

Интеллектуальная обработка медицинской информации для использования в экспертной системе

Е. В. Чебаненко¹, Л. А. Денисова²

Омский государственный технический университет

¹evchebanenko@gmail.com, ²denisova@asoiu.com

Аннотация. В статье предложен подход к обработке и анализу медицинской информации для выявления групп пациентов по степени приверженности к терапевтическому вмешательству. Выполнен анализ данных с использованием такой интеллектуальной информационной технологии как метод нечеткой кластеризации, возможности которого раскрываются применительно к слабо формализованной исходной информации. В работе использованы результаты эмпирического исследования, посвященного изучению качества жизни больных гипертонической болезнью и факторов их приверженности к лечению. Производилась оценка приверженности к терапевтическому вмешательству в количественном эквиваленте по трем критериям. Рассмотрены критерии ожидаемой эффективности медицинского сопровождения и модификации образа жизни. Также принимался во внимание критерий ожидаемой эффективности лекарственной терапии. Результатом настоящего исследования является выявление кластеров для категорий пациентов, обладающих схожим типом восприятия терапии в совокупности по этим трем критериям. Кроме того, на основании результатов кластерного анализа получены функции принадлежности пациентов к группам с максимально схожим типом приверженности к лечению. Рассмотрены перспективы разработки советующей экспертной системы для врача-кардиолога при назначении лекарственной терапии.

Ключевые слова: критерий эффективности медицинского сопровождения; приверженность к лечению; метод нечеткой кластеризации; экспертная система

I. ВВЕДЕНИЕ

В настоящее время медицинские экспертные системы, базирующиеся на современных информационных технологиях, находят все более широкое применение, как в медицинских исследованиях, так и в реальной клинической практике [1–3]. В медицинских организациях разрабатываются новые подходы к назначению и сопровождению терапевтических процедур. Собираемые и порождаемые при этом данные требуют быстрой и эффективной обработки.

Следует отметить, что в медицинском сообществе большое внимание уделяется вопросам приверженности пациентов к медицинскому сопровождению как

непосредственно влияющему на качество и результат лечения. Кроме того, является актуальным как построение прогнозирующих моделей заболеваний, так и разработка способов влияния на отношение пациента к применяемой терапии. Также важно создание подходов к оценке эффективности и рациональности медицинского вмешательства [1–3].

В работе рассмотрен подход к обработке и анализу данных о приверженности пациентов к терапевтическому вмешательству посредством применения математического аппарата нечеткой логики [4–14]. Нечеткие интеллектуальные технологии обработки и использования знаний обычно применяются в условиях неопределенности и неполноты знаний об объекте, нечеткости описаний.

Мы используем метод нечеткой кластеризации, возможности которого раскрываются применительно к слабо формализованным исходным данным [6–10, 13–15]. Как известно, задача кластеризации заключается в разбиении обширного набора объектов на отдельные группы, обладающие общими свойствами. Особенностью алгоритма нечеткой кластеризации (по сравнению с четкими методами) является то, что полученные в результате разбиения объектов кластеры соответствуют нечетким множествам. При этом каждый элемент выборки может относиться к нескольким кластерам с определенной степенью принадлежности (в отличие от четкой кластеризации, где элемент принадлежит только одному кластеру).

Преимущество использования метода нечеткой кластеризации заключается в возможности количественной оценки уровня принадлежности пациентов к группам при назначении медицинского сопровождения и в получении информации для дополнительного анализа.

II. ПОСТАНОВКА ЗАДАЧИ

Многочисленные клинические исследования в области терапии больных хроническими заболеваниями демонстрируют, что одним из важных вопросов является выяснение отношения больного к проводимой терапии, ценности получаемых позитивных и негативных эффектов, наконец, к результату лечения [1, 3].

