

О применении интеллектуальной системы прогнозирования развития диабета у больных хроническим панкреатитом

Аннотация. В статье описаны два эксперимента, проведенные с помощью интеллектуальной системы типа ДСМ (ИС-ДСМ) прогнозирования развития сахарного диабета у больных хроническим панкреатитом. Оба эксперимента проведены по одной схеме, но второй применялся к подмассиву исходных данных (удалили «шум»). Проведен анализ изменений в результатах, формулируются новые задачи для дальнейшего улучшения результатов.

Ключевые слова: ДСМ-метод, интеллектуальные системы, эмпирические закономерности, сахарный диабет, хронический панкреатит.

Введение

Интеллектуальная система для прогнозирования развития сахарного диабета у больных хроническим панкреатитом реализует ДСМ-метод автоматической поддержки научных исследований (ДСМ-метод АПНИ). Условиями применимости ДСМ-метода АПНИ являются [2]:

1. Возможность формализации сходства примеров из базы фактов (БФ);
2. Наличие в БФ положительных и отрицательных примеров (фактов) изучаемого эффекта;
3. Существование в БФ причинных зависимостей для позитивных и негативных фактов;
4. Наличие регулярностей, сохраняющихся при расширении БФ.

Наша задача удовлетворяет этим требованиям применимости. Интеллектуальная система ИС-ДСМ состоит из следующих компонент:

- базы фактов (БФ, $БФ = БФ^+ \cup БФ^- \cup БФ^{\tau}$, соответственно для $\pm$ -фактов и τ -фактов (неопределенностей);
- базы знаний (БЗ) – порожденных гипотез о причинах и гипотез о предсказаниях;
- решателя задач;
- комфортного интерфейса пользователя.

Решатель задач осуществляет ДСМ-рассуждение, которое состоит из итеративного применения правил индуктивного вывода для порождения гипотез о ($\pm$) - причинах изучаемого эффекта (наличия сахарного диабета или его отсутствия) и правил вывода по аналогии для предсказания соответствующих эффектов. ДСМ-рассуждение, помимо правил правдоподобного вывода для индукции и аналогии, содержит процедуру абдуктивного принятия порожденных гипотез [2].

После применения ДСМ-рассуждения Решатель задач выполняет еще один этап – *ДСМ-исследование*, которое применяется для расширений баз фактов и реализует распознавание регулярностей, сохраняющихся при расширении БФ. Регулярности распознаются посредством анализа сохранения типов истинностных значений $v (v \in \{1, -1\})$ порождаемых гипотез (о причинах и предсказаниях). Таким образом, ДСМ-исследование обнаруживает эмпирические закономерности (ЭЗК) [2]. Они устанавливаются относительно последовательности вложенных баз фактов ($БФ_{0,0} \subset БФ_{0,1} \subset \dots \subset БФ_{0,n}$) для выбранной стратегии ДСМ-рассуждения $Str_{x,y}$ и объясняемости этих БФ (абдуктивной сходимости), представимой функцией ρ^{σ} , $\sigma \in \{+, -\}$ [2].

Заметим, что эмпирические закономерности по силе можно разделить на эмпирические тенденции (ЭТ) и эмпирические законы (ЭЗ) [1]. В нашем исследовании мы рассматривали только эмпирические законы.

Ниже дадим краткую характеристику эмпирических законов (ЭЗ) (определение ЭТ и ЭЗ с помощью языка JSM-L дано в [1]).

Пусть объекты (подобъекты) и эффекты представимы посредством исходных множеств $U^{(1)}$ и $U^{(2)}$, соответственно. V, Z, W – переменные, такие что областью определения V, Z и W являются $2^{U^{(1)}}$ и $2^{U^{(2)}}$, соответственно, Q – константа для переменной W . Также пусть у нас есть расширяющиеся вложенные БФ: $БФ_{0,0} \subset БФ_{0,1} \subset \dots \subset БФ_{0,s} \subset \dots \subset БФ_{0,n}$. Посредством языка JSM-L [1] определение «семейства эмпирических законов» $\dot{L}_{1,2}^\sigma(V, Z, W)$, $\sigma \in \{+, -\}$, для заданного эффекта Q формулируется следующим образом:

$\dot{L}_{1,2}^\sigma(V, Z, Q) \Rightarrow \dot{L}_1^\sigma(Z, Q) \& \dot{L}_2^\sigma(V, Q) \& (V \subset Z)$, где $\sigma \in \{+, -\}$, то

1. Предикат $\dot{L}_1^\sigma$ применяется к гипотезам о предсказании и использует функцию сохранения ξ_1^σ , значение которой должно равняться 1. Это означает, что гипотеза на всех расширениях массива получила одинаковые истинностные значения.

