



С. В. Денисенко

Диагностика и прогнозирование состояний при экстракорпоральном оплодотворении на основе онтологии знаний

Национальная медицинская академия последипломного образования
имени П. Л. Шупика, г. Киев

Ключевые слова: диагностика и прогнозирование состояний, экстракорпоральное оплодотворение, медицинская онтология, онтология, ориентированная на задачу, вспомогательные репродуктивные технологии, ориентированные графы.

Рассмотрены теоретические основы и концептуальные возможности создания онтологии знаний по экстракорпоральному оплодотворению. Онтология представлена многоуровневой структурой, каждый уровень которой предназначен для решения конкретной задачи – диагностики, прогнозирования, выбора метода лечения. Соединение предметной и задачной онтологий дает возможность существенного сокращения объемов информации на каждом уровне и соответственного уменьшения объемов вычислительных операций.

Діагностика та прогнозування станів при екстракорпоральному заплідненні на основі онтології знань

С. В. Денисенко

Розглянули теоретичні основи та концептуальні можливості створення онтології знань з екстракорпорального запліднення. Онтологія представлена багаторівневою структурою, кожен рівень котрої призначений для вирішення конкретного завдання – діагностики, прогнозування, вибору методу лікування. Поєднання предметної онтології та онтології завдань дає можливість істотного скорочення обсягів інформації на кожному рівні і відповідного зменшення обсягів обчислювальних операцій.

Ключові слова: діагностика та прогнозування станів, екстракорпоральне запліднення, медична онтологія, онтологія, що орієнтована на завдання, допоміжні репродуктивні технології, орієнтовані графи.

Запорізький медичний журнал. – 2014. – №2 (83). – С. 137–140

Diagnosis and prognosis of states in invitro fertilization based on the ontology knowledge

S. V. Denysenko

The present study focuses on the theoretical principles and conceptual possibilities of creating of the ontology knowledge of the in vitro fertilization. The multi-level ontology, each level of which is aimed to solve specific task, i.e. diagnosis, prediction, choice of treatment, is suggested. Compound of the subject ontology and task-oriented ontology enables significant reduction in the volume of information on each level.

Key words: diagnosis and prediction of states, in vitro fertilization, medical ontology, task-oriented ontology, assisted reproductive technologies, oriented graphs.

Zaporozhye medical journal 2014; №2 (83): 137–140

Современное общество неизбежно движется в сторону информатизации все большего и большего количества своих функций. Медицина как направление, обоснованно считающееся консервативным, весьма осторожно продвигается в направлении информатизации. Очевидно, многое зависит от готовности отдельных направлений принять те или иные изменения. В то же время проблема структуризации знаний является одной из перспективных и первоочередных задач.

Одним из важнейших направлений, где наиболее актуален интерес к медицинской онтологии как одной из перспективных стратегий, является репродукция человека. Фундаментальные исследования в этой области привели к тому, что в последнее время была открыта новая эпоха в лечении бесплодия – эпоха вспомогательных репродуктивных технологий (ВРТ), которые вывели проблему лечения бесплодия из тупика и позволили добиться успеха большому числу супружеских пар, обреченных на бездетность. К настоящему времени стандартная процедура экстракорпорального оплодотворения (ЭКО) и его модификации (ЭКО-ОД, ИКСИ) стали почти «рутинными» методами, применяемыми в ситуациях, при которых оказываются бессильными кон-

сервативные и хирургические средства восстановления естественной фертильности.

В связи с этим существует необходимость разработки и систематизации критериев, способствующих диагностике состояния женщин и прогнозированию эффективных путей достижения беременности при наиболее часто встречающихся формах бесплодия. В частности, важна диагностика генетических аномалий у эмбрионов до момента их имплантации в стенку матки. Такую диагностику можно проводить на отдельных клетках эмбрионов, полученных в результате процедуры экстракорпорального оплодотворения, до переноса эмбрионов в матку. Кроме того, при лечении бесплодия методом ЭКО, как и при любом другом методе, возможны различные осложнения, причем каждый его этап связан с относящимся к нему рисками. Такое прогнозирование, с одной стороны, поможет использовать экстракорпоральное оплодотворение для лечения бесплодия с максимальной результативностью, с другой, – способствовать существенному сокращению бесполезных затрат на проведение тех или иных процедур.

