на ее вход поступают признаки пациентов в последовательном режиме и их диагнозы (при наличии).

Дальнейшая работа информационной технологии зависит от того известен ли диагноз и если известен, то репрезентативная выборка данных или нет (то есть от значения, содержащегося в  $d$ ).

В ситуациях, когда выборка является репрезентативной информационная технология анализа медицинских данных переходит в режим контролируемого обучения, который может быть реализован путем использования трех адаптивных нейро-фаззи систем. Первой является адаптивная диагностическая нейро-фаззи система реального времени DNFS [4], что имеет свойства менять свою архитектуру при необходимости, то есть эволюционировать. Эволюцией считается изменение архитектуры системы, а именно количества ее входов и выходов (признаков и диагнозов). В ситуациях, когда в процессе диагностирования у пациента изменяется количество симптомов (признаков) или появляется новый диагноз архитектура DNFS меняется с помощью адаптации-эволюции, в конечном итоге превращает ее в систему DNFS<sub>n+1,m+1</sub> [5]. Предложенная глубинная гибридная система вычислительного интеллекта с архитектурной адаптацией для нечеткого медицинского диагностирования DNFS<sub>n+1,m+1</sub> характеризуется высоким быстродействием за счет оптимальных алгоритмов обучения и возможностью продолжать процесс диагностирования без переобучения всей архитектуры системы. Вторым вариантом является адаптивное быстрое диагностирование с помощью многомерного нео-фаззи нейрона [6], что является одной из самых быстрых систем диагностирования за счет использования оптимального алгоритма обучения. Третьим вариантом является адаптивная вейвлет диагностическая нейро-фаззи система реального времени [7], которая имеет высокую точность за счет использования модификации экспоненциально взвешенного рекуррентного метода наименьших квадратов.

В ситуациях, когда выборка не является репрезентативной информационная технология анализа медицинских данных переходит в режим активного обучения и ассоциации, то есть переключается между режимами обучения-самообучения при необходимости. В этом режиме диагноз лишь некоторых пациентов является известным, но при этом количество этих известных диагнозов недостаточна для обучения систем, основанных на контролируемом обучении (нерепрезентативная обучающая выборка). Первым вариантом является использование метода нечеткой классификации данных медико-биологических исследований, работает в условиях дефицита информации и способен проводить классификацию на экстра-малом количестве известных диагнозов [8]. Вторым вариантом является быстрое медицинское диагностирование с помощью нейро-фаззи автоассоциативной памяти [9], которая способна проводить ассоциации по признакам нескольких пациентов, чтобы выявить степень близости пациента с неизвестным диагнозом к нему. Третьим вариантом является система онлайн нейро-фаззи медицинского диагностирования с активным обучением [10], что автоматически переключается между режимами контролируемого обучения и самообучения в зависимости от того размечены данные попадают на ее вход или неразмеченную.

В ситуациях, когда выборка является не размеченной, то есть диагнозы ни одного из пациентов неизвестны информационная технология анализа медицинских данных переходит в режим самообучения. Первой системой в рамках такого режима является модифицированный метод нечеткой кластеризации медицинских данных [11], позволяет проводить последовательное диагностирование в режиме самообучения с использованием кластеров-диагнозов произвольной формы. Еще одна система в рамках

режима самообучения используется в ситуациях когда медицинские признаки могут содержать аномальные выбросы, помехи и другие артефакты. Для этого используется метод адаптивной робастной нечеткой вероятностной кластеризации на основе манхэттенской метрики [12]. Если количество кластеров-диагнозов заранее неизвестно и существует вероятность того, что их количество будет сопоставимо с количеством данных, то следует использовать нечеткую кластеризацию на основе ассоциативно-кластеризующей нейро-фаззи модели [13], в которой на первом этапе все данные распределяются между вершинами гиперкуба, а потом с помощью функции принадлежности специальной формы некоторые вершины могут быть объединены в одну с расчетом значения этой функции.