2. Предикат $\dot{L}_2^\sigma$ применяется к гипотезам о причинах и использует функцию сохранения ξ_2^σ , значение которой должно равняться 1. Это означает, что гипотеза на всех расширениях массива получила одинаковые истинностные значения.

3. Предикат $(V \subset Z)$ означает, что предсказание из п.1 было получено с помощью гипотезы о причинах из п.2.

4. Предикаты $\dot{L}_1^\sigma(V, Z, W)$ и $\dot{L}_2^\sigma(V, Z, W)$ содержат условие ограничения функции степени абдуктивного объяснения (объясняемости) ρ^σ : $\tilde{\rho}^\sigma \leq \rho^\sigma(s) \leq 1$, $0 \leq s \leq n$, $\sigma \in \{+, -\}$, $\tilde{\rho}^\sigma$ – выбранный порог. Смысл этой функции неформально можно описать как отношение числа σ -примеров исходной БФ, имеющих абдуктивное объяснение посредством гипотез о σ -причинах, к числу всех σ -примеров БФ [2].

Введем следующие определения:

1. Пусть V, Z, W – переменные, значения которых выполняют предикат $\dot{L}_{1,2}^\sigma$, тогда будем называть значение V *причиной ЭЗ*, значение Z – *носителем ЭЗ*, значение W – *эффектом*.

2. Назовем *реализацией ЭЗ* тройку $\langle A, C, Q \rangle$, где $\dot{L}_{1,2}^\sigma(A, C, Q)$ и A, C, Q – константы (значения V, Z, W соответственно).

3. Назовем *механизмом ЭЗ* пару $\langle A, Str_{x,y} \rangle$, $\dot{L}_{1,2}^\sigma(A, Z, W)$ и A – константа для V , $Str_{x,y}$ – средство (стратегия) порождения причины.

4. Назовем *эмпирическим законом (ЭЗ)* пару $\langle A, Q \rangle$, где $\dot{L}_{1,2}^\sigma(A, Z, Q)$ и A, Q – константа для V и W .

1. База фактов

Задачей исследования являлось предсказание сахарного диабета панкреатогенного генеза (3-го типа). Этот тип развивается при осложненном течении хронического панкреатита. В исходный массив данных, составленный в Первом гастроэнтерологическом отделении Центрального НИИ гастроэнтерологии Департамента здравоохранения. Москвы, вошли сведения о больных, у которых поставлены диагнозы диабет панкреатогенного генеза, диабет непанкреатогенного генеза и сведения о больных, у которых диабет не был диагностирован.

В ходе работы с экспертами эти истории болезни были структурированы и собраны в массив данных на специально разработанном языке представления данных [2]. Пример формализованной истории болезни приведен в Приложении 1.

В процессе исследования массив историй болезни пополнялся, что позволило организовать данные в расширяющийся массив вложенных баз фактов ($БФ_{0,1} \subset БФ_{0,2} \subset \dots \subset БФ_{0,n}$) и применить процедуры ДСМ-исследования для анализа эмпирических закономерностей в этих массивах. Для начального набора данных $БФ_{0,0}$ были добавлены два расширения $БФ_{0,1}$ и $БФ_{0,2}$ ($БФ_{0,0} \subset БФ_{0,1} \subset БФ_{0,2}$). На полученных массивах данных было проведено два исследования.

Первое исследование. (+)-примерами в первом исследовании выступили пациенты, у которых был поставлен диагноз сахарный диабет панкреатогенного генеза – «свойство» в терминологии ДСМ-метода. (–)-примерами – пациенты, у которых не был поставлен диагноз диабет или был поставлен диагноз диабет непанкреатогенного генеза (Рис. 1).

$БФ_{0,0}$ содержит данные обследования 81 больного, из них 20 были (+)-примерами, а 61 – (–)-примерами. На прогноз были представлены данные о 10 больных.



Рис.1. Расширения БФ для первого эксперимента

БФ_{0,1} содержит данные обследования 123 больных (к обучающей выборке из 81 больного были присоединены данные о 42 больных. Из них 33 были (+)-примерами, а 90 – (-)-примерами. На прогноз были представлены данные о тех же 10 больных.

