

УДК 004.89

ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ ОНТОЛОГИЧЕСКОГО ПОДХОДА К ДИАГНОСТИКЕ ОРФАННЫХ ЗАБОЛЕВАНИЙ¹

Б.А. Кобринский (*kba_05@mail.ru*)
Н.А. Благосклонов (*nblagosklonov@gmail.com*)
Федеральный исследовательский центр «Информатика
и управление» РАН, Москва

Н.С. Демикова (*ns_d@rambler.ru*)
Российская Медицинская Академия Непрерывного
Профессионального Образования, Москва

В.В. Грибова (*gribova@iacp.dvo.ru*)
Е.А. Шалфеева (*shalf@iacp.dvo.ru*)
М.В. Петряева (*margaret@iacp.dvo.ru*)
Институт автоматизации и процессов управления
ДВО РАН, Владивосток

Работа представляет первые результаты компьютерной диагностики редких заболеваний с использованием онтологической системы. Алгоритм включает обработку модальностей признаков с их коэффициентами в сочетании с комплексными характеристиками факторов уверенности признаков по возрастным группам пациентов. Это позволяет в дальнейшем получить узкий дифференциально-диагностический ряд заболеваний на долабораторном этапе обследования.

Ключевые слова: онтология, дифференциальная диагностика, редкие болезни, модальности признаков, факторы уверенности, интеллектуальная система поддержки принятия решений.

Введение

К группе орфанных (редких) болезней в России принято относить заболевания, которые встречаются не чаще, чем 10 случаев на 100 000

¹ Работа выполнена при частичной финансовой поддержке РФФИ (проект 17-07-00956).

населения (в соответствии с Федеральным законом от 21.11.2011 N323-ФЗ). В крупнейшей мировой базе генетически-обусловленных болезней Online Mendelian Inheritance in Man (OMIM) по состоянию на 26 марта 2019 года содержится информация о 6413 заболеваниях [OMIM, 2019]. Диагностика этой патологии вызывает у врачей серьезные трудности в виду необходимости учета разнообразных проявлений заболеваний на разных этапах ее развития. В последний период особенно актуальным является учет особенностей первых проявлений болезни (манифестации) для группы метаболических болезней накопления, где в качестве лечения применяется ферментозаместительная терапия и начаты экспериментальные работы по генной терапии [Das, 2013; Parini et al., 2010; Ellinwood et al., 2004]. Своевременно начатое лечение способно предупредить или затормозить дальнейшее прогрессирование патологии.

Для принятия диагностических решений необходимо учитывать множество факторов. Качество оказываемой медицинской помощи зависит уже не только от уровня подготовки (компетентности) врачебного персонала, но и от систем, осуществляющих поддержку решений врача [Purcell, 2005; A-Jasmi et al., 2010; Alves et al., 2016]. Используемые в настоящее время в клинической генетике системы [Baraitser et al., 2001; Köhler et al., 2009; POSSUM, 2019] не позволяют достаточно эффективно решать проблему диагностики многочисленных редких наследственных болезней в раннем возрасте.

Целью данной публикации является описание прототипа онтологической системы для поддержки принятия диагностических решений на модели наследственных лизосомальных болезней накопления.

Высокая степень сходства исследуемых болезней предполагает выдачу узкого дифференциально-диагностического ряда, позволяющего провести последующее направленное обследование, что важно в отношении ускорения распознавания заболевания и экономии средств дорогостоящих реактивов. Рассмотрение решения этой проблемы возможно с использованием методов искусственного интеллекта.

1. Представление знаний об орфанных заболеваниях

Извлечение знаний на 1 этапе осуществлялось из литературных источников и представлялось в форме текстологических карт [Kobrinskii et al., 2018]. В дальнейшем когнитолог с экспертами использовали их в процессе обсуждения модальности и мер доверия к признакам. На 2 этапе эксперты формировали оценки факторов уверенности для манифестации признаков и их выраженности для разных возрастных групп.

