

УДК УДК 004.02; 004.8

# Исследование и разработка алгоритма извлечения эвристических знаний с помощью методов нечеткой логики в задачах диагностики заболеваний

стр. 58 – 62

Абдулаева З.И., Курбанбаева Д.Ф.

Северо-Западный государственный медицинский университет им. И.И. Мечникова

---

**Контактные данные:** Абдулаева З.И., e-mail: zina@bk.ru

---

**Резюме.** Статья посвящена исследованию и разработке диагностического процесса рака предстательной железы информационной модели для диагностики заболеваний на основе использования нечетко-логического матричного агрегатного вычислителя. Проанализированы сложности и обоснованы перспективы применения алгоритмов извлечения эвристических знаний с помощью методов нечеткой логики.

**Ключевые слова:** нечёткая логика, нечётко-логические системы и модели, матричный агрегатный вычислитель (МАВ), онкология, рак предстательной железы.

---

## Research and development of an algorithm for extracting heuristic knowledge using methods of fuzzy logic in problems diagnosing diseases

pages 58 – 62

Abdoulaeva Z.I., Kurbanbaeva D. F.

---

**Summary.** The article is devoted to the use of fuzzy-logical matrix aggregate calculator. The complexity of the diagnostic process is analyzed describes the methods of fuzzy logic.

**Keywords:** Fuzzy logic, fuzzy logic systems and models, matrix aggregate calculator (MAC), oncology, prostate cancer.

### Введение

Экспертные знания в информационных системах существенно разнятся по уровню надёжности и качества. При этом, чем менее освоена предметная область исследования, тем ценнее уже даже не знания, а слабые свидетельства о возможности такие знания получать. Тем

самым, наполнение экспертных систем плавно смещается от знаний проявленных и надёжно подтверждённых в область знаний неявных (имплицитных), в область стартовых научных гипотез и плохо проверенных экспертных суждений. Неопределённость сопровождающая данные знания, является максимальной, и качество этой неопре-

делённости таково, что она бросает вызов всем существующим методам обработки эмпирических данных и подтверждения научных гипотез.

Как отмечалось в [1], классические вероятностные методы имеют право на существование там, где исходная выборка данных подчиняется требованиям массовости и однородности, но если хоть один из этих критериев нарушается, то о применении вероятностных методов в чистом виде говорить не приходится. Разумеется, в экономических или социальных системах ничего похожего мы не наблюдаем. В данных системах хорошо зарекомендовал себя нечётко-логический подход, который оперирует уже не с вероятностями, а с возможностями, и модели построенные с применением теории нечётких множеств имеют практическое преимущество. В [2,4,11] уже были предложены методы моделирования с применением нечёткой-логики, такие как система сбалансированных показателей (ССП) на основе нечётких правил и МАВ (матричный агрегатный вычислитель), которые неплохо проявили себя в экономических моделях. При этом экономические модели обладают большим числом количественных показателей, например, финансовых, и нечётко-логический блок здесь интерпретирует количественные данные. Цель же проводимого исследования состояла в том, чтобы проанализировать поведение информационной модели с применением МАВ в условиях крайне ограниченного числа показателей, подавляющее число из которых являются эмпирическими и построены исключительно на субъективных (имплицитных) данных, результаты которого и изложены в данной статье. Для исследования поведения модели МАВ при работе с имплицитными данными и знаниями была выбрана медицина, как наука выстроенная на основе обобщения многолетнего опыта работы и интуиции специалистов. В настоящей статье под имплицитными знаниями и данными понимаются такие, которые не прошли полноформатной опытной проверки, относительно которых есть расхождения в экспертных суждениях и по которым не сформирована полноценная статистическая база.

Кроме того, на сегодняшний день имеется значительное количество медицинских информационных систем, основным отличием которых является присутствие интеллектуализированных систем принятия решения для консультативной помощи при диагностике и прогнозировании. Эти системы доказали свою состоятельность применительно к заболеваниям, где сформирована многолетняя база надёжных медицинских наблюдений и существуют надёжные экспертные знания в части профилактики и лечения заболеваний. Применительно к новым областям научного исследования, статистические данные являются скудными, а экспертные суждения – противоречивыми. Это делает актуальной задачу активного процесса по переработке неполной и разнородной статистической информации в надёжные знания, с последующей верификацией этих знаний. К числу таких технологий,

именуемых в зарубежной практике Data Mining, относятся системы лингвистической классификации, нечётко-логического вывода и дефазификации, в том числе и матричный агрегатный вычислитель (МАВ). Соответствующие интеллектуальные технологии кардинально расширяют инструментарий традиционных экспертных систем.