БФ_{0,2} содержит данные обследования 155 больных (к обучающей выборке из 123 больных были присоединены данные о 32 больных. Из них 38 были (+)-примерами, а 115 – (-)-примерами. На прогноз были представлены данные о тех же 10 больных.

Второе исследование. В первом исследовании (-)-примерами выступали примеры, имеющие разную природу: пациенты, у которых не был поставлен диагноз Диабет и пациенты, у которых был поставлен диагноз Диабет непанкреатогенного генеза. Также было подсчитано, что количество примеров со вторым диагнозом составляло небольшой процент от общего количества всех примеров. Мы предположили, что удаление этого «шума» из исходного массива может улучшить результаты проводимого исследования. Таким образом, во втором исследовании (+)-примерами выступили пациенты с диабетом панкреатогенного генеза, а (-)-примерами – пациенты без диабета.

В новом исследовании БФ_{0,0} содержало данные обследования 75 больных, из них 21 (+)-пример и 54 – (-)-примера. БФ_{0,1} содержало данные обследования 114 больных (34 (+)-примера и 80 (-)-примеров). БФ_{0,2} содержало данные обследования 138 больных (39 (+)-примера и 99 (-)-примеров). На прогноз в каждом из случаев были представлены данные о тех же 10 больных, что и в первом исследовании (Рис. 2).

2. Ход исследования

Для исследований были выбраны стратегии $\text{Str}_{x,y}$, где $x,y \in \{a^\sigma, (ab)^\sigma, (ad_0)^\sigma, (ad_0b)^\sigma\}$, где a, ab, ad_0, ad_0b – представления (имена) M^σ -предикатов сходства, сходства с запретом на контрпример, различия с запретом на контрпример, M^σ -предикаты используются для определения правил индуктивного вывода [3]. Эти стратегии представимы посредством решетки интенционалов и экстенционалов для правил индуктивного вывода $(I)^\sigma$, $\sigma \in \{+, -, 0, \tau\}$ [3]. Будем обозначать множество выбранных стратегий Str .

Сформулируем общую схему процедур, реализующих обнаружение эмпирических закономерностей.



Рис.2. Расширения БФ для второго эксперимента

Этап 1. Проводим ДСМ-рассуждения с использованием каждой стратегии $\text{Str}_{x,y}$ из $\overline{\text{Str}}$ для каждого из расширений БФ_{0,i}, $i = 0, 1, 2$.

Этап 2. Проводим ДСМ-исследование для обнаружения ЭЗ.

1. Для каждого рассуждения подсчитываем значения объясняемости (ρ^σ) исходного массива (приведены в Табл. 1 Приложений 1 и 2 для первого и второго исследования соответственно) и проверяем подформулу из определения ЭЗ: $\tilde{\rho}^\sigma \leq \rho^\sigma(s) \leq 1$;

2. Для каждого рассуждения все полученные гипотезы (о причинах и предсказаниях) группируем в последовательности полученных ими истинностных значений $v, v \in \{+1, -1, 0, \tau\}$, где $+$ и $-$ - есть сокращения для типов истинностных значений «+1» и «-1». Эти последовательности назовем «цепочками» (множество $\pi_i^{(2)}$ в [2]). Для каждой из полученных «цепочек» вычисляем значение функций сохранения истинностных значений $\tilde{\xi}_i^\sigma$, $\sigma \in \{+, -\}$, $i = 1, 2$.

Табл. 1.

# п/п	БФ _{0,0}	БФ _{0,1}	БФ _{0,2}	Кол-во	$\tilde{\xi}_2^+$	$\tilde{\xi}_2^-$	$\tilde{\xi}_2^0$	$\tilde{\xi}_2^\tau$
1	+	+	+	1	1	#	#	#
2	-	-	-	6	#	1	#	#
3	+	+	τ	1	0,67	#	#	#

Пример составленных на этом шаге таблиц приведен в Табл. 1. Каждой строке таблицы соответствует группа гипотез, принимающая одинаковые истинностные значения для каждого из расширений базы. Например, первой строке таблицы соответствует группа гипотез, получившая в ДСМ-рассуждении для каждого из расширений массива данных тип истинностных значений «+». Таких гипотез в группе одна. Значение функции сохранения истинностного значения $\tilde{\xi}_2^+$ равно 1, для остальных σ , $\sigma \in \{-, 0, \tau\}$ функция $\tilde{\xi}_2^\sigma$ не определена (обозначается #).