В процессе построения прототипа экспертной системы для диагностики орфанных болезней на основе онтологической системы [Gribova et al., 2017] была выявлена целесообразность модификации

атрибутов симптомов диагностируемых заболеваний и оценок факторов уверенности, описанных в [Кобринский, 2018]. Понятие наличия признаков было заменено понятием модальности (лат. *modus* — мера, оценка высказывания). Релевантные признаки были подразделены на главные, необходимые и второстепенные, что определяется их модальностью [GARD, 2019]. Признаки не только были распределены по модальности, что в определенной степени учитывало частоту их встречаемости, но и сопровождалось коэффициентами уверенности. Кроме того, были внесены изменения в шкалы мер доверия экспертов к манифестации и выраженности признаков, в большей степени учитывающие способы их определения врачами, что отражено в разделе 2 данной статьи.

2. Шкалы и формулы для оценки характеристик наследственных метаболических болезней

Формирование шкал модальности (M) признаков с их коэффициентами и формул для комплексной оценки совокупности факторов уверенности для лизосомальных болезней являлось необходимым условием при построении диагностических правил нечетко проявляющихся орфанных заболеваний.

Подбор количественных характеристик модальности подвергался изменениям в процессе анализа интегральных оценок. В результате модальность признаков была представлена градациями 5, 4 и 2, соответствующими лингвистическим характеристикам – главные, необходимые, второстепенные.

Манифестация (m) признака характеризовалась мерой доверия (уверенности) экспертов в том, что данный признак определённого заболевания манифестирует в конкретном возрасте (возрастной группе). Определяется в интервале $[-1; 1]$, где:

- -1 характеризует невозможность проявления данного признака в данном возрасте по физиологическим причинам, либо реализация признака завершилась в предыдущих возрастных группах;
- 0 характеризует норму, т.е. признак чисто теоретически может быть в данном возрасте, но либо не характерен для данного заболевания, либо ещё не успевает проявиться в силу особенностей течения заболевания;
- $[0,1; 1]$ характеризует меру доверия экспертов к тому, что признак проявится (манифестирует) в одном из возрастных периодов.

Особенностью m является то, что значения, отличные от -1 , в сумме не могут превышать 1 . В каждом возрастном периоде указывается уверенность того, что признак манифестирует (обнаруживается) именно в данном возрасте. Факторы уверенности в манифестации по

предшествующим четырем возрастным периодам имеют значения m_1, m_2, m_3, m_4 , дополняют друг друга, за исключением отрицательных величин.

Выраженность (s) характеризует уверенность экспертов в том, что данный признак встречается в конкретной возрастной группе с определенной степенью проявления. Определяется в интервале $[0; 10]$.

Изменение выраженности по возрастным периодам косвенно указывает на скорость развития патологических проявлений. Для каждого возрастного периода имеем: s_1, s_2, s_3, s_4 .

Для получения комплексной количественной оценки признака было предложено использовать следующую формулу:

$$P_i = M_i \cdot m_i \cdot s_i$$

где: P_i – параметр (признак), M_i – модальность (modality) признака, m_i – фактор уверенности манифестации (manifestation), s_i – фактор уверенности выраженности (severity).

3. Основные характеристики онтологии диагностики орфанных заболеваний

Одним из современных подходов к созданию систем поддержки принятия врачебных решений является онтологический. Одной из онтологий медицинской диагностики является Онтология знаний об идентификации заболеваний [Грибова и др., 2018]. Основными ее особенностями являются:

А. Возможность формирования симптомокомплексов, отражающих как динамику патологического процесса, так и текущее состояние организма.

Б. Возможность формирования разных симптомокомплексов заболеваний с учетом категорий пациентов (пол, возраст, особые состояния, например, беременность и др.) и разных подходов к диагностике (наиболее щадящий, быстрый или экономически оправданный и т.п.).

В. Возможность использования синдромального подхода к диагностике.

Г. Наличие простых и составных признаков (составной – это набор характеристик с множеством значений), расширенных модальностями, экспертными оценками, факторами выраженности значения признака.

Д. Возможность проведения дифференциальной диагностики заболеваний и выбора соответствующих методов лечения с учетом этиологии, патогенеза, развития процесса во времени.

Е. Учет причинно-следственных связей изменений под влиянием внешних воздействий, оказываемых на организм на разных этапах заболевания.