Если говорить о медицинской практике, то в настоящее время накоплено значительное число исследований о возможностях применения аппарата нечеткой логики для решения задач диагностики рака предстательной железы [8,9]. В настоящем исследовании данные для анализа были также взяты из раздела онкологии, в частности по заболеванию «Рак предстательной железы». Онкология - одно из критических направлений медицины (второе место в мире по смертности), а рак предстательной железы (РПЖ) находится на втором месте (по данным ВОЗ), но уже в общей структуре заболеваний раком у мужчин [5]. Необходимо отметить, что данные, которые используются при диагностике РПЖ в подавляющем большинстве, являются эмпирическими, даже те измерения, которые носят количественный характер, являются динамически изменяющимися и не имеют чётких границ. В виде примера можно привести то, что единственный значимый показатель при диагностике РПЖ до проведения биопсии [7], так называемый ПСА-скрининг (ПСА – простатический специфический антиген) заключением USPSTF [10] ещё в 2012 году получил отрицательные рекомендации: «Потенциальные риски перевешивают потенциальные выгоды для пациентов, у которых рак простаты ещё не диагностирован или не проходящих лечение». А во избежание угрозы неправильного определения стадии развития РПЖ, пациентам официально рекомендуется получать «Второе мнение» (обращаться к нескольким врачам). Таким образом, затруднена не только постановка диагноза, но и установление стадии развития болезни, до взятия биопсии. Отсюда следует, что мы имеем дело с неформализованной задачей, которая обладает неоднозначностью, неполнотой и противоречивостью исходных данных, а также большой размерностью пространства решения. В этом отношении, медицина вправе рассчитывать на помощь других наук. Прежде всего, это информатика, кибернетика, статистика, специализированные разделы математики (в т.ч. нечёткие множества и мягкие вычисления, а также нечёткие знания). В работе [5] зарубежными исследователями показаны преимущества нечетко-логических моделей как средства дополнительного обоснования диагноза РПЖ в сравнении с традиционными методами: логистическая регрессия и вероятностные методы, нейронные сети. По-нашему мнению, принципиально важно создать нечетко-логическую интегральную информационную модель, в которой представить все необходимые данные о состоянии пациента и динамике его болезни для усиления диагностики.

Таблица 1.

Информативные методы исследования предстательной железы

Метод	Направление метода
Анамнез	Сбор симптомов и их интерпретация
Пальцевое ректальное исследование	Выявление наличия опухоли, оценка железы
Анализ крови на ПСА	Описывает присутствие в крови онкомаркера
УЗИ предстательной железы	Выявление опухоли, её описание
Биопсия*	Для гистологического анализа состава ткани опухоли

\* Диагноз рака предстательной железы не ставится без проведения биопсии.

### Построение и калибровка информационной модели

Сверхзадача построения информационной модели – это консолидировать все существенные факты, относимые как к самому пациенту в целом, так и к диагностируемому у него заболеванию (РПЖ). Все имеющиеся в обиходе у врача информативные исследовательские методы, с помощью которых можно собрать необходимые факты, показаны в таблице 1, первые четыре из которых являются основанием для проведения биопсии (таблица 1).

В статье предлагается структурировать информационную модель по принципу набора из пяти многоуровневых иерархических деревьев, каждое из которых отвечает соответствующему возрасту пациента:

- до 40 лет;
- 40 – 49 лет;
- 50 – 59 лет;
- 60 – 69 лет;
- 70 лет и выше.

Все экземпляры деревьев, полученные в ходе диагностирования и лечения пациентов, хранятся в компьютерной базе знаний и проходят групповую обработку, на предмет калибровки моделей и методов диагностики. Исходный возрастной квалификационный разрез предполагает, что каждое из «возрастных деревьев» будет структурировано по-своему, и наборы весов каждого из факторов, участвующие в методиках агрегирования исходной информации, будут своими, в соответствии с возрастной спецификой пациента.