3. С помощью составленных на шаге 2 таблиц получаем ЭЗ следующим образом:

Для гипотез о предсказаниях отбираем те гипотезы, которым соответствуют «цепочки» с одинаковыми истинностными значениями на всех расширениях массива (или, что то же самое, значение $\tilde{\xi}_1^\sigma$ для соответствующего σ равняется 1). Эти гипотезы выполняют предикат

$\tilde{L}_1^\sigma(Z, Q)$ ($\sigma \in \{+, -\}$), где Q – константа, означающая возникновение сахарного диабета для (+)-примеров и невозникновение сахарного диабета для (-)-примеров).

Для гипотез о причинах отбираем те гипотезы, которым соответствуют цепочки с одинаковыми истинностными значениями на всех расширениях массива (значение $\tilde{\xi}_2^\sigma$ для соответствующего σ равняется 1). Эти гипотезы выполняют предикат $\tilde{L}_2^\sigma(V, Q)$ ($\sigma \in \{+, -\}$), где Q – константа, означающая возникновение сахарного диабета для (+)-примеров и невозникновение сахарного диабета для (-)-примеров).

Находим пары, выполняющие квазикомпозицию $\exists Z \tilde{L}_{1,2}^\sigma(V, Z, Q) \Leftrightarrow \exists Z (\tilde{L}_1^\sigma(Z, Q) \& \tilde{L}_2^\sigma(V, Q) \& (V \subset Z))$, где $\sigma \in \{+, -\}$. Получаем множество причин $V = \{V | \exists Z \tilde{L}_1^\sigma(Z, W) \& \tilde{L}_2^\sigma(V, W) \& (V \subset Z)\}$.

4. Для каждой стратегии $\text{Str}_{x,y}$ из $\overline{\text{Str}}$ можем указать множество причин V' ($V' \subseteq V$), которые порождаются в рассуждении с использованием этой стратегии, т.е. построить отображение $\tilde{g}: \overline{\text{Str}} \rightarrow 2^{V'}$. Это отображение представлено в Табл. 2 и Табл. 3 Приложений 2, 3.

5. Оцениваем причины (механизмы ЭЗ), отобранные на шаге 3, и предсказания, осуществленные на последнем из расширений БФ – БФ_{0,3}.

Предсказания оценивались через сравнение с их действительными истинностными значениями, предоставленными нам после проведения рассуждений. Для каждой стратегии на последнем из расширений БФ можно подсчитать количество правильных предсказаний (обозначаем l_0), количество грубых ошибок (a) – типов истинностных значений 1 вместо -1 и -1 вместо 1, количество негрубых (b) – типов истинностных значений 0 вместо 1 или -1, количество отказов от предсказаний (c) – типов истинностных значений τ вместо -1 и -1 [2]. Таким образом, можем отсортировать используемые в исследовании стратегии в зависимости от количества правильных предсказаний и типов полученных ошибок и выбрать наиболее успешные. Результаты для всех стратегий представлены в Табл. 4 Приложений 1 и 2. Количество ошибок в таблицах приведено отдельно для (+) и для (-)-примеров.

Список гипотез о причинах был предъявлен для анализа экспертам. Каждую гипотезу предлагалось оценить по надежности числовой оценкой от «4» до «1», где наибольшая оценка

означала, что гипотеза представляется наиболее состоятельной, средние оценки – что отобранные признаки могут иметь место, но неоднозначно свидетельствуют о наличии диабета с осложнением, а наименьшая оценка означала, что гипотеза не имеет смысла. Примеры гипотез с наибольшей и наименьшей оценкой приведены в Приложении 4.

3. Сравнение результатов двух исследований

По результатам проведенного исследования были отмечены следующие улучшения:

1. Объясняемость (ρ^σ) расширяющихся массивов для всех стратегий монотонно возрастает, т.е. выполняется условие $\tilde{\rho}^\sigma \leq \rho^\sigma(0) \leq \dots \leq \rho^\sigma(n) \leq 1$ для всех стратегий из $\overline{\text{Str}}$, в то время как в первом исследовании такое условие не выполнялось для $\text{Str}_{a,ad0}$ и $\text{Str}_{ad0,ad0}$.

2. Для стратегий $\text{Str}_{ab,ad0b}$ и $\text{Str}_{ad0b,ad0b}$ при оценке гипотез о предсказании добавилось одно правильное (-)-предсказание (увеличилось значение правильных предсказаний l_0 и, соответственно, уменьшилось количество ошибок типа с).