Ж. Возможность задавать необходимые условия для каждого заболевания: факторы, без которых невозможно заболевание (например, пол, время года и др.).

Среди указанных особенностей онтологии имеются необходимые для орфанных заболеваний.

4. Принципы функционирования онтологической диагностической системы

Для диагностики орфанных заболеваний возможны две ситуации:

1. Пациент впервые с подозрением (с признаками некоторых нарушений) обращается к врачу и требуется предложить адекватные гипотезы о заболеваниях, которые можно заподозрить.

2. Пациент имеет предварительный диагноз и требуется на основе имеющихся признаков и/или их динамики убедиться в наличии именно этого заболевания либо сформировать дифференциальный ряд для последующего уточнения другими специалистами и направленного проведения специальных исследований.

В зависимости от потребности практики для программного сервиса диагностики может быть сформирована база знаний (БЗ) с симптомокомплексами, отражающими либо текущее состояние организма, представленное набором возрастных категорий, либо с симптомами, отражающими динамику патологического процесса, либо с симптомокомплексами обоих типов. Каждый признак имеет шкалу его оценки, состоящую из модальности его проявления и экспертной количественной характеристики в виде факторов уверенности. Онтология медицинской диагностики-IACPaaS дает возможность определить связь заключений с объективными данными в истории болезни.

Используемая коллективом (авторами) форма представления онтологий и знаний на их основе (иерархические семантические сети) и применяемый инструментарий IACPaaS [Gribova et al., 2017] позволяют коллективно создавать программные сервисы диагностики (интегрируя решатели с БЗ) и эффективно модернизировать базы знаний. В то же время, поиски семантического подобия в онтологиях представляют собой полезный способ использования семантической структуры фенотипических аномалий человека при решении проблемы дифференциальной диагностики.

5. Первые результаты эксперимента по применению онтологической системы в предметной области наследственной патологии

Для анализа результатов диагностики были взяты реальные истории болезни пациентов из отечественных и зарубежных открытых источников

с верифицированными диагнозами. Истории болезни вводились через сервис ввода истории болезни платформы IACPaas и затем подавались на вход сервису диагностики орфанных заболеваний. Результатом работы сервиса является как подробное объяснение соответствия признаков пациента знаниям о заболевании, так и соответствия признаков пациента знаниям о разных сходных заболеваниях.

Особый интерес представляет распознавание болезней при неизвестном диагнозе. Пример дифференциальной диагностики в таком случае продемонстрировал вполне приемлемые результаты для долабораторного этапа идентификации нозологической природы заболевания.

При запуске Сервиса Поиска гипотез для ребенка 5 лет с подозрением на лизосомальное заболевание [Gabrielli et al., 2016] было получено семь гипотез болезней из группы мукополисахаридозов. В пользу диагноза IS (синдром Шейе) указывали 4 принципиально важных признака (грубые черты лица, помутнение роговицы, гепатомегалия, тугоподвижность крупных суставов) при отсутствии полезных для подтверждения диагноза сведений о 2х признаках (задержка роста и снижение слуха).

В пользу диагноза мукополисахаридоз, тип IH (синдром Гурлер) система выявила 5 принципиально важных признаков (грубые черты лица, помутнение роговицы, тугоподвижность крупных суставов, гепатомегалия, спленомегалия), но для однозначного подтверждения требовалась информация еще о 3х признаках (короткая шея, макроглоссия, умственная отсталость) (рис.1).

Таким образом, в пользу диагноза Шейе имелись 4 принципиально важных признака, а в пользу синдрома Гурлер 5 существенных признаков, но диагноз синдрома Шейе подтвержден 75% признаков (6 из 8), а синдрома Гурлер – 45% (7 из 15). Остальные 5 нозологий из группы мукополисахаридозов менее вероятны. Такой вариант соответствует имеющему место в практике и является удовлетворительным.

6. Дискуссия

Семантические поиски сходства в онтологиях с использованием метрики осложняются тем фактом, что симптомы могут иметь различную степень специфичности, а также может иметь место наличие признаков, не связанных с данной болезнью [Köhler et al., 2009]. В зависимости от опыта врача и наличия специального оборудования (ЭХО-КГ, щелевая лампа и др.) клинические отклонения могут быть описаны более или менее подробно. Для преодоления этого нами был разработан упомянутый выше комплекс из факторов уверенности с нестандартными интервалами оценок и коэффициенты модальности признаков. Также апробируются различные режимы работы онтологической системы.