Для выяснения этого вопроса профессором Н. А. Николаевым выполнялось эмпирическое исследование, начатое в 2003 г., посвященное изучению качества жизни больных гипертонической болезнью и факторов их приверженности к лечению [1]. При этом приверженность пациента к постоянной лекарственной терапии и ожидаемая эффективность лечения прогнозировались на основании результатов анкетирования. Индивидуальные особенности приверженности к лечению у больных артериальной гипертензией могут быть выявлены, оценены количественно и использованы для прогноза и мониторинга ожидаемой эффективности лечения.

Будем использовать индексы ожидаемой эффективности, предложенные Н. А. Николаевым в методике [1]. Следуя этой методике можно оценить приверженность к терапевтическому вмешательству в количественном эквиваленте по трем критериям:

- ожидаемая эффективность модификации образа жизни (I_{EUWL});
- ожидаемая эффективность лекарственной терапии (I_{EMT});
- ожидаемая эффективность врачебного сопровождения (I_{EMS}).

Задача настоящего исследования заключается в выявлении категорий пациентов, обладающих максимально схожим типом восприятия терапии в совокупности по этим трем критериям ожидаемой эффективности.

Анализ данных о приверженности пациентов к терапевтическому вмешательству выполнялся посредством применения метода нечеткой кластеризации. Кроме того, на основании результатов кластерного анализа были получены функции принадлежности пациентов к группам с максимально схожим типом восприятия терапии для создания экспертной системы. В заключении работы рассмотрены перспективы разработки советующей экспертной системы для врача-кардиолога при назначении лекарственной терапии.

III. АЛГОРИТМ ВЫЯВЛЕНИЯ ГРУПП ПАЦИЕНТОВ ПО ТИПАМ ВОСПРИЯТИЯ ЛЕЧЕНИЯ

Для решения поставленной задачи выявления групп пациентов, обладающих максимально схожими свойствами восприятия лечения с учетом всех трех критериев, воспользуемся алгоритмом нечетких c -средних. Особенностью алгоритма нечетких c -средних является то, что данные представляются в виде нечетких множеств, и каждый элемент может относиться к нескольким кластерам с определенной степенью принадлежности [10, 13, 14]. Применительно к поставленной задаче кластерная структура представляет собой матрицу нечеткого разбиения пациентов M , где каждая i -я строка ($i = \overline{1, l}$) выражает степень принадлежности i -го пациента к кластеру A_j ($j = \overline{1, c}$), формируемая согласно степени

приверженности по каждому из трех критериев. В формуле обозначены: l – это количество пациентов, а c – число кластеров. Степень принадлежности $\mu_{ij} \in [0, 1]$.

То есть матрица нечеткого разбиения пациентов имеет вид:

$$M = [\mu_{ij}] = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1c} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2c} \\ \dots & \dots & \dots & \dots \\ \mu_{l1} & \mu_{l2} & \dots & \mu_{lc} \end{bmatrix}$$

Рассмотрим подробнее реализованный алгоритм.

Шаг первый. Определяем число кластеров (принимаем $c=3$, первоначально интерпретируем кластеры, как группы по степени приверженности) и весовой коэффициент элементов (принимаем $m=2$, далее этот параметр используется для расчета центров кластеров). Также принимаем условие завершения алгоритма, исходя из желаемой точности решения (принимаем $\varepsilon=1e-5$).

На втором шаге генерируется матрица нечеткого разбиения M , при этом необходимо исходить из следующих соображений. Требуется, чтобы все объекты были распределены по каждому кластеру в соответствии с выражением

$$\sum_{j=1, c} \mu_{ij} = 1, i = \overline{1, l}$$

и, кроме того, ни один кластер не является пустым множеством или содержит все элементы:

$$0 < \sum_{i=1, l} \mu_{ij} < l, j = \overline{1, c}.$$

Шаг третий. Рассчитываем центры кластеров по выражению:

$$C_j = \sum_{i=1, l} ((\mu_{ij})^2 M_i) / \sum_{i=1, l} (\mu_{ij})^2, j = \overline{1, c},$$

где $M_i \in A_j$ – i -й объект j -го кластера.