3. Гипотеза о причине, единственная получившая наименьшую оценку эксперта в первом эксперименте (№117182315 в Приложении 4), ушла из списка ЭЗ для второго эксперимента.

4. Во втором эксперименте появились две положительные гипотезы о причинах (№154401(73051), №694862(233212) в Приложении 4), которые являются ЭЗ для всех стратегий из $\overline{\text{Str}}$, в то время как в первом исследовании таких гипотез не было.

Заключение

В ходе ДСМ-исследования для каждой стратегии $\text{Str}_{x,y}$ и для каждого расширения $\text{БФ}_{0,i}$ было подсчитано значение теста на обобщенный метод (количество примеров, объясняемых гипотезами противоположного знака) (подробнее сформулирован в [4]) отдельно для (+)- и (-)-примеров.

Значения этого теста (ТОб) приведены в Табл. 1 Приложений 2 и 3 для положительных и отрицательных примеров отдельно. Значение теста на обобщенный метод имеет ненулевое значение, что говорит о возможном успешном применении разбиения массивов исходных данных ($\text{БФ}_{0,0}, \dots, \text{БФ}_{0,s}$) на подмассивы, к которым могут быть отдельно применены метод запрета на контрпримеры и обобщенный ДСМ-метод [4], что в конечном итоге может улучшить предсказание.

Литература

1. Финн В.К. Обнаружение эмпирических закономерностей в последовательностях баз фактов посредством ДСМ-рассуждений//Научно-техническая информация. Серия 2. Информационные процессы и системы. №8. 2015, С.1-29.
2. Шестерникова О.П., Агафонов М.А., Винокурова Л.В., Панкратова Е.С., Финн В.К. Интеллектуальная система прогнозирования развития сахарного диабета у больных хроническим панкреатитом//Искусственный интеллект и принятие решений. № 4. 2015, с. 12-50.
3. Финн В.К. Дистрибутивные решетки индуктивных процедур//Научно-техническая информация. Сер. 2. 2014. №11. С. 1-30.
4. Финн В.К., Шестерникова О.П. О новом варианте обобщенного ДСМ-метода//Искусственный интеллект и принятие решений. № 1. 2016, с. 57-64.

Приложения

Вместо полного указания стратегии $\text{Str}_{x,y}$ будем использовать следующие порядковые номера:

- | | | | |
|----------------------------|-----------------------------|-------------------------------|------------------------------|
| 1) $\text{Str}_{a,a}$, | 5) $\text{Str}_{a,ab}$, | 9) $\text{Str}_{a,ad0}$, | 13) $\text{Str}_{ad0b,ad0b}$ |
| 2) $\text{Str}_{ab,a}$, | 6) $\text{Str}_{ab,ab}$, | 10) $\text{Str}_{ab,ad0}$, | 14) $\text{Str}_{a,ad0b}$, |
| 3) $\text{Str}_{ad0,a}$, | 7) $\text{Str}_{ad0,ab}$, | 11) $\text{Str}_{ad0,ad0}$, | 15) $\text{Str}_{ab,ad0b}$, |
| 4) $\text{Str}_{ad0b,a}$, | 8) $\text{Str}_{ad0b,ab}$, | 12) $\text{Str}_{ad0b,ad0}$, | 16) $\text{Str}_{ad0,ad0b}$ |

Аналогично, 10 примеров для предсказания и соответствующие им гипотезы о предсказании будем нумеровать числами от 1 до 10.

