

Применение интеллектуальной системы типа ДСМ для анализа клинических данных

Л. В. Винокурова¹, М. А. Агафонов¹, Г. Г. Варванина¹, В. К. Финн²,

Е. С. Панкратова², Д. А. Добрынин²

¹Центральный научно-исследовательский институт гастроэнтерологии, Москва, Россия

²Всероссийский институт научной и технической информации РАН, Москва, Россия

КЛЮЧЕВЫЕ СЛОВА

хронический панкреатит, интеллектуальная система, анализ клиники, ДСМ-метод, сахарный диабет

В последние годы увеличилось число пациентов с осложненными формами хронического панкреатита (ХП) (20–80 больных на 100 тыс. населения), что приводит к постоянному увеличению тяжести этих осложнений и заставляет чаще прибегать к хирургическому лечению [9]. С дифференциально-диагностической целью, в т. ч. для исключения рака поджелудочной железы (ПЖ), этим больным проводят общеклиническое, лабораторное и инструментальное обследование [6, 10, 12]. Однако достоверность сведений о клинических признаках заболевания очень часто зависит от таких субъективных параметров, как опыт врача и контактность пациента. Применение компьютерной томографии расширило диагностические возможности выявления осложнений при ХП. Развитие медицинских технологий обусловило появление новых инструментальных методик диагностики заболеваний органов панкреатодуоденальной зоны, в т. ч. эндоскопической ультрасонографии с проведением тонкоигльной биопсии. При этом субъективного фактора оценки результатов исследований так же сложно избежать [11].

Классификация 2007 г. M-ANNHEIM включает оценочную балльную систему для определения тяжести заболевания. Подобные системы были разработаны для болезни Крона и некоторых заболеваний печени, при которых эта система помогает правильно выбрать тактику лечения и оценить прогноз заболевания [13].

Градиация панкреатической боли проводится по сумме особенностей течения болевого синдрома и необходимости назначения различных анальгетиков. Например, если пациенту необходимо назначать мощные наркотические анальгетики (2 балла), и у него отмечаются периодические атаки острого панкреатита (3 балла), то тяжесть боли оценивают в 5 баллов. В расчет индекса тяжести ХП включаются все оперативные вмешательства на ПЖ и все тяжелые осложнения панкреатита, начиная с первых проявлений заболевания и далее на протяжении всего анамнеза. Если у пациента имели место два и более тяжелых осложнения, то каждое из них отдельно должно включаться в расчет индекса тяжести ХП. После заполнения оценочной балльной системы M-ANNHEIM необходимо суммировать баллы. Сумма баллов и является индексом тяжести M-ANNHEIM. За последнее время на помощь практикующему врачу все чаще приходят математические методы анализа, математическое моделирование патологических процессов [1, 7].

Материалы и методы

В секторе интеллектуальных систем ВИНТИ РАН создана интеллектуальная система типа ДСМ (ИнтДСМ) [3, 4, 5, 7], основанная на ДСМ-методе автоматического порождения гипотез. Существующая ИнтДСМ включает в себя: ИнтДСМ = Решатель задач + Информационная среда (база фактов (БФ) и база знаний (БЗ)) + Интеллектуальный интерфейс (диалог + представление результатов + научные работы с системой).

Решатель ИнтДСМ основан на ДСМ-методе автоматического порождения гипотез, реализующем автоматизированные правдоподобные рассуждения. Разрабатываемые правдоподобные рассуждения образуют формализованную эвристику извлечения зависимостей причинно-следственного типа из баз структурированных фактов.

Гипотезы о (\pm)-причинах, извлекаемых из БФ, порождаются посредством специально созданной процедуры индукции, основанной на сходстве объектов — источнике или причине наличия (отсутствия) изучаемого эффекта. Предсказание этого эффекта осуществляется посредством аналогии, использующей гипотезы о (\pm)-причинах, содержащихся в БЗ и порожденных индукцией. И, наконец, ДСМ-рассуждение завершается абдуктивной процедурой — объяснением начального состояния БФ, которое является или достаточным основанием для принятия гипотез, или средством расширения БФ для итерации ДСМ-рассуждения, если существуют необъясненные факты из БФ.