Когда возраст пациента задан, мы рассматриваем соответствующее дерево факторов как многоуровневую иерархию (рис. 1).

Данная иерархия имеет следующие уровни представления:

- **Уровень 0.**  $I_0$  – интегральный индекс РПЖ. Действительное число, отсчитываемое от 0 до 4.
- **Уровень 1.** Индексы:
  - $I_{1.1}$  – индекс прямых измерений (оценивается от 0 до 4);
  - $I_{1.2}$  – индекс наблюдений (также оценивается от 0 до 4).
- **Уровень 2.** Блоки факторов. В том числе:
  - $I_{2.0}$  – индекс УЗИ, составляется на основе показаний ТРУЗИ и Эластографии (оценивается от 1 до 10);
  - $I_{2.1}$  – ГЛС – индекс, определяемый по шкале Глисона (оценивается от 1 до 10 по результатам биопсии);
  - $I_{2.2}$  – сводный индекс ПСА, составленный на основе количественных измерений факторов ППСА, ПСАТЗ, ППСАТЗ и др. Оценивается от 1 до 10 методом матричного агрегатного вычислителя (МАН), который будет описан позднее;
  - $I_{2.3}$  – сводный индекс TNM, составленный из измерений факторов Т, N и М по методу МАН. Оценивается от 1 до 10;
  - $I_{2.4}$  – индекс ПРИ (пальцевого ректального исследования). Оценивается от 0 до 1;
  - $I_{2.5}$  – сводный индекс риска. Оценивается от 0 до 1 по сводной оценке признаков факторов;
  - $I_{2.6}$  – ССИ – сводный симптоматический индекс. Оценивается от 0 до 1 (со слов пациента);
  - $I_{2.7}$  – СПИ – сводный психосоматический индекс. Оценивается от 0 до 1 со слов пациента и по результатам анкетирования.

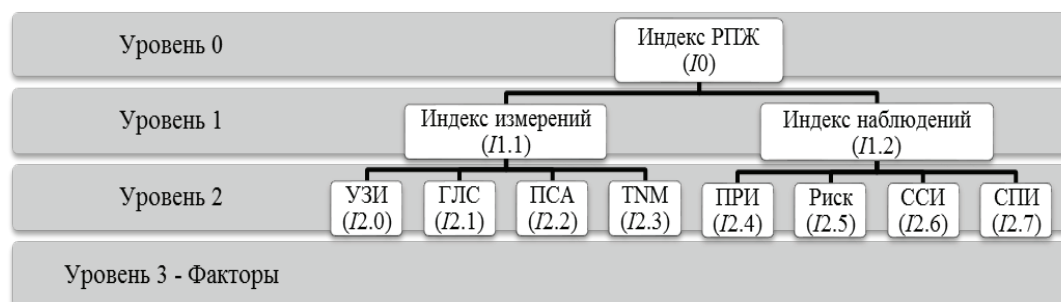


Рис. 1. Иерархическое дерево факторов информационной модели

Здесь необходимо отметить, что индексы  $I_{2.1}$  и  $I_{2.3}$  используются при диагностике РПЖ только по результатам биопсии, поэтому в добиопсийной диагностике они принимают значение «0».

- **Уровень 3.** Количественные измерения и признаковые наблюдения. Факторов, находящихся на данном уровне, найдено пока всего 34 (для сравнения: экономическая модель содержит минимум 200 факторов). Из этих 34 факторов – 20 носят чисто описательный характер. При этом связи между факторами прослеживаются слабо, и при первоначальном рассмотрении они кажутся независимыми друг от друга, однако на деле, поскольку человеческий организм является полностью связанной кибернетической системой, то рассматриваемые факторы модели следует увязывать друг с другом в нечётко-логическую причинно-следственную сеть. Поскольку человеческий организм находится в гомеостазе с внешней средой и саморегулируется, то факторы информационной модели в любой момент времени находятся в состоянии сбалансированности. Выход определённых факторов из состояния баланса свидетельствует о начале заболевания, и момент такого выхода следует отмечать на ранних стадиях, с помощью механизмов специального нормирования. Прямой аналогией в этом плане является хорошо известная в экономике система сбалансированных показателей (ССП).