Приложение 1

Пример формализованных данных: пациент №1, поставленный на предсказание

1- Возраст пациента	41-60 лет
2- Индекс массы тела	16,1-18,5 недостаточная (дефицит) масса тела
3- ХКП (хронический кальцифицирующий панкреатит)	да
4- ХГ, контроль (хронический гастрит)	
5- ХХ (хронический холецистит), контроль	
6- СРК (синдром раздраженного кишечника), контроль	
7- ПН (панкреонекроз)	да
8- ХП (рак)	
9- ЖКБ (желчекаменная болезнь)	
10- ДД (дуоденальная дистрофия)	
11- IPMN (доброкачественная внутрипротоковая опухоль поджелудочной железы)	
12- Цистаденома	
13- Вирсунголитиаз	да
14- Деформация панкреатического протока	есть
15- Увеличение размеров поджелудочной железы	головка
16- Резекционные операции в анамнезе	ПДР
18- Алкоголь в анамнезе	да
19- Табакокурение в анамнезе	да
20- Длительность заболевания	более 15 лет
21- Дренирующие операции в анамнезе	
22- Боль	2
23- Кисты	да
24- Кальцинаты	да
25- Панкреатическая гипертензия	да
26- Гликемический профиль ммоль/л	8,1-10
27- Нарушение толерантности к глюкозе	
28- Экзокринная недостаточность	2
29- ХЦК (холецистокинин)	0,16-3
30- секретин	0,023-0,223
31- ГИП (глюкозозависимый инсулиотропный пептид)	0,002-2,002
32- ВИП (вазоактивный инсулиновый гормон)	0,001-4,0
33- соматостатин	0,02-0,42
34- гастрин	0,02-20
35- с-пептид	0,1-2
36- SETR	да
37- PRSS	да
38- SPINK	
39- ХБК-хронический билиарный панкреатит	
40- Сахарный диабет	

Приложение 2

Результаты первого исследования

Табл. 1. Значения объясняемости (абдуктивной сходимости ρ) и теста на обобщенный метод (ТОб) для $\text{Str}_{x,y} \in \overline{\text{Str}}$ и $\text{БФ}_{0,1}, \text{БФ}_{0,2}, \text{БФ}_{0,3}$

Значения объясняемости (абдуктивной сходимости)						
Str _{x,y}	Расширения базы фактов					
	БФ _{0,0}		БФ _{0,1}		БФ _{0,2}	
	БФ _{0,0} ⁺	БФ _{0,0} ⁻	БФ _{0,1} ⁺	БФ _{0,1} ⁻	БФ _{0,2} ⁺	БФ _{0,2} ⁻
1	1	1	1	1	1	1
2	0.85	1	0.97	1	0.97	1
3	1	1	1	1	1	1
4	0.85	1	0.97	1	0.97	1
5	1	0.79	1	0.89	1	0.90
6	0.85	0.79	0.97	0.89	0.97	0.90
7	1	0.79	1	0.89	1	0.96
8	0.85	0.79	0.97	0.89	0.97	0.90
9	1	0.87	1	1	1	0.99
10	0.85	0.93	0.97	1	0.97	1
11	1	0.87	1	1	1	0.99
12	0.85	0.93	0.97	1	0.97	1
13	1	0.66	1	0.68	1	0.68
14	0.85	0.59	0.97	0.68	0.97	0.68
15	1	0.59	1	0.68	1	0.68
16	0.85	0.59	0.97	0.68	0.97	0.68

Значения теста на обобщенный метод						
Str _{x,y}	Расширения базы фактов					
	БФ _{0,0}		БФ _{0,1}		БФ _{0,2}	
	БФ _{0,0} ⁺	БФ _{0,0} ⁻	БФ _{0,1} ⁺	БФ _{0,1} ⁻	БФ _{0,2} ⁺	БФ _{0,2} ⁻
1	1	0.7	1	0.67	1	0.67
2	1	0	1	0	1	0
3	1	0.7	1	0.64	1	0.67
4	1	0	1	1	1	0
5	0	0.77	0	0.72	0	0.72
6	0	0	0	0	0	0
7	0	0.77	0	0.72	0	0.72
8	0	0	0	0	0	0
9	1	0.70	1	0.72	1	0.59
10	1	0	1	0	1	0
11	1	0.70	1	0.72	1	0.72
12	1	0	1	0	1	0
13	0	0.77	0	0.72	0	0.72
14	0	0	0	0	0	0
15	0	0.77	0	0.72	0	0.72
16	0	0	0	0	0	0

Табл. 2. Таблица соответствия положительных ЭЗ стратегиям ($g: \overline{\text{Str}} \rightarrow 2^V$)

$\text{Str}_{x,y}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Идентификаторы ЭЗ⁺																
154401	1		1		1		1		1		1		1		1	
694862		1		1	1	1	1	1	1	1	1	1	1	1	1	1
239, 1509, 5028, 124738, 126721, 131411 , 131880, 167074, 347776, 358062, 372242, 690520, 692023, 693860, 694194, 696198, 697367																
3668, 4858, 4943, 130942, 688015, 688182, 688349, 688683, 689017, 693693, 695196					1		1						1		1	

В клетках таблицы содержатся булевские значения: «1» – «Истина», «0» – «Ложь». Значение «Истина» в клетке на пересечении строки с ЭЗ и столбца со стратегией означает, что эта причина реализована этой стратегией. Гипотезы, выделенные жирным шрифтом, приведены в Приложении 4.