Особенности ДСМ-метода:

1. Извлечение знаний типа «причина — следствие» основано на принципе «сходство фактов порождает сходство эффектов и их повторяемость» (этот принцип отличен от вероятностного подхода к анализу данных «повторяемость эффектов определяет сходство фактов»).

2. ДСМ-метод, будучи нестатистическим методом анализа данных, в состоянии учитывать индивидуальные особенности изучаемых объектов исследования.

3. ДСМ-метод способен порождать полезные гипотезы на малых массивах данных.

4. ДСМ-метод работает с открытыми массивами данных (а не с замкнутыми таблицами), указывая на необходимость расширения БФ, если таковая возникает (это осуществляется посредством абдуктивного объяснения БФ).

Условия применимости. ДСМ-метод применим к исследуемой предметной области, т. к. здесь выполняются следующие условия:

1. Возможность структурирования данных и формального определения сходства фактов (из БФ).

2. Наличие положительных и отрицательных примеров (\pm)-примеров в БФ.

3. Наличие в БФ неявно заданных зависимостей причинно-следственного типа ((\pm) -причины изучаемых эффектов).

Постановка задачи

В настоящей статье описано применение ИнтДСМ для прогнозирования развития сахарного диабета панкреатогенного генеза у больных ХП.

Результаты исследования

Создание БФ

Первым этапом совместной работы была разработка подсистемы представления знаний и создание БФ — одной из составных частей ИнтДСМ. Атрибутами этой БФ являются сведения о больных, представленные в медицинских картах и описанные в соответствии с языком представления данных.

Для задачи прогнозирования сахарного диабета панкреатогенного генеза были выделены следующие признаки-атрибуты: возраст больного, индекс массы тела, вредные привычки (алкоголь и табакокурение), длительность заболевания, наличие кальцифицирующего ХП, вирсунголитиаза, панкреатическая гипертензия, резекционные или дренирующие операции в анамнезе, панкреонекрозы в анамнезе, гликемический профиль, уровень С-пептида; оценивали размеры ПЖ, ее контуры, наличие кальцинатов, ширину главного панкреатического протока, концентрации гастроинтестинальных гормонов в периферической крови (глюкозозависимый инсулинотропный полипептид, соматостатин, холецистокинин, секретин, гастрин), наличие развернутой клинической картины сахарного диабета и т. д. (всего 39 признаков). В БФ выполняются требования структурированности данных.

В основу работы положены исследования 158 больных с осложненным ХП. Больные были разделены на 2 возрастные группы: от 20 до 60 лет (119 больных) и от 61 до 74 лет (39 больных).

Настройка на предметную область

Для возможности применения ИнтДСМ к конкретной задаче необходима настройка, включающая следующее:

1. Разработка языка представления данных. Эффективность анализа результатов различных исследований зависит от максимально полного описания обследований больного. Анализируя эти многочисленные медицинские факты, представленные в БФ, из всех возможных выбирается один тип данных: указывается один из возможных качественных признаков, представленных списком.

Пример 1. Резекционные операции в анамнезе: А11 — гПДР; А12 — нПДР; А13 — дистальная резекция; А14 — операция по Бегеру.

Пример 2. Индекс массы тела: А11=16 и меньше — выраженный дефицит массы; А12=16,1–18,5 — недостаток (дефицит) массы тела; А13=18,5–25 — норма; А14=25,1–30 — избыточная масса тела и т. д.

Пример 3. Глюкозозависимый инсулинотропный полипептид: А11 — меньше 0,002 нг/мл; А12 — 0,002–2 нг/мл; А13 — 2–4 нг/мл; А14 — 4–6 нг/мл.

Таким образом, был создан язык представления медицинских данных, разработаны программы ввода медицинской информации. Полученная система содержит автоматизированные истории болезни пациентов. Форма

ввода медицинских данных очень удобна для врача. Можно просматривать БФ, добавлять информацию о заболевании по мере ее получения, дополнительно включать результаты различных исследований. При необходимости возможно наполнение БФ новыми терминами и понятиями.

2. Определение понятия «объект» и «свойство» в терминологии ДСМ-метода. В нашей задаче объект представляет собой упорядоченный набор из 39 клинических, инструментальных и лабораторных признаков. Свойство в данном случае одно — сахарный диабет панкреатогенного генеза.