На четвертом шаге рассчитывается расстояние между объектами из матрицы M и центрами кластеров. Применительно к исследуемым данным каждая строка матрицы содержит индексы приверженности конкретного пациента, каждый столбец матрицы значение индексов I_{EUWL} , I_{EMT} , I_{EMS} , соответственно.

При определении расстояния между объектами мы используем Евклидову метрику, являющуюся геометрическим расстоянием:

$$D_{ij} = \sqrt{\|M_i - C_j\|^2}; j = \overline{1, c}; i = \overline{1, l}$$

На пятом шаге выполняется расчет элементов матрицы нечеткого разбиения пациентов

$$\mu_{ij} = \left(D_{ji}^2 \times \sum_{k=1,c} \frac{1}{D_{ki}^2} \right)^{-1}.$$

На шестом шаге сравниваются матрицы нечеткого разбиения пациентов текущего и предыдущего шага, при условии, если $\|M - M'\| < \varepsilon$, алгоритм завершается, в противном случае выполняется переход к *третьему шагу*. Здесь M' – матрица нечеткого разбиения на предыдущей итерации алгоритма.

Для проведения кластерного анализа мы использовали программную процедуру *fcm* вычислительного инструментария MATLAB / FuzzyLogicToolbox [10–12].

IV. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

А. Исходные данные

Исходные данные сформированы на основе опросника профессора Н.А. Николаева [1], с помощью которого возможно оценить такой качественный параметр как приверженность в количественном эквиваленте. В исследовании рассмотрены критерии ожидаемой эффективности медицинского сопровождения, модификации образа жизни и лекарственной терапии.

Ожидаемая эффективность лечения измерялась количественно. При этом по ее значения в диапазоне 1-1,99 балла свидетельствуют о хорошей эффективности, 2-3,99 балла свидетельствуют об удовлетворительной эффективности, а значение 4 балла и более говорят о неудовлетворительной ожидаемой эффективности вмешательств. Данные для исследуемой группы пациентов частично представлены в таблице 1.

ТАБЛИЦА I ТАБЛИЦА ИНДЕКСОВ ПРИВЕРЖЕННОСТЕЙ К ТЕРАПИИ

Пациенты	I_{EMS}	I_{EUWL}	I_{EMT}
1	1,57	1,24	1,50
2	1,57	1,69	1,19
3	4,59	3,46	2,00
4	1,50	1,38	1,64
...
60	1,36	3,00	1,07

В проведенное исследование было включено 160 пациентов с сердечно-сосудистыми заболеваниями возрастом от 40 до 85 лет.

В. Результаты кластерного анализа

Для проведения кластерного анализа индивидуальные особенности приверженности к лечению пациентов (которые быть выявлены и оценены количественно) использованы для разбиения на группы по ожидаемой эффективности лечения.

Данный подход демонстрирует возможность выявления группы пациентов со схожими критериями, что позволяет осуществлять классификацию в отсутствии предварительных сведений. Кроме того, такой подход позволяет впоследствии строить функции принадлежности пациентов к группам и делать вывод о приверженности новых пациентов/

С. Экстракция функций принадлежности

В результате исследований выявлены три группы пациентов с ярко выраженной принадлежностью к определенному виду лечения. Будем учитывать, что согласно методике [1] из критериев (индексов) ожидаемой эффективности индекс I_{EMT} имеет наибольшее значение по важности, индекс I_{EUWL} – среднее значение, а индекс I_{EMS} – наименьшее значение для принятия решений по выбору терапии. Каждый нечеткий кластер можно отобразить в нечеткое правило [9, 12, 13], что планируется выполнить в дальнейшем при построении экспертной системы. Соответственно величина степени принадлежности каждого значения входной переменной (то есть величина индекса приверженности) будет учитываться при формировании каждого нечеткого правила. На основе кластерного анализа и выявленных центров кластеров получены функции принадлежности термов для пациентов с разной степенью приверженности к лечению (рис. 2). Функции принадлежности получены проецированием степеней принадлежности соответствующего кластера (строка матрицы нечеткого разбиения M) на оси индексов ожидаемой эффективности.