Для оценки работы всех предложенных в рамках информационной технологии систем следует ввести критерий эффективности  $PerCr$ , который состоит из точности работы системы и времени обработки данных одного пациента –  $Time$ . Но поскольку точность является параметром, который максимизируется, а время обработки – минимизируется, то для корректного формирования критерия эффективности вместо точности следует выбрать параметр ошибки системы  $Fault$ :

$$PerCr = \lambda_1 Fault + \lambda_2 Time, \quad (1)$$

где  $\lambda_1, \lambda_2$  – весовые коэффициенты, избираемых при условии, что  $\sum \lambda = 1$ .

Для возможностей сравнения критериев между собой, а также адекватного его нормирования следует ввести пределы измерений параметров ошибки при диагностировании ( $Fault$ ) и времени обработки одного пациента ( $Time$ ). Ошибка работы системы должна быть нормирована от 0 до 1, таким образом, чтобы нулевому значению соответствовало 0% неверно поставленных диагнозов, а единице – 100% неверно поставленных системой диагнозов. Время обработки одного пациента следует рассчитывать в зависимости от параметров компьютера, на котором проводится эксперимент и примененного программного обеспечения (колеблется от 15 мкс до 1 мс). Таким образом, критерий (1) является таким, который минимизируется и при условии  $\lambda_1 = \lambda_2 = 0,5$  лучшей является ситуация, когда его значение равно нулю – нулевая ошибка работы системы и нулевому времени обработки данных одного пациента. Наихудшим значением критерия является значение единицы, соответствует ситуации 100 процентной ошибки работы системы и времени обработки, равной 1 мс. Исходя из вышеприведенного и проведенных экспериментальных исследований можно сделать вывод, что для задач диагностирования оптимальным является значение критерия эффективности в пределах от 0 до 0,25, приемлемым – от 0.25 до 1 при условии контроля точности диагностирования.

#### ЛІТЕРАТУРА

1. de Oliveira, J., Pedrycz, W. *Advances in Fuzzy Clustering and its Applications*. John Wiley&Sons Ltd. – 2007., 454p.
2. Berka, P., Rauch, J., & Zighed, D. *Data mining and medical knowledge management cases and applications*. New-York: Hershey. – 2009, 440p.
3. Giannopoulos, E. *Data mining in medical and biological research*. New York: ITAe. – 2008, 331p.
4. Бодянский Е.В., Винокурова Е.А., Мулеса П.П., Перова И.Г. Диагностирующая нейро-фаззи-система и ее адаптивное обучения в задачах интеллектуальной обработки данных медико-биологических исследований / Системні технології. Регіональний міжвузівський збірник наукових праць. – Випуск 2(91). – Дніпропетровськ, 2014. – с. 125-135.
5. Pliss, I. Perova “Diagnostic Neuro-Fuzzy System and Its Learning in Medical Data Mining Tasks in Conditions of Uncertainty about Numbers of Attributes and Diagnoses” *Automatic Control and Computer Sciences*, 2017, 51(6), pp.391-398. DOI: 10.3103/ S0146411617060062
6. S. M. K. Mahmoud, I. Perova, I. Pliss, Multidimensional neo-fuzzy-neuron for solving medical diagnostics tasks in online-mode // *Journal of Applied Computer Science*, 2017, Vol. 25, № 1, pp. 39-48
7. Ye. Bodyanskiy, I. Perova, O. Vynokurova, I. Izonin “Adaptive Wavelet Diagnostic Neuro-Fuzzy System for Biomedical Tasks” *Proc. of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv-Slavske, Ukraine, February 20 – 24, 2018, pp.299-303
8. Бодянский Е.В., Перова И.Г. Нечеткая классификация данных медико-биологических исследований в условиях дефицита информации. Системи обробки інформації. – X.: ХУПС, 2015. – Вип. 11(136), с.161-163.12.
9. Perova, Ye. Bodyanskiy Fast medical diagnostics using autoassociative neuro-fuzzy memory // *International Journal of Computing*, 16 (1), 2017, 34-40.
10. Perova, Ye. Bodyanskiy “Medical online neuro-fuzzy diagnostics system with active learning” *International Journal of Advances in Computer and Electronics Engineering Volume 2, Issue 7, July 2017*, pp. 1–10.