Цель работы

Обоснование концептуальных и теоретических основ онтологии знаний по ЭКО.



Как известно, классификация онтологий по цели создания предполагает чаще всего 4 уровня: онтология представления, онтология верхнего уровня, онтология предметной области и прикладная онтология. В настоящем исследовании рассмотрены задачи прикладной онтологии и особенно онтологии, ориентированной на задачу [1–3].

Назначение прикладной онтологии в том, чтобы описать концептуальную модель конкретной задачи или приложения. Прикладные онтологии описывают концепты, зависящие как от онтологии задач, так и от онтологии предметной области. Такие онтологии содержат наиболее специфичную информацию.

Онтология, ориентированная на задачу, содержит термины, которые используются при разработке системы, решающей конкретные задачи: определение целей, диагностика, прогнозирование, построение классификации. Она отражает специфику приложения, но может также содержать некоторые общие термины. При этом онтология задачи использует специализацию терминов, представленных в онтологиях верхнего уровня (общих онтологиях).

Считаем, что моделирование патологического процесса как таковое практически полностью отсутствует в современных онтологических моделях. Основной недостаток теоретических разработок, связанных с моделированием патологического процесса, обусловлен тем, что в большинстве исследований процесс представляется как двухмерный. В то же время характер клинических задач вовсе не предполагает использование всего массива информации, накопленной в медицине. Для каждой из решаемых постановочных задач используется до 40–50 симптомов, хотя задач в клинике много, и общий объем сведений, необходимых для обобщения, составляет тысячи симптомов, признаков, симптомокомплексов и т. д.

Таким образом, врач всегда видит и использует то, что соответствует конкретной задаче (диагностики, прогнозирования и т. п.). В то же время крайне важно обеспечить системное решение клинических задач. В этом случае общая модель знаний о заболевании должна быть представлена многоуровневой моделью [4].

Анализ разрабатываемых в последние годы систем медицинской диагностики второго класса показал, что используемые в них онтологии являются сравнительно простыми и одновременно не отражают такие повсеместно используемые врачами в их практике знания предметной области, как причины патологических процессов, лежащие в основе бесплодия; различные типы причинных связей между признаками и заболеваниями; воздействие событий на значения признаков при бесплодии; различные варианты изменений значений признаков, зависящие от анатомо-физиологических особенностей пациентов и т. д.

Помимо этого, одним из отрицательных свойств некоторых разрабатываемых систем является то, что круг их применения достаточно узок. Это обусловлено тем, что они представляют собой либо макетные версии, выполненные для исследовательских целей, либо разработаны для определенного медицинского учреждения и не доступны за рамками его локальной сети. С другой стороны, системы медицинской диагностики, предоставляющие широкий доступ к своим ресурсам с применением современных

сетевых технологий, не позволяют экспертам расширять используемые в них базы знаний.

Таким образом, актуальной задачей является разработка системы медицинской диагностики, основанной на знаниях экспертов, и модели онтологии, учитывающей все приведенные особенности медицинских знаний, в которой их модель имеет форму, наиболее близкую к представлениям экспертов и позволяет определять не только диагноз пациента, но и объяснять его.

Соотношения между неизвестными и параметрами мы разделили на несколько смысловых групп:

- 1) знания о причинно-следственных отношениях и связях;
- 2) причинно-следственные связи, являющиеся причинами значений каждого признака на разных интервалах времени;
- 3) определение для каждого признака свойства границ интервалов разбиения на оси времени, связанного с этим признаком;
- 4) причины патологических состояний, входящих в диагноз.

Мы поставили общую задачу распознавания возможных патологических процессов у пациента на основе знаний предметной области, к которым относятся значения признаков (в моменты их наблюдения), значения его анатомо-физиологических особенностей (постоянные во времени) и значения произошедших с ним событий (в моменты, когда они происходили), а также для каждого такого состояния пациента его причина и объяснение. Такая задача получается при наложении на используемую модель онтологии нескольких ограничений: у пациента фиксируется наиболее выраженный и важный патологический процесс и рассматривается лишь один период развития.