Табл. 3. Таблица соответствия отрицательных ЭЗ стратегиям ($g: \overline{\text{Str}} \rightarrow 2^V$)

Комментарий см. к Табл. 2.

$\text{Str}_{x,y}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Идентификаторы ЭЗ⁻																
98862, 99640, 100418, 108198,	1	1	1	1					1	1	1	1				
1526151	1	1	1	1												
2370477		1		1					1		1					
100807 , 516144, 539344, 564151, 1549329, 2036199, 2148924, 2161369, 2258245, 2272340, 2282419, 2365792, 3161644, 6674224, 22783570 , 22980952, 22987123, 23289353, 23388089, 23394260, 23400431, 26308390, 26525757, 40951161, 117153607, 117182315 , 117185119, 117188923, 117923651																

Табл. 4. Таблица для применяемого множества стратегий (типы ошибок)

Str _{x,y}	l ₀		a		b		c	
	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)
(6), (8)	3	5			1	1		
(14), (16)	4	2		1				3
(2),(4),(10),(12)		5			4	1		
(5),(7)	3	1			1	5		
(13),(15)	4			3		2		1
(1), (3), (9), (11)		1			4	5		

Приложение 3

Результаты второго исследования

Табл. 1. Значения объясняемости (абдуктивной сходимости ρ) и теста на обобщенный метод (ТОб) для Str_{x,y} ∈ Str̄ и БФ_{0,1}, БФ_{0,2}, БФ_{0,3}

Значения объясняемости (абдуктивной сходимости)							Значения теста на обобщенный метод (ТОБ)						
Str _{x,y}	Расширения базы фактов						Str _{x,y}	Расширения базы фактов					
	БФ _{0,0}		БФ _{0,1}		БФ _{0,2}			БФ _{0,0}		БФ _{0,1}		БФ _{0,2}	
	БФ _{0,0} ⁺	БФ _{0,0} ⁻	БФ _{0,1} ⁺	БФ _{0,1} ⁻	БФ _{0,2} ⁺	БФ _{0,2} ⁻		БФ _{0,0} ⁺	БФ _{0,0} ⁻	БФ _{0,1} ⁺	БФ _{0,1} ⁻	БФ _{0,2} ⁺	БФ _{0,2} ⁻
1	1	1	1	1	1	1	1	1	0.74	1	0.69	1	0.68
2	0.95	1	0.97	1	0.97	1	2	1	0	1	0	1	0
3	1	1	1	1	1	1	3	1	0.74	1	0.66	1	0.68
4	0.95	1	0.97	1	0.97	1	4	1	0	1	0	1	0
5	1	0.8	1	0.86	1	0.95	5	0	0.81	0	0.74	0	0.73
6	0.95	0.8	0.97	0.86	0.97	0.95	6	0	0	0	0	0	0
7	1	0.8	1;	0.86	1	0.95	7	0	0.81	0	0.74	0	0.73
8	0.95	0.8	0.97	0.86	0.97	0.95	8	0	0	0	0	0	0
9	1	0.85	1	1	1	1	9	1	0.74	1	0.74	1	0.73
10	0.95	0.93	0.97	1	0.97	1	10	1	0	1	0	1	0
11	1	0.85	1	1	1	1	11	1	0.74	1	0.74	1	0.73
12	0.95	0.93	0.97	0.99	0.97	1	12	1	0	1	0	1	0
13	1	0.57	1	0.67	1	0.75	13	0	0.87	0	0.72	0	0.73
14	0.9	0.57	0.97	0.67	0.97	0.75	14	0	0	0	0	0	0
15	1	0.57	1	0.67	1	0.75	15	0	0.81	0	0.72	0	0.73
16	0.9	0.57	0.9	0.68	0.97	0.75	16	0	0	0	0	0	0

Табл. 2. Таблица соответствия положительных ЭЗ стратегиям ($g: \overline{\text{Str}} \rightarrow 2^V$)

Str _{x,y}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Идентификаторы ЭЗ ⁺	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
73051, 233212	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
111406, 123598, 208129, 215778, 228270 , 892299, 892705, 897577, 901637, 912968		1		1	1	1	1	1		1		1				
224, 239, 9610, 9868, 9997, 13609, 76812, 91428, 92571, 123979, 213900, 216404, 218908, 219534, 224514, 228896, 891284, 891487, 891893, 893720, 895547, 899607, 899810, 900013, 900825, 902043					1		1									
9352, 9481, 890878, 891690, 892908					1		1						1		1	

В клетках таблицы содержатся булевские значения: «1» – «Истина», «0» – «Ложь». Значение «Истина» в клетке на пересечении строки с ЭЗ и столбца со стратегией означает, что эта причина реализована этой стратегией. Гипотезы, выделенные жирным шрифтом приведены в Приложении 4.