11. И.Г. Перова Адаптивная обработка данных медико-биологических исследований методами вычислительного интеллекта // Восточно-европейский журнал передовых технологий. – Х.: 2014. – №1(67), с. 24-28.
12. Perova, Ye. Bodyanskiy Adaptive fuzzy clustering based on Manhattan metrics in medical and biological applications // Вісник національного університету "Львівська політехніка" №826, 2015, p. 8-12.
13. Perova, I., Bodyanskiy, Y., Sachenko, A., Karpinski, M., Rudyk, P. Fuzzy clustering of biomedical datasets using BSB-neuro-fuzzy-model // CEUR Workshop Proceedings, Lviv, 2018, pp. 21-28.

## NORMALIZED METHOD FOR THREATS ASSESSING FOR DISTRIBUTED WIRELESS SYSTEMS

**Buriachok V.L.<sup>1</sup>, Sokolov V.Yu.<sup>2</sup>, Skladannyi P.M.<sup>3</sup>**

*Borys Grinchenko Kyiv University, Kyiv, Ukraine*

<sup>1</sup>ORCID: 0000-0002-4055-1494, v.buriachok@kubg.edu.ua

<sup>2</sup>ORCID: 0000-0002-9349-7946, v.sokolov@kubg.edu.ua

<sup>3</sup>ORCID: 0000-0002-7775-6039, p.skladannyi@kubg.edu.ua

The subjective process of obtaining the probability of a threat can be divided into three stages:

- preparatory (the object of research is formed: the set of events and the initial analysis of the properties of this set; one is selected for methods of obtaining subjective probability; the preparation of an expert or a group of experts);
- obtaining grades (using the chosen method; obtaining results in a numerical form, possibly contradictory);
- analysis of the obtained assessments (research results of the survey; clarification of the answers of experts).

Sometimes the third stage is not carried out if the method itself uses axioms of probable distribution, which in itself is close to experts' estimates. Conversely, this stage becomes especially important if results are obtained from a group of experts.

It is also possible to separate two approaches to multi criteria assessment of the efficiency of wireless networks:

- associated with bringing the set of individual performance indicators to a single integral indicator;
- uses methods of the theory of multi-choice choice and decision-making (with a significant number of individual performance indicators, approximately equally important).

The proposed normalized method for assessing the degree of security assurance operates with at least three characteristics, which allows for a comparative analysis of heterogeneous information systems.

The degree of security provides a rough estimate of the effectiveness of the information security system. The method operates with the subjective coefficients of weight  $i^{\text{th}}$  characteristic  $W_i$  and the ball values of each characteristic  $G_i$ , which is determined by expert's estimates. The formula for the degree of security is as follows:

$$S = \frac{1}{N} \sum_{i=1}^N W_i \cdot G_i \quad (1)$$

where  $N$ —amount of the characteristics.

The method has two drawbacks: it is impossible to compare systems with different sets of characteristics and it does not take into account the dependence of the weighting factor and the value of the characteristic of the characteristic itself [1,2]. The author of the paper proposes to use normalized characteristic  $S^*$  to assess the degree of the system security, and at the same time, to consider the subjective factors of the importance of the  $i^{\text{th}}$  characteristic and the ball value of each characteristic, as a function of the characteristics:

$$\begin{cases} W_i = f_W(x_i), \\ G_i = f_G(x_i) \end{cases} \quad (2)$$

where  $f_W$  and  $f_G$ —function of the characteristic  $x_i$ .

The general formula for monotonous  $f_W$  and uncertain function  $f_G$  is as follow:

$$S^* = \frac{1}{N} \sum_{i=1}^N W^*(x_i) \cdot G^*(x_i) \quad (3)$$

where  $W^*(x_i)$ —normalized weighting factor of subjective estimation from  $x_i$ :

$$W^*(x_i) = \left| \frac{f_W(x_i)}{\max[f_W(x_i)]} \right| \quad (4)$$