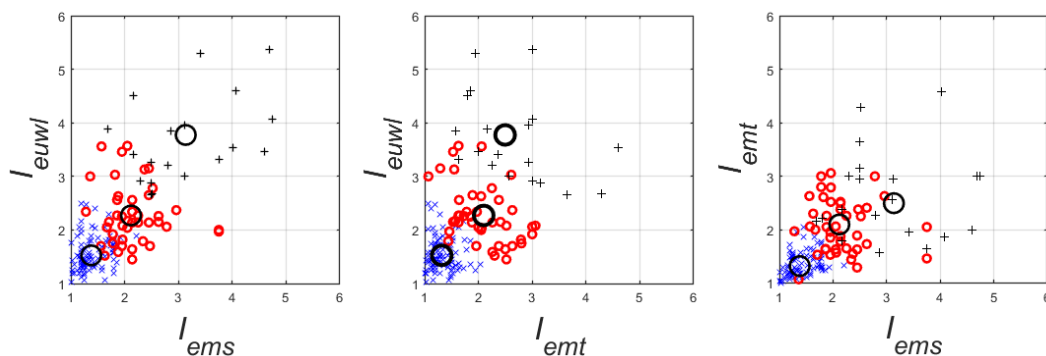


Рис. 1. Результаты кластеризации по принятым критериям. Степень ожидаемой приверженности к терапии: x – высокая, o – удовлетворительная, - - неудовлетворительная; символом O обозначены центры кластеров

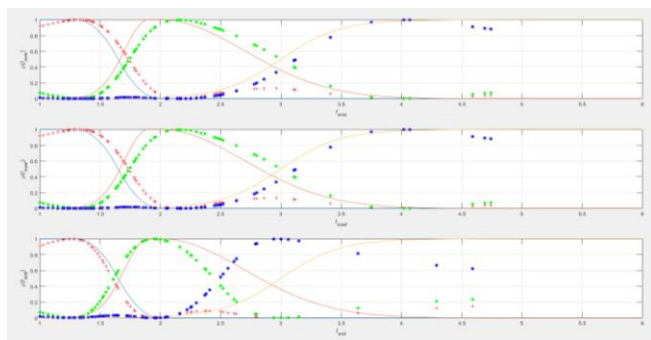


Рис. 2. Функции принадлежности термов для пациентов с разной степенью приверженности к лечению

V. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Предложенная нечеткая система может применяться для оценки уровня приверженности пациентов при первичном обращении, как советующая система поддержки и сопровождения. Кластерный анализ применительно к обработке количественной оценки приверженности пациентов является первоначальным этапом для построения прогнозирующих моделей предлагаемой терапии и прогнозирования исходов и разработки экспертной советующей системы при назначении терапии врачом-кардиологом.

Анализируя результаты кластеризации, следует обратить внимание на следующие группы данных, для которых целесообразно применить дополнительный анализ, и, возможно, в результате получить рекомендации при назначении медицинского сопровождения:

- «спорные» объекты, обладающие равной степенью принадлежности к нескольким кластерам, а также «выбросы», то есть объекты с крайне низкой принадлежностью ко всем имеющимся кластерам;
- объекты, непосредственно принадлежащие кластеру, но при этом с ярко выраженным преобладанием одного либо двух индексов. Например, пациент имеет высокую приверженность к лекарственной терапии, но при этом крайне низкие приверженности по модификации образа жизни и медицинскому сопровождению.