Теперь эскизно рассмотрим существо МАВ. На вход МАВ поступает набор количественных факторов со своими измерениями. Метод МАВ проходит следующие этапы [3]:

- ранжирование факторов по убыванию предпочтения одних факторов другим. Назначение факторам весов  $\{p\}$  (например, по схеме Фишберна);
- фазификация (лингвистическое распознавание) количественного уровня факторов, с присвоением качественных градаций;
- настройка системы узловых точек  $\{y\}$ , соответствующих виду шкалы интегрального выходного фактора;
- структурирование матрицы, где по строкам – факторы со своими весами  $\{p_i\}$ , по столбцам – качественные градации уровней факторов со своими узловыми точками  $\{y_j\}$ , на пересечении строки и столбца – уровень принадлежности  $\{m_{ij}\}$   $i$ -го фактора  $j$ -й качественной градации;
- определение выходного индекса  $I$  по формуле матричного перемножения;

$$I = p^T \times \mu \times y^T = \sum_{(i)} p_i \sum_{(j)} \mu_{ij} y_j \quad (1)$$

Индекс  $T$  в формуле матричного перемножения выражает операцию транспонирования вектора – столбца  $p$  в вектор – строку и вектор – строку  $y$  в вектор – столбец.

Когда методы МАВ отработаны по всем кустам дерева факторов, можно переходить к интегральной оценке. Здесь нами предлагаются следующие формулы:

$$II.1 = (I_{2.1} + I_{2.2} + I_{2.3}) \times 4 / 27 - 4/9; \quad (2)$$

здесь входные шкалы 1..10 линейно переводятся в выходную шкалу 0..4;

$$II.2 = I_{2.4} + I_{2.5} + I_{2.6} + I_{2.7}; \quad (3)$$

а здесь входные шкалы 0..1 линейно переводятся в выходную шкалу 0..4.

Наконец, интегральный фактор  $I_0$  стадии РПЖ:

$$I_0 = \max \{II.1, II.2\}, \quad (4)$$

то есть стадийность, исходя из принципа осторожности, определяется по наихудшему (по наиболее опасному для пациента) сценарию. Классификатор для показателя  $I_0$  – это равномерная серая шкала Пospелова [6], определённая на носителе  $[0, 4]$  как система из пяти трапециевидных нечётких чисел, с системой узловых точек  $y = \{0,4; 1,2; 2,0; 2,8; 3,2\}$ .

Дробное значение фактора  $I_0$  выражает степень неопределённости, которая неотменимо присутствует при распознавании стадии РПЖ. Например, значение  $I_0 = 1,5$  говорит о том, что врач не в состоянии точно определить стадию заболевания, и для более точного определения стадии необходимо проводить дополнительные исследования.

Калибровка построенной информационной модели осуществляется на основе её «обкатки» на историях болезни реальных пациентов (требуется накопленная статистика на уровне нескольких сотен историй болезни). Таким образом, диагностическая система проходит обучение (полная аналогия с нейронными сетями). Сводя зафиксированные в истории болезни измерения/наблюдения и диагностируемые стадии РПЖ, необходимо оптимизировать систему весов  $\{p\}$  факторов модели таким образом, чтобы расхождение между диагностикой стадии РПЖ в историях болезни, выполненной опытными врачами, и оценкой  $I_0$  по модели, было бы минимальным. Для многомерной оптимизации в непрерывном поле значений  $\{p\}$  уместно использовать традиционные градиентные методы, причём начальной точкой оптимизации является система равных весов, которая незримо участвует в формулах (2) и (3).

## Заключение

Точные и надежное принятие решения в онкологическом прогнозе могут помочь в планировании подходящей терапии и хирургии. Мы зафиксировали стартовое представление о путях формирования информационной модели критического заболевания на примере РПЖ. Данная модель соответствует текущему состоянию исследований по заявленному профилю. Аналогичную работу можно провести по другим критическим заболеваниям (иные разновидности рака, сердечно-сосудистые заболевания и др.). Агрегирование факторов в системе иерархии показателей в основном проводится методом МАВ, но есть и другие разновидности агрегирования.