3. Задание операции сходства. Для выбранного типа данных результатом операции сходства является совпадение признаков.

4. Задание отношения вложения. Отношения вложения определяется по наличию конкретных признаков.

Настройка на экспериментальное исследование

Настройка системы на экспериментальное исследование включает в себя следующее:

1. Выбор стратегии: простой метод сходства, метод сходства с запретом на контрпримеры (невложение полученных методом сходства гипотез в исходные примеры противоположного знака) отдельно для (+)- и (-)-примеров.

2. Подбор нужного количества родителей.

Возможный критерий оценки подбора параметров и стратегии эксперимента — применение процедуры «доопределение по одному»: последовательно каждому объекту выборки присваивается значение «неопределенность», производится доопределение этого объекта средствами ДСМ-системы с выбранными параметрами и сравнивается доопределенное значение с существующим. Подсчитывается общее количество правильных и неправильных доопределений. Выбираются параметры пп. 1–2, при которых при применении процедуры «доопределение по одному» будет наилучший результат, т. е. наибольшее количество правильных доопределений и наименьшее количество неправильных. Эти параметры настройки логично использовать для доопределений вновь заносимых в БФ пациентов, у которых необходимо поставить диагноз.

Результаты компьютерных исследований

Данные для компьютерных исследований были предоставлены сотрудниками Центрального научно-исследовательского института гастроэнтерологии. Первое компьютерное исследование по диагностике сахарного диабета панкреатогенного генеза проводилось на данных обследования 81 больного ХП, у 21 из которых был поставлен диагноз сахарный диабет панкреатогенного генеза — «свойство» в терминологии ДСМ-метода. На прогноз были представлены данные о 42 больных.

Для этого компьютерного исследования с использованием критерия «доопределение по одному» была выбрана стратегия и параметры эксперимента:

1. простой метод с запретом на контрпримеры;

2. количество родителей «+» — 12;

3. количество родителей «-» — 22.

На первом этапе средствами ДСМ-системы правилами правдоподобного вывода 1-го рода (индукцией) на обучающей выборке из 81 больного порождались гипотезы 1-го рода: «Наличие у больного конкретного набора признаков есть причина наличия или отсутствия диагноза сахарный диабет панкреатогенного генеза». Например, набор следующих признаков: «наличие кальцифицирующего ХП; резекционные операции ПЖ в анамнезе; панкреонекрозы в анамнезе; длительность ХП; показатели гликемического

профиля; показатели гастроинтестинальных гормонов, степень выраженности экзокринной недостаточности, болевого синдрома, наличие или отсутствие мутаций».

Порожденные гипотезы являлись фрагментами БЗ.

На втором этапе с использованием гипотез 1-го рода правилами 2-го рода (п.п.в. 2 — аналогией) доопределялось наличие или отсутствие сахарного диабета панкреатогенного генеза у 42 больных, представленных на прогноз.

В результате компьютерных исследований 10 больных были доопределены положительно, т. е. им был поставлен диагноз сахарный диабет панкреатогенного генеза, 28 больных были доопределены отрицательно, т. е. у них не прогнозировался сахарный диабет панкреатогенного генеза, причем 19 из этих больных были старшей возрастной группы. Результаты доопределений совпали с реальными медицинскими данными. Таким образом, в данном компьютерном исследовании точность доопределений — 100%, а полнота — 90%.

Второе компьютерное исследование проводилось на данных обследования 123 больных (к обучающей выборке из 81 больного были присоединены данные о 42 больных, которые в первом компьютерном исследовании были даны на прогноз). У 33 больных был поставлен диагноз сахарный диабет панкреатогенного генеза, а у 89 — отсутствие этого диагноза. Результаты обследования 35 больных были даны на прогноз.

Для этого случая с использованием критерия «доопределение по одному» была выбрана стратегия и параметры эксперимента:

1. простой метод с запретом на контрпримеры;
2. количество родителей («+») — 20;
3. количество родителей («-») — 30.

Результаты компьютерного исследования: в 8 случаях — правильное положительное доопределение, в 24 случаях — правильное отрицательное доопределение, в 1 случае — неправильное положительное доопределение. Таким образом, точность данного исследования — 95%, а полнота — 92%.