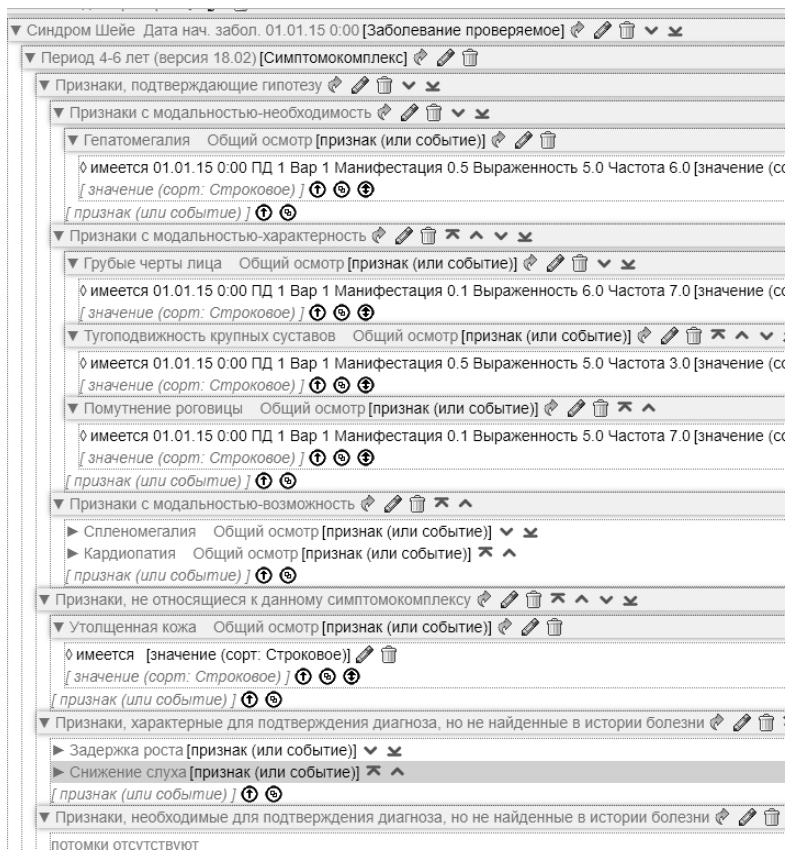


Рис.1. Фрагмент результата работы сервиса поиска гипотез при синдроме Шейе (без предварительной врачебной гипотезы).

Заключение

Высокая степень сходства исследуемых лизосомальных болезней накопления, что характерно и для большинства других наследственных орфанных болезней, предполагает получение не единственной возможной нозологической формы, а узкого дифференциально-диагностического ряда, позволяющего врачу провести последующее направленное лабораторное обследование, что важно в отношении ускорения распознавания заболеваний. Использование онтологической системы позволяет выдвинуть или подтвердить гипотезу врача с одновременным указанием на отсутствие необходимых для уточнения окончательного диагноза признаков.