Множеству хранящихся в онтологии симптомов и признаков с заданным на нем бинарным отношением, индуцируемым ассоциативными связями между различными элементами, взаимно однозначно ставится в соответствие ориентированный связный граф G . Каждому элементу информации a ($a \in \Omega$) соответствует вершина i графа G , а любые две вершины i, j ($i, j \in G$) соединены дугой, если соответствующие элементы a_i, a_j , $a_i \in \Omega$ находятся в отношении, определенном на Ω . Множество Ω , а следовательно, граф G – конечны. Граф без контуров, вершины графа интерпретируются как состояние организма (рис. 1).

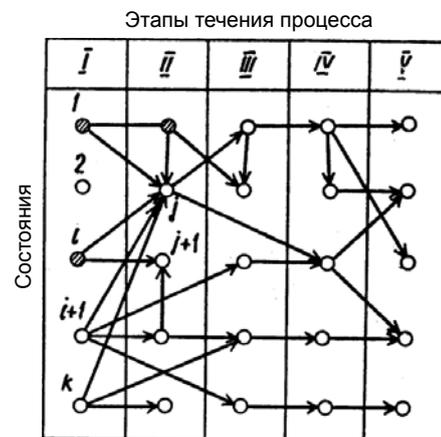


Рис. 1. Метод представления течения заболеваний в виде графа.



Если у пациента имеется несколько состояний, интерпретируемых как некоторое множество вершин графа G , то процесс развития патологического процесса рассматривается как передача информации попарно инцидентным его вершинам и выборе из множества задействованных в процессе вершин последовательности, удовлетворяющей определенным условиям.

Каждой вершине x графа G поставлены в соответствие пара действительных неотрицательных чисел: ее вес и порог $-S_x$ и n_x .

Вопросы представляют собой k -мерные векторы $\vec{v} = (v_1, v_2, \dots, v_k)$ где $v_i = \frac{1}{0}$, а k – число вершин графа. Ответ, выдаваемый системой, также k -мерный вектор $\vec{\omega} = (\omega_1, \dots, \omega_k)$, где $\omega_i = \frac{1}{0}$.

Каждой дуге ij графа G соотнесено действительное неотрицательное число p_{ij} ($0 \leq p_{ij} \leq 1$) которое можно интерпретировать как вероятность перехода из вершины i к вершине j .

Принципиально диагностическая или прогностическая система содержит два оператора: F и Q . Оператор F служит для предварительного отбора в G информации, относящейся к заданному вопросу \vec{v} , и ставит в соответствие вектору \vec{v} граф $T \vec{v}$ ($T \vec{v} \leq G$) При помощи оператора Q в графе $T \vec{v}$ строится ответ $\vec{\omega}$.

Для каждой вершины x определяется числовая функция Ψ_x^v , зависящая от веса вершин, являющихся началом дуг, конец которых x , и от вопроса \vec{v} . Оператор F может быть реализован при помощи последовательности функций $E = \{l_i(x)\}$, определяемых индуктивно.

Пусть системе задан вопрос \vec{v} . Тогда для $x \in G$

$$e_i(x) = \begin{cases} Y_1 & \text{при } v_x = 1, \text{ в противном случае, где } Y_1 = \{y/x \sim \rightarrow y, \\ 0 & \end{cases}$$
 $y \in G, \Psi_x^v > \pi_y\}.$

Пусть определены функции $e_m(x)$ для всех $m \leq i$. Тогда для $x \in Y_i$ $e_{i+1}(x) = Y_{i+1}$, где $Y_{i+1} = \{y/x \sim \rightarrow y, y \in G, \Psi_x^v > \pi_y\}$.

Значение оператора F – граф $T \vec{v}$, вершины которого принадлежат множеству $G \cup Y$ – дуги, соединяющие вершины, те же, что и в G . Если $T \vec{v}$ не содержит ни одного пути, которому принадлежат все вершины x , таких, что $V_x = 1$, то система на вопрос не дает ответа.