Обсуждение результатов

Предлагаемая интеллектуальная система являет-

ся инструментом поддержки медицинских исследований со сложно структурированными данными и множеством фактов, необозримых без использования компьютерных технологий. Результаты применения ДСМ-метода как средства анализа медицинских данных демонстрируют полезность этого метода, являющегося новым инструментом доказательной медицины.

Распространенный сейчас термин «evidence-based medicine», к сожалению, переведен как «доказательная медицина», а не как «медицина, основанная на очевидных фактах». Методы, относящиеся к evidence-based medicine, основаны на установленных фактах и используют эти факты как аргументы при принятии решений в медицине, в частности, для диагностики.

Но смысл этого термина состоит в том, что решения, принимаемые врачами, должны быть аргументированными (но не доказанными — ведь авторы говорят о вероятных прогнозах, т. е. о гипотезах!), что означает, что принимаемые решения используют аргументы, извлеченные из клинических данных.

В самом деле, гипотезы о причинах эффектов, обнаруженные в БФ посредством выявления сходства, являются аргументами или контраргументами в пользу порождаемой интеллектуальной системой гипотезы о наличии или отсутствии изучаемого эффекта.

Сказанное является аргументом в пользу утверждения о том, что ДСМ-метод является полезным аппаратом доказательной медицины [7, 8].

Таким образом, подтверждается возможность использования ДСМ-метода автоматического порождения гипотез в области медицинской диагностики, где недостаток формальных знаний может быть «скомпенсирован» богатым фактическим материалом. В некоторых областях медицины, находящиеся на стадии феноменологического описания и накопления данных, развитые логико-математические методы восстанавливают причинно-следственные зависимости, служат для пополнения знаний на основе имеющихся фактических данных и являются средством поддержки интеллектуальных возможностей экспертов-медиков.

ЛИТЕРАТУРА

1. Алгоритм прогнозирования рецидивирующего непрерывного течения язвенного колита у детей на основе математического моделирования / Э. Н. Федулова, Е. И. Шубанина, Е. А. Жукова [и др.] // Фарматека. — 2014. — № 6. — С. 74–79.
2. ДСМ-метод автоматического порождения гипотез: логические и эпистемологические основания / Под ред. О. Аншакова. — М.: Либроком, 2009. — 432 с.
3. Интеллектуальные системы для анализа медицинских данных. Ч. 1 / В. К. Финн, В. Г. Блинова, Е. С. Панкратова, Е. Ф. Фабрикантова // Врач и информационные технологии. — 2006. — № 5. — С. 62–70.
4. Интеллектуальные системы для анализа медицинских данных. Ч. 2 / В. К. Финн, В. Г. Блинова, Е. С. Панкратова, Е. Ф. Фабрикантова // Врач и информационные технологии. — 2006. — № 6. — С. 50–60.
5. Интеллектуальные системы для анализа медицинских данных. Ч. 3 / В. К. Финн, В. Г. Блинова, Е. С. Панкратова, Е. Ф. Фабрикантова // Врач и информационные технологии. — 2007. — № 1. — С. 51–57.
6. Механизм развития панкреатической недостаточности при хроническом панкреатите / Л. В. Винокурова, И. Е. Трубицына, Е. А. Дубцова [и др.] // Клиническая геронтология. — 2014. — Т. 20. № 1–2. — С. 46–49.
7. О применении интеллектуальной компьютерной системы

- для анализа клинических данных больных меланомой / И. Н. Михайлова, Е. С. Панкратова, Д. А. Добрынин [и др.] // Российский биотерапевтический журнал. — 2010. — Т. 9, № 2. — С. 54.
8. Флетчер Р. Клиническая эпидемиология. Основы доказательной медицины / Р. Флетчер, С. Флетчер, Э. Вагнер / Пер. с англ. под общ. ред. С. Е. Башинского, С. Ю. Варшавского. — М.: Медиа Сфера, 2004. — 347 с.
9. Харнас С. С. Дооперационная и интраоперационная диагностика хирургических заболеваний поджелудочной железы / С. С. Харнас, Ю. А. Кулезнева, Д. Лачман // Хирургия. — 2003. — № 3. — С. 62–65.
10. Хронический панкреатит и рак поджелудочной железы / Л. Б. Лазебник, Л. В. Винокурова, Н. И. Яшина [и др.] // Клиническая и экспериментальная гастроэнтерология. — 2012. — № 7. — С. 3–9.
11. Frequent detection of pancreatic lesions in asymptomatic high-risk individuals / M. I. Cantó, R. H. Hruban, E. K. Fishman [et al.] // Gastroenterology. — 2012. — Vol. 142, No 4. — P. 796–804.
12. Krejs G. J. Pancreatic cancer: epidemiology and risk factors / G. J. Krejs // Dig. Dis. — 2010. — Vol. 28, No 2. — P. 355–358.
13. Schneider A. The M-ANNHEIM classification of chronic pancreatitis: introduction of a unifying classification system based on review of previous classification of the disease / A. Schneider, J.-M. Löhr, M. V. Singer // J. Gastroenterol. — 2007. — Vol. 42. — P. 101–119.