Представляется целесообразным сформулировать и решить задачу по модернизации реализованного метода нечеткой кластеризации, способного отражать специфику исследуемых данных.

VI. ЗАКЛЮЧЕНИЕ

Мы предложили и исследовали метод выявления групп пациентов кардиологического профиля согласно степени приверженности к терапевтическому вмешательству. При проведении исследования мы учитывали то, что индивидуальные особенности приверженности к лечению у больных могут быть выявлены, оценены количественно

и использованы для прогноза и мониторинга ожидаемой эффективности лечения.

В результате исследования были сформированы три группы пациентов по степени приверженности к терапии. Получены сведения о границах значений каждого критерия. Построены функции принадлежности по каждому показателю для дальнейшего использования в экспертной системе.

Следует подчеркнуть, что использование предлагаемого метода обработки медицинской информации совместно с методиками, предназначенными для работы в условиях клинической практики, позволяет разрабатывать интеллектуальные сопровождающие системы для пациентов кардиологического профиля. Это позволяет надеяться, что такая экспертная система значительно повысит эффективность работы врача-кардиолога при назначении лекарственной терапии.

СПИСОК ЛИТЕРАТУРЫ

- [1] Николаев Н.А. Руководство по клиническим исследованиям внутренних болезней. М: Издательский дом Академия Естествознания. 2015.
- [2] Guidelines for the Management of Arterial Hypertension: The Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC) J. Hypertens 25, 2007, pp.105-1187.
- [3] Gallup G., The Gallup poll: Public opinion 1978 (Wilmington, Delaware: Scholarly Resources), 1979, p. XLIV.
- [4] Zadeh L.A., Fuzzy sets Information and Control. 8, 1965, pp 338-53.
- [5] Sugeno M and Yasukawa T 1993 Fuzzy-logic-based approach to qualitative modeling IEEE Transactions on Fuzzy Systems. 1 no 1 pp 7–31.
- [6] Ren M, Liu P, Wang Z and Yi J 2016 A Self-Adaptive Fuzzy c-Means Algorithm for Determining the Optimal Number of Clusters Journal of Computational Intelligence and Neuroscience, pp. 5-8.
- [7] Kotov A. and Krasilnikov N. Clustering of data URL: <http://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf>. (reference date: 05.02.2017), pp. 13-5.
- [8] Jiangsheng Y. Clustering Methods Applications of Multivariate Statistical Analysis (Beijing: Institute of Computational Linguistics) URL: <http://www.archive.org/web/20030422001200/icl.pku.edu.cn/yujs/papers/pdf/clster.pdf> (reference date: 07.03.2017), pp. 15–20.
- [9] Roubos J. and Setnes M. Learning Fuzzy Classification Rules from Labeled Data (Delft: Delft University of Technology) URL: http://www.researchgate.net/publication/222657076_Learning_fuzzy_classification_rules_from_labeled_data/links/09e4150879582c877f000000 (reference date: 05.05.2017), 2006, pp. 7–13.
- [10] Штовба С. Проектирование нечетких систем средствами MATLAB. М.: Горячая линия–Телеком, 2007, с. 187-223.
- [11] Denisova L.A., Nadtochii P.N. and Raskin E.M. Realization of the Automatic Control System with Fuzzy-Logic Static Error Compensation in the Computer-Aided Design Environment Teprol / Automation and Remote Control 76 no 1, 2015, pp. 157-165.
- [12] Meshcheryakov V. and Denisova L. Computer-aided design of the fuzzy control system using the genetic algorithm Dynamics of Systems, Mechanisms and Machines, Dynamics 2016, 2016, p. 7819000.
- [13] Zhang M., Zhang W., Sicotte H. and Yang P. A new validity measure for a correlation-based fuzzy c-means clustering algorithm / Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 4 3865-8.
- [14] Knyazev N.L. and Denisova L.A. The detection methods of dynamic objects / J. Phys.: Conf. Ser. 2018, p. 944 012051.