Оператор Q реализован арифметической функцией. Ее выбор определяется особенностями задачи и степенью сложности патологического процесса. В простейшем случае, если под вершиной графа понимать состояние больного, то прогнозирование течения процесса можно осуществлять с помощью: $\pi = p_{ij} S_i - \sum p_{kj} S_k$, где p_{ij} – вероятность перехода из имеющегося состояния пациента в прогнозируемое; p_{kj} – вероятность перехода из необнаруженных состояний пациента в прогнозируемое.

Недостатком такого подхода в двумерной модели является необходимость осуществления полного перебора значений параметров для описания всевозможных состояний, что резко увеличивает объем графа.

Следует подчеркнуть, что онтологические представления дают возможность существенно уменьшить объем инфор-

мации за счет многоуровневости представления патологического процесса.

В более сложных случаях, когда под вершиной понимают градации симптомов, показателей и т. д., арифметическая функция может быть такой:

$$\pi_j = \sum_i p_{ij} S_i - \sum_k p_{kj} S_k$$

Если рассмотреть наиболее важный для практики случай с использованием лечебных мероприятий, то уровень графа Q определяет состояние пациента у в фиксированный момент времени t . Между интервалами значений параметров больного и вершинами графа установлено взаимно однозначное соответствие. Каждая дуга графа Q отвечает возможному изменению значения параметров – переходу из интервала на t -м уровне в интервал t_{m+1} уровня при применении лечения l . Каждой дуге графа Q соотнесено действительное неотрицательное число p^l_{ij} ($0 \leq p^l_{ij} \leq 1$), которое можно в данном случае интерпретировать как вероятность соединения вершины i t_{m-1} уровня с вершиной j t -го уровня при применении лечения l .

Количество уровней и временные интервалы выбираются из соображений учета качественных изменений в организме при течении заболевания. Соответственно последним изменяются вероятности осложнений, состояний и т.п.

Промежутки между уровнями выбираются так, чтобы внутри промежутка вероятности осложнений были примерно равны, и за время, соответствующее каждому промежутку, наблюдали эффект лекарственных средств.

Введем следующие обозначения. Пусть $q^l_{\sigma(ij)}$ – вероятность появления осложнения, а в результате дрейфа параметров при переходе из i -й вершины в l -ю под влиянием воздействия l ; S_σ – вес осложнения σ ; $\tau^{\lambda(l)}_{\sigma(ij)}$ – вероятность компенсации дрейфа параметров с помощью лечения λ (l); ΔS_σ – отклонение от нормы веса S_σ (вес нормы понимается условно равным 1). Определим также величину R^l_{ij} как отклонение выбранной стратегии от оптимальной. Она для одного параметра равна: $R^l_{ij} = q^l_{\sigma(ij)} \Delta q_\sigma (1 - \tau^{\lambda(l)}_{\sigma(ij)})$.

Тогда характеристику перехода из вершины i в j в смысле приближения соответствующего параметра к норме с учетом вероятности этого перехода и отклонения от оптимальной стратегии можно представить в виде следующего выражения:

$$\pi^l_{ij} = V^l_{ij} + t \sum_k S_k \sum_l p^l_{\omega}$$

где t – коэффициент, подбираемый в процессе обучения и определяющий, насколько увеличивается вероятность перехода в j из k вершин (т. е. не обнаруженных у пациента). Подобный алгоритм позволяет выбрать как основное, так и компенсирующее (профилактика и терапия осложнений) лечение. Обучение при применении описанного алгоритма сводится как к изменению величины t , так и π^l_σ по следующему правилу:

$$\pi^l_\sigma \begin{cases} \pi^l_\sigma + \Delta\pi, T_\sigma \geq T^l_\sigma, al_\sigma = 0 \\ \pi^l_\sigma - \Delta\pi, T_\sigma < T^l_\sigma, al_\sigma = 1 \\ \pi^l_\sigma, T_\sigma < T^l_\sigma, al_\sigma = 0 \text{ или } T_\sigma \geq T^l_\sigma, al_\sigma = 1 \end{cases}$$

$$\text{где } T_\sigma = t \sum_k S_k \sum_l p^l_{kj}$$



Стратегия управления лечением организма заключается в максимизации описанного функционала. Диагностика или предсказание течения заболевания понимается в этом случае как указание определенной последовательности состояний организма при выбранном лечении.