Список литературы

- [Грибова и др., 2018] Грибова В.В., Петряева М.В., Окунь Д.Б., Шалфеева Е.А. Онтология медицинской диагностики для интеллектуальных систем поддержки принятия решений // Онтология проектирования. 2018. Т.8. №1(27).
- [Кобринский, 2018] Кобринский Б.А. Триединство факторов уверенности в задачах медицинской диагностики // Искусственный интеллект и принятие решений. 2018. №2.
- [A-Jasmi et al., 2010] A-Jasmi F., Moldovan L., Clarke J.T. Hunter disease eClinic: interactive, computer-assisted, problem-based approach to independent learning about a rare genetic disease // BMC Med Educ. 2010. 10:72.
- [Alves et al., 2016] Alves R., Piñol M., Vilaplana J., Teixidó I., Cruz J., Comas J., Vilaprinyo E., Sorribas A., Solsona F. Computer-assisted initial diagnosis of rare diseases // Peer J. 2016. 4.
- [Baraitser et al., 2001] Baraitser M., Winter R.M. London Dysmorphology Database, London Neurogenetics Database & Dysmorphology Photo Library on CD-ROM. 3rd ed. – Oxford: Oxford University Press, 2001.
- [Das, 2013] Das S.K. Inborn Errors of Metabolism: Challenges and Management // Indian J Clin Biochem. 2013. 28(4).
- [Ellinwood et al., 2004] Ellinwood N.M., Vite C.H., Haskins M.E. Gene therapy for lysosomal storage diseases: the lessons and promise of animal models // J Gene Med. 2004. 6.
- [Gabrielli et al., 2016] Gabrielli O., Clarke L.A., Ficcadenti A., Santoro L., Zampini L., Volpi N., Coppa G.V. 12 year follow up of enzyme-replacement therapy in two siblings with attenuated mucopolysaccharidosis I: the important role of early treatment // BMC Medical Genetics. 2016. 17:19.
- [GARD, 2019] GARD. Genetic and Rare Diseases Information Center. <https://rarediseases.info.nih.gov/>, last accessed 2019/04/15
- [Gribova et al., 2017] Gribova V., Kleschev A., Moskalenko P., Timchenko V., Fedorischev L., Shalfeeva E. The IACPaaS cloud platform: Features and perspectives // Computer Technology and Applications (RPC), 2017 Second Russia and Pacific Conference on. - IEEE, 2017.
- [Kobriniskii et al., 2018] Kobriniskii B.A., Blagosklonov N.A. Hybrid approach to knowledge extraction: textual analysis and evaluations of experts // Open Semantic Technologies for Intelligent Systems. Материалы Международной научно-технической конференции. Вып.2. Республика Беларусь, Минск, 15-17 февраля 2018. – Минск: БГУИР, 2018.
- [Köhler et al., 2009] Köhler S., Schulz M.H., Krawitz P., Bauer S., Dölken S., Ott C.E., Mundlos C., Horn D., Mundlos S., Robinson P.N. Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies // American Journal Human Genetics. 2009. 85(4).
- [OMIM, 2019] Online Mendelian Inheritance in Man® An Online Catalog of Human Genes and Genetic Disorders. <https://www.omim.org/>, last accessed 2019/04/15
- [Parini et al., 2010] Parini R., Andria G. Lysosomal storage diseases: early diagnosis and new treatments // Mariani Foundation paediatric neurology. Vol.23. – Chicago: John Libbey Eurotext, 2010.

[POSSUM, 2019] POSSUM. Pictures of Standard Syndromes and Undiagnosed Malformations. <https://www.possun.net.au/>, last accessed 2019/04/15.

[Purcel, 2005] Purcell G. What makes a good clinical decision support system // BMJ. 2005. Vol.330.

THE POSSIBILITY OF APPLYING THE ONTOLOGICAL APPROACH TO THE DIAGNOSIS OF ORPHAN DISEASES

Boris A. Kobrinskii (*kba_05@mail.ru*)

Nikolay A. Blagosklonov (*nblagosklonov@gmail.com*)
Federal Research Center “Computer Science and Control”
of RAS, Vavilova str. 44, kor.2, Moscow, 119333, Russian
Federation

Nataliya S. Demikova (*ns_d@rambler.ru*)

Russian Medical Academy of Continuous Professional
Education, Barrikadnaya str. 2/1, Moscow, 125993, Russian
Federation

Valeriya V. Gribova (*gribova@iacp.dvo.ru*)

Elena A. Shalfeeva (*shalf@iacp.dvo.ru*)

Margaret V. Petryaeva (*margaret@iacp.dvo.ru*)

Institute of Automation & Control Processes FEB RAS,
Radio str. 5, Vladivostok, 690041, Russian Federation

The paper presents the first results of computer diagnostics of rare diseases using the ontological system. The algorithm includes the processing of feature modalities with their coefficients in combination with complex characteristics of feature confidence factors by age groups of patients. This makes it possible to further obtain a narrow differential diagnostic series of diseases at the pre-laboratory stage of the examination.

Keywords: ontology, differential diagnostics, rare diseases, modality of symptoms, confidence factors, intelligent decision support system.