УДК 616-074/-078:614.1]:004.81

ПРИМЕНЕНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ТИПА ДСМ ДЛЯ АНАЛИЗА КЛИНИЧЕСКИХ ДАННЫХЛ. В. Винокурова¹, М. А. Агафонов¹,
Г. Г. Варванина¹, В. К. Финн²Е. С. Панкратова², Д. А. Добрынин²¹Центральный научно-исследовательский институт гастроэнтерологии, Москва, Россия²Всероссийский институт научной и технической информации РАН, Москва, Россия**Ключевые слова:** хронический панкреатит, интеллектуальная система, анализ клиннки, ДСМ-метод, сахарный диабет

В последнее время на помощь практикующему врачу все чаще приходят математические методы анализа и математическое моделирование патологических процессов.

Цель исследования – применить метод ИНТДСМ для прогнозирования развития сахарного диабета панкреатогенного генеза у больных хроническим панкреатитом.**Материалы и методы.** В секторе интеллектуальных систем ВИНТИ РАН создана интеллектуальная система типа ДСМ (ИНТДСМ), основанная на ДСМ-методе автоматического порождения гипотез. Существующая ИНТДСМ включает в себя: ИНТДСМ = Решатель задач + Информационная среда (база фактов и база знаний) + Интеллектуальный интерфейс (диалог + представление результатов + научение работе с системой).**Результаты.** На первом этапе средствами ДСМ-системы правилами правдоподобного вывода 1-го рода (индукцией) на обучающей выборке из 81 больного порождались гипотезы 1-го рода; наличие у больного конкретного набора признаков есть причина наличия или отсутствия диагноза сахарный диабет панкреатогенного генеза.

На втором этапе с использованием гипотез 1-го рода правилами 2-го рода доопределялось наличие или отсутствие сахарного диабета панкреатогенного генеза у 42 больных, представленных на прогноз.

В результате компьютерных исследований 10 больных были доопределены положительно, т. е. им был поставлен диагноз сахарный диабет панкреатогенного генеза, 28 больных были доопределены отрицательно, т. е. у них не прогнозировался сахарный диабет панкреатогенного генеза. Результаты доопределений совпали с реальными медицинскими данными. Таким образом, в данном компьютерном исследовании точность доопределений – 100%, а полнота – 90%.

Заключение. Предлагаемая интеллектуальная система является инструментом поддержки медицинских исследований со сложно структурированными данными и множеством фактов, необозримых без использования компьютерных технологий. Результаты применения ДСМ-метода как средства анализа медицинских данных демонстрируют полезность этого метода, являющегося новым инструментом доказательной медицины.

УДК 616-074/-078:614.1]:004.81

ЗАСТОСУВАННЯ ІНТЕЛЛЕКТУАЛЬНОЇ СИСТЕМИ ТИПУ ДСМ ДЛЯ АНАЛІЗУ КЛІНІЧНИХ ДАНИХЛ. В. Винокурова¹, М. О. Агафонов¹,Г. Г. Варваніна¹, В. К. Фінн²,О. С. Панкратова², Д. А. Добрынин²¹Центральный научно-исследовательский институт гастроэнтерологии, Москва, Россия²Всероссийский институт научной и технической информации РАН, Москва, Россия**Ключові слова:** хронічний панкреатит, інтелектуальна система, аналіз клініки, ДСМ-метод, цукровий діабет

Останнім часом на допомогу практикуючому лікарю все частіше приходять математичні методи аналізу і математичне моделювання патологічних процесів.