Табл.3. Таблица соответствия отрицательных ЭЗ стратегиям ($g: \overline{\text{Str}} \rightarrow 2^V$)
Комментарий см. к Табл.2.

$\text{Str}_{x,y}$ Идентификаторы ЭЗ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
13465, 306925, 322701, 324091, 954885, 968138	1	1	1	1					1		1	1				
303545	1	1	1	1												
595216		1		1								1				
311657, 472011, 476985, 588594, 841042, 1716322, 1717931, 8270657, 8489900, 8497634, 8669678, 8716082, 8750885, 10008740, 10103614 , 10183375, 10214311, 10222045, 10318720, 10326454, 14361602, 14905007, 15300932, 17306555, 17747920, 17997826, 18117435, 20886645, 29469330, 30511979, 30784658, 86476672 , 86480230, 86484091, 86486046, 86488001, 86489956, 86516190, 86523421, 86526010, 86527965, 86529920, 86531875, 86533830, 86537325, 86539280, 86541235, 86543190, 86548710, 86552291, 86554246, 86572318, 87665560, 87775717, 87831824, 87839628, 88072880																
					1	1	1	1								

Табл.4. Таблица для применяемого множества стратегий (типы ошибок)

$\text{Str}_{x,y}$	I_0		a		b		c	
	+	-	+	-	+	-	+	-
(6),(8)	3	5			1	1		
(14), (16)	4	3		1				2
(2), (4), (10), (12)		5			4	1		
(13), (15)	4			2		3		1
(5), (7)	3	1			1	5		
(1), (3), (9),(11)		1			4	5		

Приложение 4

Примеры наиболее интересных причин ЭЗ

В приложении представлены гипотезы о причинах ЭЗ, упоминающиеся в статье, а также гипотезы, получившие лучшие и худшие экспертные оценки. Гипотезы идентифицируются по порядковому номеру, полученному при эксперименте. Если номер гипотезы во втором эксперименте отличается от первого, то он указан в скобках.

№154401 (73051)

Гипотеза о причине возникновения эффекта, порождаемая всеми рассуждениями во втором эксперименте.

33- соматостатин	0,02-0,42
35- с-пептид	0,1-2

№694862(233212)

Гипотеза о причине возникновения эффекта, порождаемая всеми рассуждениями во втором эксперименте и большинством рассуждений в первом. Получила наибольшую оценку эксперта.

28- Экзокринная недостаточность	2
33- соматостатин	0,02-0,42
35- с-пептид	0,1-2

№131411(228270)

Гипотеза о причине возникновения эффекта, получившая наибольшую оценку эксперта.

14- Деформация панкреатического протока	есть
29- ХЦК (холецистокинин)	0,16-3
33- соматостатин	0,02-0,42
35- с-пептид	0,1-2

№22783570 (86476672)

Гипотеза о причине отсутствия эффекта, получившая наибольшую оценку эксперта.

26- Гликемический профиль ммоль/л	4,5-6,5
29- ХЦК (холецистокинин)	0,16-3
32- ВИП (вазоактивный инсулиновый гормон)	0,001-4,0
33- соматостатин	0,02-0,42
34- гастрин	0,02-20

№100807(10103614)

Гипотеза о причине отсутствия эффекта, получившая наибольшую оценку эксперта.

26- Гликемический профиль ммоль/л	4,5-6,5
32- ВИП (вазоактивный инсулиновый гормон)	0,001-4,0
33- соматостатин	0,02-0,42
34- гастрин	0,02-20

№117182315

Гипотеза о причине отсутствия эффекта, получившая наименьшую оценку эксперта в начальном эксперименте. Не порождается во втором эксперименте.

20- Длительность заболевания	5-10
28- Экзокринная недостаточность	1

Шестерникова Ольга Павловна. Начальник отдела ООО «Прогтек». Окончила Российский государственный гуманитарный университет в 2010 году. Автор трех печатных работ. Область научных интересов: компьютерные интеллектуальные системы, программирование, анализ данных. E-mail: oshesternikova@gmail.com