В результате прогнозируемый ответ состоит из r векторов, каждый из них описывает состояние и все лечебные воздействия, которые должны быть выбраны на каждом этапе течения заболевания (для каждого уровня графа Q). Здесь r – число уровней графа Q , следующих за уровнем, которому принадлежит исходное состояние.

Взаимодействие прогностической системы с внешней средой можно рассматривать по схеме, согласно которой задаются два режима работы: режим экзамена (вопрос – ответ) и режим обучения (вопрос – ответ – оценка).

В нашей системе базовый уровень диагностической (прогностической) информации содержит 346 вершин. На

диагностическом уровне используется значительно меньшее число симптомов и признаков – 120. Наконец, на прогностическом уровне содержится чуть большее число единиц информации – 170. Существенное сокращение объема используемых для решения практических задач признаков заметно сокращает вычислительные операции.

Выводы

1. Предложено соединение предметной и задачно-ориентированной онтологий для структурирования знаний в области экстракорпорального оплодотворения.

2. Рассмотрена концептуальная возможность использования ориентированных графов для задач диагностики состояний пациентов и прогнозирования исходов при вспомогательных репродуктивных технологиях.

3. Предложена многоуровневая онтология, каждый уровень которой предназначен для решения конкретной задачи – диагностики, прогнозирования, выбора метода лечения.

Список литературы

1. Problems of inductive formation of knowledge in the ontology of medical diagnosis / А.С. Клещев, С.В. Смагин // *Automatic Documentation and Mathematical Linguistics*. – 2012. – Vol. 46. – № 1. – С. 8–21.
2. Петряева М.В. Острый панкреатит в формальном представлении для банка знаний медицинской диагностики / М.В. Петряева, М.Ю. Черняховская // *Интеллектуальные системы*. – 2011. – №3(29). – С. 96–103.
3. Harper J. What next for preimplantation genetic screening (PGS)? A position statement from the ESHRE PGD Consortium steering committee / J. Harper, E. Coonen, M. De Rycke, F. Fiorentino [et al.] // *Human Reproduction*. – 2010. – № 2. – С. 1–3.
4. Денисенко С.В. Принципи та логіка побудови бази знань по допоміжним репродуктивним технологіям / С.В. Денисенко // *Медицина інформатика та інженерія*. – 2013. – № 4. – С. 20–23.

References

1. Kleschew, A. S., & Smagin, S. V. (2012). Problems of inductive formation of knowledge in the ontology of medical diagnosis.

Сведения об авторе:

Денисенко С.В., к. мед. н., соискатель каф. медицинской информатики, Национальная медицинская академия последипломного образования имени П.Л. Шупика, E-mail: k-minf05@nmapo.edu.ua.

Automatic Documentation and Mathematical Linguistics, 46(1), 8-21.

2. Petryaeva, M. V. & Chernyakhovskaya, M. Yu. (2011) Ostryj pankreatit v formal'nom predstavlenii dlya banka znaniy medicinskoj diagnostiki [Acute pancreatitis in a formal presentation to the bank of knowledge in medical diagnostics] *Intellektualnyie sistemy*, 3(29), 96–103 [in Russian].
3. Harper, J., Traeger-Synodinos, J., Rycke, M. D., Coonen, E., Vesela, K., SenGupta, S., et al. (2010). What Next For Preimplantation Genetic Screening (PGS)? A Position Statement From The ESHRE PGD Consortium Steering Committee. *Human Reproduction*, 25(4), 821-823.
4. Denysenko, S. V. (2013) Pryntsyppy ta lohika ponudovy bazy znan po dopomizhnym reproduktyvnym tekhnolohiiam [The principles and logic of building a knowledge base for Assisted Reproductive Technology] *Medychna informatyka ta inzheneria*, 4, 20–23 [in Ukrainian].

Поступила в редакцию 03.03.2014 г.