Мета дослідження – застосувати метод ІНТДСМ для прогнозування розвитку цукрового діабету панкреатогенного генезу у хворих на хронічний панкреатит.**Матеріали та методи.** У секторі інтелектуальних систем ВІНТИ РАН створена інтелектуальна система типу ДСМ (ІНТДСМ), заснована на ДСМ-методі автоматичного породження гіпотез. Існуюча ІНТДСМ включає в себе: ІНТДСМ = Вирішувач завдань + Інформаційне середовище (база фактів і база знань) + Інтелектуальний інтерфейс (діалог + представлення результатів + навчання роботи з системою).**Результати.** На першому етапі засобами ДСМ-системи правилами правдоподобного виведення 1-го роду (індукцією) на навчальній вибірці з 81 хворого породжувалися гіпотези 1-го роду; наявність у хворого конкретного набору ознак є причина наявності або відсутності діагнозу цукровий діабет панкреатогенного генезу.

На другому етапі з використанням гіпотез 1-го роду правилами 2-го роду довізнається наявність або відсутність цукрового діабету панкреатогенного генезу у 42 хворих, представлених на прогноз.

В результаті комп'ютерних досліджень 10 хворих було довізначено позитивно, тобто їм був поставлений діагноз цукровий діабет панкреатогенного генезу, 28 хворих було довізначено негативно, тобто у них не прогнозувався цукровий діабет панкреатогенного генезу. Результати довізнавання збігалися з реальними медичними даними. Таким чином, у даному комп'ютерному дослідженні точність довізнавання – 100%, а повнота – 90%.

Висновок. Пропонована інтелектуальна система є інструментом підтримки медичних досліджень зі складно структурованими даними і безліччю фактів, неозорих без використання комп'ютерних технологій. Результати застосування ДСМ-методу як засобу аналізу медичних даних демонструють корисність цього методу, що є новим інструментом доказової медицини.**THE USE OF INTELLIGENT JSM-TYPED SYSTEM FOR THE ANALYSIS OF CLINICAL DATA**L.V. Vinokurova¹, M. A. Agafonov¹,G. G. Varvanina¹, V. K. Fynn²,E. S. Pankratova², D. A. Dobrynin²¹Central research institute of gastroenterology, Moscow, Russia²RAS Russian Institute of scientific and technical information, Moscow, Russia**Key words:** chronic pancreatitis, intelligent system, clinical analysis, JSM-method, diabetes. Mathematical methods of analysis and mathematical modeling of pathological processes more often come to help the general practitioners today.**Aim of research** is to apply a JSM-method of data analysis in intelligent systems to predict the development of diabetes of pancreatogenic genesis in patients with chronic pancreatitis.**Materials and methods.** The intelligent JSM-typed system (IntJSM), based on the JSM-method of automatic generation of hypotheses, was created in the sector of intelligent systems of RAS VINITI. Existing IntJSM includes: IntJSM = Solver of tasks + Information environment (base of the facts and knowledge base) + Intelligent interface (dialogue + representation of results + learning to work with system).**Results.** At the first stage, by means of JSM-system by rules of a plausible conclusion of the 1st sort (induction), using the training sample of 81 patients, hypotheses of the 1st sort were generated: presence of a certain set of features in the patient was the reason for presence or absence of the diagnosis diabetes of pancreatogenic genesis.

At the second stage, with the use of hypotheses of the 1st sort by rules of the 2nd sort, presence or absence of diabetes of pancreatogenic genesis was predetermined in 42 patients presented for the prognosis.

As a result of computer research, 10 patients were predetermined positively, i.e. the diabetes of pancreatogenic genesis was diagnosed, 28 patients were predetermined negatively, i.e. they weren't predicted to have the diabetes of pancreatogenic genesis. Results of extension of a definition coincided with real medical data. Thus, the accuracy of predetermination in this computer research was equal to 100%, and completeness – to 90%.

Conclusion. The offered intelligent system is the instrument of support of medical researches with difficult structured data and facts being vast without use of computer technologies. Results of the JSM-method use, as means for the analysis of medical data, demonstrate usefulness of this method, which is the new tool of evidence-based medicine.