Следующие шаги нашего исследования:

- калибровка построенной информационной модели на основе большого количества имеющихся в на-

шем распоряжении обезличенных статистических данных;

- прицельное исследование способов ранней диагностики 1-й стадии РПЖ, без проведения биоп-

сии, с выделением набора отвечающих симптомов и структурирования их в иерархическую систему, по аналогии с тем, как это сделано в настоящей работе.

#### **БИБЛИОГРАФИЯ:**

1. **Абдулаева З.И.** Применение нечётких множеств и мягких вычислений в медицинской статистике // НоваИнфо - NovaInfo.Ru – 2016 – № 51. URL: <http://novainfo.ru/article/7714>
2. **Абдулаева З.И.** Разработка методов управления рисками инновационной деятельности / Дисс. на соискание уч. ст. канд. экон. наук. – СПб, Санкт-Петербургский университет управления и экономики, 2013. – 200 с. - Также на сайте: [http://zina.ifel.ru/docs/Diss\\_AZ.pdf](http://zina.ifel.ru/docs/Diss_AZ.pdf).
3. **Абдулаева З.И., Недосекин А.О.** Стратегический анализ инновационных рисков / - СПб: Изд. СПбГПУ, 2013. – 146 с.
4. **Недосекин А.О., Абдулаева З.И.** Комплексная матричная модель организации, функционирующей в условиях информационных угроз // Информация и Космос. – 2015. – № 1. – С. 32-37.
5. **Абдулаева З.И.** Информационная модель для диагностики критических заболеваний // NovaInfo.Ru (Электронный журнал.) – 2017– № 64. URL: <http://novainfo.ru/article/12724>
6. **Поспелов Д.А.** «Серые» и/или «чёрно-белые» // Прикладная эргономика специальный выпуск «Рефлексивные процессы» . – 1994. – №1. – С. 29 – 33.
7. **Топузов М.Э.** Пути оптимизации ранней диагностики рака предстательной железы. – Дисс. на соиск. уч. ст. доктора медицинских наук / М.Э. Топузов. – СПб.: СЗГМА им. И.И. Мечникова, 2008. – 241 с.
8. **Kuo, R.-J. , Huang, M.-H., Cheng, W.-C., Lin, C.-C.c., Wu, Y.-H.** Application of a two-stage fuzzy neural network to a prostate cancer prognosis system (Article) // Artificial Intelligence in Medicine. – 2015. – Volume 63, Issue 2. – Pages 119-133.
9. **Seker, H., Odetayo, M.O., Petrovic, D., Naguib, R.** A fuzzy logic based-method for prognostic decision making in breast and prostate cancers (Article) // IEEE Transactions on Information Technology in Biomedicine. – 2003. – Volume 7, Issue 2. – Pages 114-122.
10. **USPSTF – United States Preventive Services Task Force** «Talking With Your Patients About Screening for Prostate Cancer (02.07.2012)». URL: <https://www.uspreventiveservicestaskforce.org>
11. **Vinogradov, V.V., Abdoulaeva Z.I.** Fuzzy-set economic stability analysis model of mineral complex of the Russian Federation // Proceedings of the 19th International Conference on Soft Computing and Measurements. – 2016. – P. 489-490.

#### **АВТОРЫ:**

1. **Абдулаева Зинаида Игоревна** – кандидат экономических наук, доцент кафедры медицинской информатики и физики СЗГМУ им. И.И. Мечникова Минздрава России.  
**Abdoulaeva Z.I.** – Candidate of Economical Sciences, Associate Professor of the Department of Medical Informatics and Physics, SZGMU named after I.I. Mechnikov of the Ministry of Health of Russia.  
e-mail: [zina@bk.ru](mailto:zina@bk.ru)
2. **Курбанбаева Динара Фархатовна** – кандидат экономических наук, доцент кафедры медицинской информатики и физики СЗГМУ им. И.И. Мечникова Минздрава России.  
**Kurbanbaeva D.F.** – Candidate of Economical Sciences, Associate Professor of the Department of Medical Informatics and Physics, SZGMU named after I.I. Mechnikov of the Ministry of Health of Russia.  
e-mail: [kurbanbaeva\\_dina@mail.ru](mailto:kurbanbaeva_dina@mail.ru)